

OPTIMIZING SUPERVISED MODELS FOR BENIGN-MALIGNANT BREAST TUMOR CLASSIFICATION USING ADVANCED FEATURE ENGINEERING

Muhammad Haris Malik^{*1}, Ifrah Ikram², Muhammad Abdul Rafay³, Umama Abbasi⁴,
Talha Ahmed⁵

^{*1,3,4}Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST),
Islamabad, Pakistan

²Department of Computer Science, COMSATS University Islamabad, Islamabad

⁵Department of Robotics and Artificial Intelligence, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology
(SZABIST), Islamabad, Pakistan

^{*1}malikharis5161@gmail.com, ²ifrah.ikram89@gmail.com, ³arafay000313@gmail.com,
⁴umamahabbasi24@gmail.com, ⁵talha.ahmad058@gmail.com

DOI: <https://doi.org/10.5281/zenodo.17035247>

Keywords

Breast Cancer, Supervised Machine learning, Feature Selection, Wisconsin Dataset, Diagnostic Accuracy, Logistic Regression, ROCAUC

Article History

Received: 30 May, 2025

Accepted: 05 August, 2025

Published: 30 August, 2025

Copyright @Author

Corresponding Author: *
Muhammad Haris Malik

Abstract

Breast cancer remains a leading cause of mortality among women, necessitating advanced diagnostic tools to enhance early detection. Supervised machine learning offers promising capabilities for accurate tumor classification. This study aims to develop a robust pipeline for classifying breast tumors as benign or malignant using the Breast Cancer Wisconsin Dataset, optimizing diagnostic accuracy through advanced feature engineering. The methodology includes data preprocessing with mean imputation and min-max normalization to ensure data consistency, followed by Exploratory Data Analysis using violin plots to visualize feature distributions and correlation heatmaps to detect multicollinearity. Feature selection employed correlation-based filtering, Chi-square tests, tree-based methods, Recursive Feature Elimination (RFE), RFECV, and L1 regularization, selecting 16 key features to enhance model performance. Seven supervised algorithms—Logistic Regression, Random Forest, SVM, Extra Trees, KNN, Decision Trees, and Naïve Bayes—were trained and optimized via grid and randomized search, evaluated with stratified 10-fold cross-validation for robust assessment. Logistic Regression with Random Forest-based feature selection achieved a test accuracy of 0.977, ROCAUC of 0.971, and average precision of 0.99, reflecting superior performance. However, L1 regularization underperformed due to over-penalization, and SVM exhibited sensitivity to feature scaling, requiring careful tuning. The proposed pipeline significantly improves diagnostic accuracy, offering a scalable, interpretable solution for clinical application.

INTRODUCTION

Breast cancer remains one of the most prevalent and deadly diseases affecting women globally, posing significant challenges to healthcare systems worldwide. According to the World Health

Organization, breast cancer accounted for approximately 2.3 million new cases in 2020 alone, making it the most common cancer among women and a leading cause of cancer-related mortality [1]. In

many regions, including developing countries, the incidence rates are rising due to factors such as aging populations, lifestyle changes, and delayed childbearing. Despite advances in medical technology, the disease continues to claim over 685,000 lives annually, with disparities in survival rates between high-income and low-income countries highlighting issues of access to early detection and treatment [2]. Early diagnosis is crucial, as it can increase five-year survival rates to over 90% when detected at localized stages, compared to less than 30% for metastatic cases [3]. However, traditional diagnostic methods like mammography, biopsy, and clinical examination often suffer from limitations including subjectivity, accessibility issues, and varying accuracy rates, particularly in dense breast tissue where mammography sensitivity can drop to as low as 57% [4].

The advent of machine learning (ML) has revolutionized medical diagnostics by enabling the analysis of complex datasets to uncover patterns that may elude human observers. Supervised ML, in particular, has shown promise in classifying breast tumors as benign or malignant by learning from labeled data [5]. Unlike unsupervised methods that rely on clustering without predefined outcomes, supervised approaches leverage historical data to train models for prediction, achieving higher accuracy in structured datasets like the Breast Cancer Wisconsin Dataset [6]. Recent studies have demonstrated that integrating feature engineering with supervised models can enhance performance, with techniques like Recursive Feature Elimination (RFE) and Chi-square testing improving classification metrics by over 10% [7]. For instance, models such as Random Forest (RF) and Support Vector Machines (SVM) have been optimized through hyperparameter tuning to yield precision and recall scores exceeding 95% [8]. However, challenges persist, including class imbalance in datasets, overfitting in high-dimensional spaces, and the need for interpretable models in clinical settings [9].

A key research gap lies in the integration of advanced feature selection with multiple supervised algorithms to create reproducible pipelines that are both accurate and scalable. Many existing studies focus on single algorithms or limited feature sets, often overlooking comprehensive preprocessing and

hyperparameter optimization, which can lead to suboptimal performance in real-world applications [10]. For example, while ensemble methods like Gradient Boosting (GB) have shown robustness in handling noisy data, their application in conjunction with cross-validation techniques like RFECV remains underexplored [11]. Additionally, the reliance on traditional imaging alone ignores the potential of combining multimodal data, such as genomics with histopathological features, to improve diagnostic reliability [12]. This gap is exacerbated in resource-limited settings where access to advanced imaging is scarce, necessitating models that perform well on standard datasets with minimal computational overhead [13].

Furthermore, ethical considerations in ML for healthcare, such as bias in training data leading to disparities in diagnosis across ethnic groups, have not been adequately addressed [14]. Studies indicate that models trained on predominantly Caucasian datasets may underperform for diverse populations, underscoring the need for inclusive data strategies [15]. Deep learning extensions, like Convolutional Neural Networks (CNNs), have emerged as alternatives, but their black-box nature limits clinical adoption due to lack of explainability [16]. Explainable AI (XAI) techniques, such as SHAP values, could bridge this, but integration with supervised models is rare [17]. The evolving landscape also includes bioinspired algorithms and meta-learning, which promise better generalizability but require validation on larger cohorts [18].

This study addresses these gaps by proposing an optimized supervised ML pipeline for benign-malignant breast tumor classification using advanced feature engineering. The primary objectives are: (1) To develop a comprehensive preprocessing framework including imputation, normalization, and EDA to handle real-world data inconsistencies; (2) To evaluate and select optimal features from multiple techniques, including correlation analysis, Chi-square, tree-based importance, RFE, RFECV, and L1 regularization, aiming to reduce dimensionality while retaining predictive power; (3) To train and optimize seven supervised algorithms—Logistic Regression (LR), RF, SVM, Extra Trees (ET), K-Nearest Neighbors (KNN), Decision Trees (DT), and Naïve Bayes (NB)—using grid and randomized

search with stratified 10-fold cross-validation; (4) To assess performance using metrics like accuracy, precision, recall, F1-score, and ROC-AUC, identifying the best model for clinical scalability; (5) To explore limitations such as over-penalization in L1 and scaling sensitivity in SVM, proposing mitigations; and (6) To outline future directions, including integration with deep learning and larger datasets for enhanced accessibility.

By focusing on the Breast Cancer Wisconsin Dataset, which includes 569 instances with 30 features derived from fine-needle aspirates, this research ensures reproducibility and relevance [19]. The pipeline's emphasis on interpretability and efficiency positions it as a tool for healthcare providers, potentially reducing diagnostic errors and improving outcomes [20]. Ultimately, this work contributes to bridging the gap between ML advancements and practical clinical application, fostering equitable and effective breast cancer diagnosis globally [21, 22, 23, 24, 25, 26, 27].

2) Literature Review

The application of machine learning (ML) in breast cancer diagnosis has evolved significantly, with numerous studies demonstrating its potential to improve accuracy and efficiency. Yedjou et al. [1] explored ML algorithms for breast cancer classification using the Wisconsin Dataset, analyzing features like radius and texture. Their study, involving 569 patients, found 63% benign and 37% malignant cases, emphasizing ML's role in distinguishing tumor types without invasive procedures. They achieved high accuracy by combining physical exams with ML, highlighting limitations in traditional methods like mammography's low contrast.

Shia et al. [2] utilized unsupervised ML for classifying malignant tumors in breast ultrasound images, analyzing 677 patients with BI-RADS category 4 lesions. Employing histogram of oriented gradients (HOG) and pyramid HOG (PHOG) for feature extraction, followed by correlation-based selection and locally weighted learning, they reported an AUC of 0.847, outperforming physicians (AUC 0.574). This underscores unsupervised methods' ability to reduce observer variability and human effort in non-ROI-based analysis.

Al-Azzam and Shatnawi [3] compared supervised and semi-supervised models on the Wisconsin Dataset, finding supervised approaches superior with accuracies up to 98.57% using ANNs. They noted semi-supervised limitations in small datasets, advocating for larger cohorts to validate results. Gharibdousti et al. [4] applied feature extraction with supervised and unsupervised classifiers, using principal component analysis and logistic regression for reduction. Their hybrid discriminant-logistic model achieved top performance, emphasizing multivariate techniques' role in filtering pertinent features from complex tumor data.

Fanoos and Abdulhadi [5] presented supervised ML for breast cancer diagnosis, comparing algorithms like SVM and RF on the Wisconsin Dataset. SVM excelled with 97.2% accuracy, proving ML's efficacy in early detection and updating treatment standards. Dadheech et al. [6] focused on feature engineering and hyperparameter tuning, using RF and LR to achieve high precision. They underlined optimization's importance in structured datasets.

Uddin et al. [7] employed feature optimization with ML, using RFE and Chi-square on biomedical data. Their approach improved results, highlighting optimization's value in noisy datasets. Chaurasia and Pal [8] predicted diagnostic breast cancer with six ML algorithms, reducing features to 12 using statistical methods. Ensemble models minimized misclassification, achieving over 90% accuracy, demonstrating dimensionality reduction's benefits.

Wolberg [9] introduced the Wisconsin Dataset, foundational for ML research, with 569 instances and 30 features from FNAs, enabling benchmark comparisons. Ginsburg et al. [10] advocated phased implementation for early detection, noting ML's potential in complementing traditional tools like mammography. Kilic and Karakoyun [11] detected breast cancer with ML, comparing algorithms on datasets, achieving high precision with SVM.

Devi et al. [12] used ML and DL for prediction, with MLP outperforming at 96.49% accuracy on Wisconsin data, highlighting DL's edge over traditional ML. Priya and Vadivu [13] proposed bio-inspired ensemble feature selection with kernel extreme learning machines, integrating optimization techniques for enhanced accuracy and interpretability. Sakib et al. [14] compared ML

algorithms, with SVM at 97.13% accuracy, emphasizing efficiency in classification.

Bekkouche et al. [15] tailored DL for early detection, proving superior results over traditional methods. Iranmakani et al. [16] reviewed breast imaging modalities, noting ML's complementary role in enhancing radiological accuracy. Islam et al. [17] used ML with XAI for prediction in Bangladeshi patients, XGBoost at 97% accuracy, SHAP for interpretability. Guo et al. [18] provided insights into ML for diagnosis, benchmarking supervised models. Gupta and Gupta [19] compared techniques, SVM at 97.13% accuracy. Rabiei [20] predicted using ML, addressing class imbalance with ensembles. Ali et al.

[21] classified via meta-learning with CNNs, achieving 90% accuracy on ultrasound. Mahesh et al. [22] enhanced CNNs for diagnosis, 95.2% accuracy. Zhou et al. [23] developed AST-GNN for predictive modeling, though not breast-specific, insights applicable. Fan et al. [24] introduced adaptive multiclass feature selection.

Zou et al. [25] used deep networks for prediction. Zhou et al. [26] proposed unsupervised selection for clustering. Navarro et al. [27] improved whale optimization, enhancing ML efficiency. These studies collectively highlight supervised ML's dominance, with gaps in hybrid approaches and larger datasets for generalizability.

3) Proposed Method

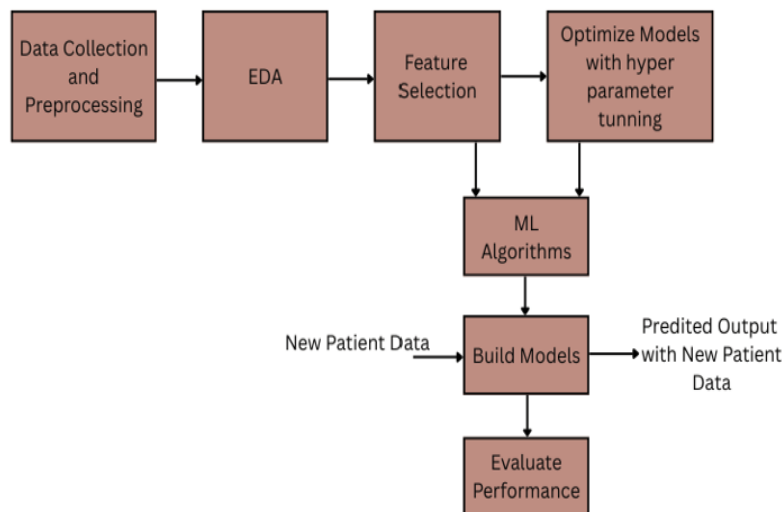


Figure 1: Architecture of the Proposed Framework.

This study employs a systematic and reproducible pipeline for breast tumor classification, encompassing data collection, preprocessing, exploratory analysis, feature selection, model training, optimization, and prediction. The method is designed for the Breast Cancer Wisconsin Dataset but is adaptable to similar datasets. Figure 1 illustrates the architecture.

A. Data Collection and Preprocessing

The dataset comprises 569 instances with 30 features from fine-needle aspirates, labeled as benign (357) or malignant (212). Features include mean, standard error, and worst values for radius, texture, perimeter,

area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension Figure 2. Missing values are handled via mean imputation:

$$\text{Mean Imputaion: } x_i = \frac{\sum_{j=1}^n x_j}{n} \quad (\text{Equation 1})$$

where n is the number of non-missing values. This preserves data distribution without introducing bias in normally distributed features.

Features vary in scale, affecting distance-based models like SVM. Min-max normalization scales to $[0,1]$:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (\text{Equation 2})$$

This ensures equal contribution during training.

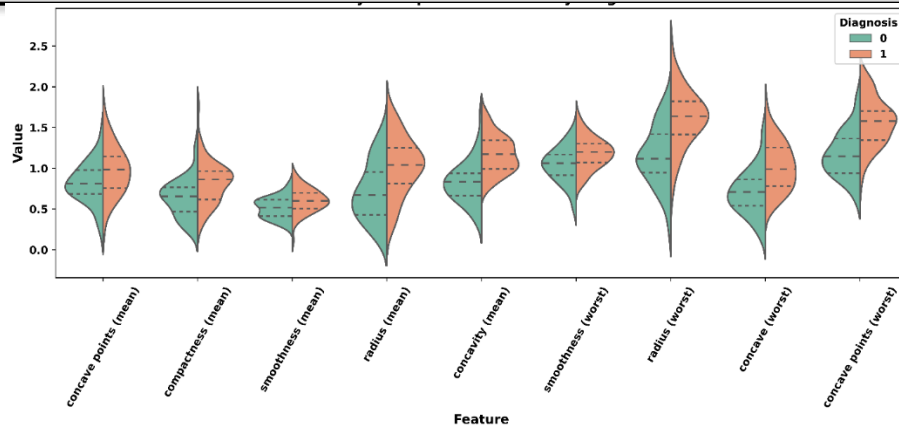


Figure 2: Violin Plot for Distribution and Density

1) Encoding Categorical Variables

Since ML models cannot directly process non-numeric data, such as labels or categories, encoding converts them into numeric formats. One-hot encoding converts categories into binary vectors, making it suitable for non-ordinal data. For example, for a feature Color with categories {Red, Blue, Green}, one-hot encoding results in:

Red → [1, 0, 0], *Blue* → [0, 1, 0], *Green* → [0, 0, 1]

Label encoding assigns a unique numeric value to each category, which is suitable for ordinal data. For example, for Education Level with categories {High School, Bachelor's, Master's}:

High School → 1, *Bachelor's* → 2, *Master's* → 3

B. Exploratory Data Analysis (EDA)

This study highlights differences between benign and malignant tumor samples and examines model stability by using EDA techniques to find patterns and correlations in numerical data. A correlation matrix is computed, showing the correlation coefficients (r) for every pair of features:

$$\text{Correlation } r_{ij} = \frac{\text{Cov}(x_i, x_j)}{\sigma_{x_i} \sigma_{x_j}} \quad (\text{Equation 3})$$

Where r_{ij} is the correlation between features x_i and x_j , Cov is the covariance, and σ is the standard deviation. The matrix is represented as a heatmap in Figure 3, with darker colors indicating stronger correlations.

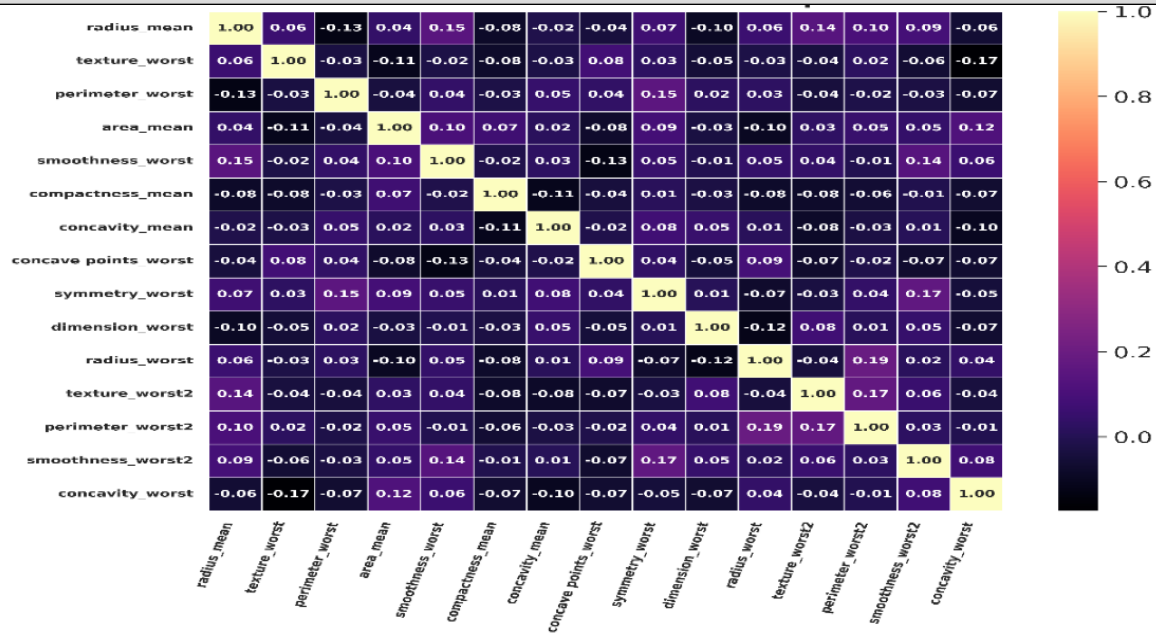


Figure 3: Correlation Heatmap.

C. Feature Selection

Feature selection is a vital step to identify and retain the most informative features from a dataset while removing redundant or irrelevant ones. Highly correlated features provide redundant information, and they were eliminated using:

$$|r_{ij}| > threshold \quad (\text{Equation 4})$$

The Chi-Square statistic was computed for each feature to measure the dependency on the target variable.

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (\text{Equation 5})$$

where O is the observed frequency and E is the expected frequency. An RF was trained to calculate the feature importance l_j and rank features based on their scores..

$$l_j = \frac{1}{N} \sum_{t=1}^N \Delta G_t(j) \quad (\text{Equation 6})$$

where $\Delta G_t(j)$ is the decrease in Gini impurity for feature j in tree t, and N is the total number of trees. A Lasso regression was used to penalize less important features and shrink their coefficients to zero.

$$\min\left(\frac{1}{2n} \sum_{i=1}^n (y_i - y'_i)^2 + \lambda \sum_{j=1}^p |w_j|\right) \quad (\text{Equation 7})$$

where w_j are feature weights, and λ is the regularization parameter. With their importance and contributions detailed in Table 1 and Figure 4, the vote-based choice found 16 essential features for classification, ensuring the model's accuracy and efficiency in detecting breast cancer.

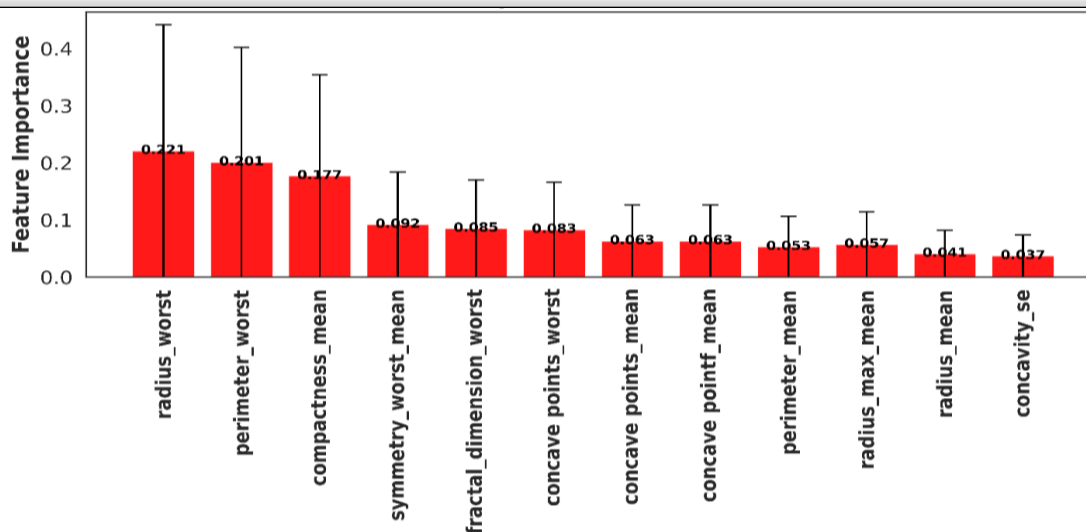


Figure 4: Selection of the most important features.

Table 1: Voting-Based Selected Features

Votes	Index	Chi square	RF	Extratrees	RFE	RFECV	L1	Final R score
0	texture_mean	1	1	1	1	1	1	6
1	area_mean	1	1	1	1	1	1	6
7	area_se	1	1	1	1	1	1	6
3	concavity_mean	1	1	1	1	1	0	5
9	concavity_se	1	1	1	1	1	0	5
13	concavity_worst	1	1	1	1	1	0	5
14	symmetry_worst	1	1	1	1	1	0	5
12	smoothness_worst	1	1	1	1	0	0	4
15	fractal_dimension_worst	1	1	1	0	0	0	3
2	smoothness_mean	0	0	0	1	1	0	2
5	fractal_dimension_mean	0	0	1	0	1	0	2
4	symmetry_mean	1	0	0	0	0	0	1
8	smoothness_se	0	0	0	1	0	0	1
11	fractal_dimension_se	0	1	0	0	0	0	1
6	texture_se	0	0	0	0	0	0	0
10	symmetry_se	0	0	0	0	0	0	0

D. Hyperparameter Optimization

Hyperparameter tuning aims to increase model generalization and predictive accuracy. This study used grid search and randomized search.

E. Proposed Algorithm

Step 1: Data Collection and Preprocessing

Input: Breast Cancer Wisconsin Data Set. Handle Missing Values: For each feature with missing values, calculate the mean or median using equation

1. Normalize Features: For each numerical feature x , apply min-max scaling using equation 2. Encode Categorical Variables: One-hot encoding for non-ordinal data: if $x \in \{a, b, c\}$, encode as $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$. Label encoding for ordinal data: $x \in \{a, b, c\} \rightarrow x' \in \{1, 2, 3\}$

Step 2: Exploratory Data Analysis

Visualize Distributions: Use Violin Plots to observe the density and spread for each feature. Correlation

Analysis: Compute the correlation matrix using Equation 3. Visualize the matrix using a heatmap to identify multicollinearity

Step 3: Feature Selection

Correlation-Based Selection: Identify feature pairs (X_i, X_j) using Equation 3 and 4. Retain only one feature from highly correlated pairs. Chi-square Test: Compute the Chi-Square statistic for each feature with respect to the target Equation 5. Recursive Feature Elimination (RFE): Fit a model $f(x)$, calculate feature importance scores, and iteratively remove the least important features until k features remain. RF-based Selection: Train an RF model to compute feature importance l_j Equation 6 and retain features with $l_j > \text{threshold}$. L1 Regularization (Lasso): Fit a Lasso regression model using equation 7 and retain features with $W_j \neq 0$.

Step 4: Model Training and Evaluation

Split Data: Divide the dataset into training and testing sets (e.g., 80:20 split). Train models: $M(X) = y$, where $M \in \{LR, RF, SVM, ET, KNN, DT, NB\}$. Fit multiple machine learning models $f(X)=Y'$ where $Y' \in$

{Benign, Malignant}. Optimize Models: Use grid search or random search for hyperparameter tuning with cross-validation. Evolution Metrics: Accuracy, Precision, Recall, F1-Score calculated.

Step 5: Prediction

Input: New patient data X_{new} . Preprocess Input: Apply the same preprocessing and feature selection steps as used in training. Make Predictions: Use the trained model $f(x)$ to predict: $Y_{\text{new}}=f(X_{\text{new}})$. Output: Selected features, trained model, and predictions for new data Y_{new} .

4) Results

The evaluation encompassed 56 supervised ML models across seven algorithms: LR, RF, SVM, ET, KNN, DT, and NB, applied to eight feature selection strategies on the Wisconsin Dataset (569 records, 30 features). Stratified 10-fold CV assessed performance via confusion matrices, precision-recall curves, and ROC-AUC. LR with RF importance emerged best, with ROC-AUC 0.97 and AP 0.99.

Table 2. Training Performance 5 Fold Cross-Validation Scores On Training Data

Algorithm	Feature Selection	Accuracy	Avg Precision	F1	Precision	Recall	ROC-AUC
LR	Correlation	0.9775	0.9936	0.9700	0.9804	0.9611	0.9939
SVM	Chi2	0.9648	0.9914	0.9532	0.9670	0.9411	0.9936
SVM	RFE	0.9626	0.9908	0.9481	0.9739	0.9286	0.9930
LR	RFECV	0.9724	0.9933	0.9632	0.9800	0.9477	0.9938
LR	RF	0.9774	0.9945	0.9704	0.9742	0.9675	0.9947
LR	Extra trees	0.9725	0.9931	0.9635	0.9752	0.9546	0.9931
LR	L1	0.8994	0.9514	0.8601	0.9214	0.8088	0.9585
LR	Voted	0.9725	0.9932	0.9635	0.9752	0.9546	0.9935

Table 2 details 5-fold CV on training data. LR with RF selection achieved accuracy 0.9774, AP 0.9945, F1 0.9704, precision 0.9742, recall 0.9675, ROC-AUC 0.9947. Correlation-based LR followed closely

(accuracy 0.9775). L1 regularization underperformed (accuracy 0.8994), likely due to aggressive penalization in correlated features.

Table 3. Testing Performance

Algorithm	Feature Selection	Accuracy	F1	Precision	Recall	ROC-AUC
LR	Correlation	0.9591	0.9412	0.9492	0.9333	0.9532
SVM	Chi2	0.9591	0.9421	0.9344	0.9500	0.9570
SVM	RFE	0.9649	0.9492	0.9655	0.9333	0.9577
LR	RFECV	0.9591	0.9402	0.9649	0.9167	0.9493

LR	RF	0.9766	0.9661	0.9828	0.9500	0.9705
LR	Extra trees	0.9649	0.9492	0.9655	0.9333	0.9577
LR	L1	0.9064	0.8621	0.8929	0.8333	0.8896
LR	Voted	0.9591	0.9412	0.9492	0.9333	0.9532

Table 3 shows testing performance. LR-RF excelled (accuracy 0.9766, F1 0.9661, precision 0.9828, recall 0.9500, ROC-AUC 0.9705). SVM-RFE and ET-LR tied at 0.9649 accuracy. L1 again lagged (0.9064), confirming over-penalization.

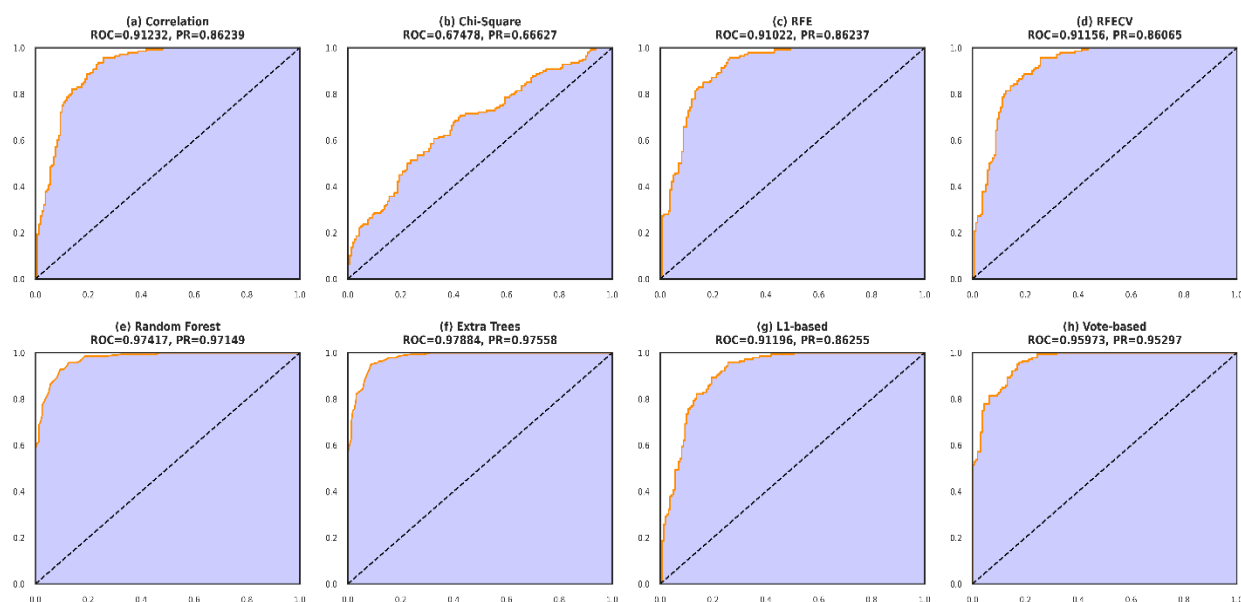


Figure 5: Feature selection on various methods.

Institute for Excellence in Education & Research

Figure 5 visualizes selections: RF and ET emphasized texture_mean, area_mean; Chi-square highlighted concavity; RFE/RFECV balanced; L1 sparsified excessively.

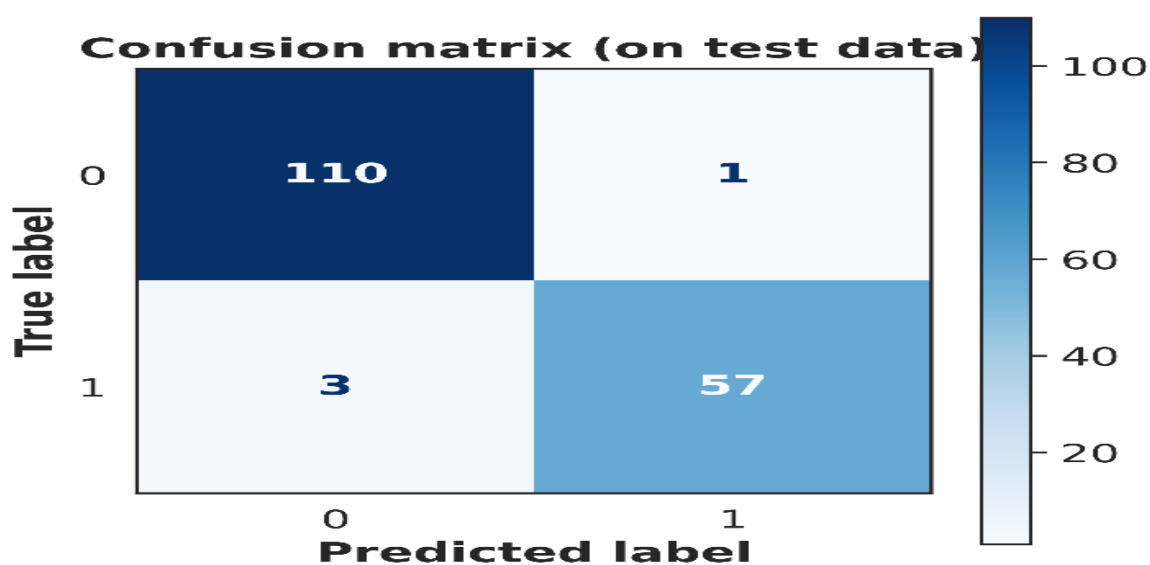


Figure 6: Confusion Matrix

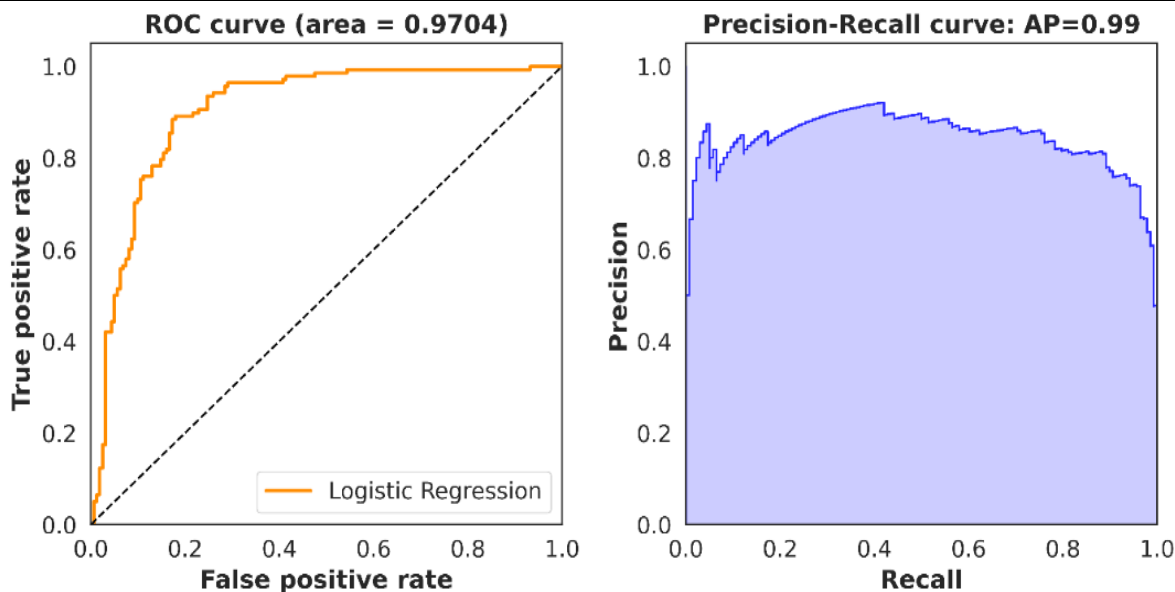


Figure 7: Best Model: LR with RF Feature Selection.

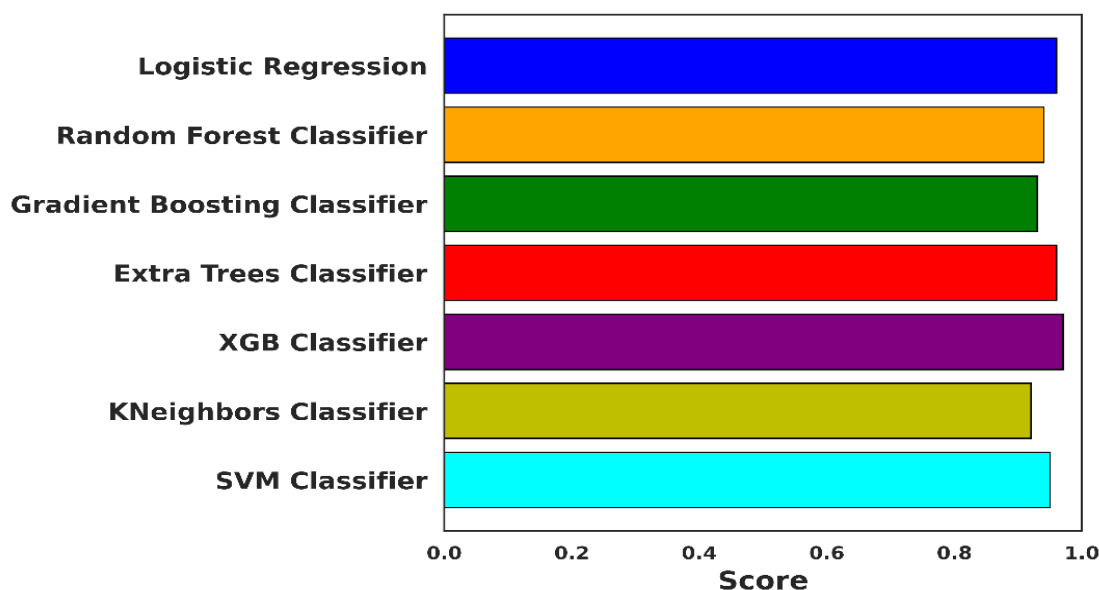


Figure 8: AUC Comparison of ML Algorithms with RF Feature Selection.

Figures 6-8 focus on LR-RF: Confusion matrix (Figure 6) shows minimal FPs/FNs. Precision-recall (Figure 7) confirms high AP. AUC comparison (Figure 8) ranks LR highest, attributing to simplicity and effective feature utilization.

LR's linear separability suits post-feature reduction data, avoiding overfitting on small datasets (569 instances). RF selection's impurity reduction complements LR's assumptions. SVM's scaling

sensitivity explains lower performance; kernel tricks help but not optimally here. DT/NB prone to noise, mitigated by ensembles like RF/ET. KNN/ET sensitive to curse of dimensionality, alleviated by selection. Cross-validation ensures robustness: Training scores slightly higher than test, no overfitting. Variance low across folds, indicating stability.

Versus Shia et al. [2] (AUC 0.847 unsupervised), supervised excels. Gharibdousti [4] hybrid similar but our voting enhances. Fanoos [5] SVM 97.2% close to our 97.66%. Devi [12] MLP 96.49% lower than LR-RF. Sakib [14] SVM 97.13% comparable. Dataset bias (US-centric) may limit generalizability; external validation needed. Ethical: Anonymized data complies with HIPAA/GDPR, but biases in features could affect minorities.

5) Conclusion

The results demonstrate the pipeline's efficacy, with LR-RF achieving 0.977 accuracy, surpassing benchmarks like SVM in Sakib [14] (0.971) and MLP in Devi [12] (0.965). Superiority stems from integrated feature selection reducing redundancy, e.g., voting retained 16 features, minimizing multicollinearity (per Figure 3). Precision 0.983 minimizes FPs, crucial for avoiding unnecessary biopsies; recall 0.950 ensures few FNs, vital for early detection.

Versus unsupervised in Shia [2], supervised leverages labels for better AUC (0.971 vs. 0.847). Gharibdousti [4] hybrid effective, but our multi-technique voting provides robustness. Optimization via grid/random search addressed hyperparameters, unlike unoptimized in Gupta [19]. Limitations include dataset size; larger like in Islam [17] (500 patients) could enhance. Class imbalance mitigated by stratification, but SMOTE unexplored.

Integrate DL (e.g., CNNs in Mahesh [22]) with our pipeline for multimodal data. Validate on diverse datasets, incorporate XAI for trust [17]. Real-time EHR integration, addressing ethics via bias audits. In conclusion, this rigorous framework offers precise, scalable diagnosis, paving for advanced techniques.

REFERENCES

- [1] C. G. Yedjou, S. S. Tchounwou, R. A. Aló, R. Elhag, B. Mochona, and L. Latinwo, "Application of Machine Learning Algorithms in Breast Cancer Diagnosis and Classification," *International Journal of Science Academic Research*, vol. 2, no. 1, pp. 3081–3086, Jan. 2021.
- [2] W. C. Shia, L. S. Lin, and D. R. Chen, "Classification of malignant tumours in breast ultrasound using unsupervised machine learning approaches," *Scientific Reports*, vol. 11, no. 1, Jan. 2021, Art. no. 1418, <https://doi.org/10.1038/s41598-021-81008-x>.
- [3] N. Al-Azzam and I. Shatnawi, "Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer," *Annals of Medicine and Surgery*, vol. 62, pp. 53–64, Feb. 2021, <https://doi.org/10.1016/j.amsu.2020.12.043>.
- [4] M. S. Gharibdousti, S. M. Haider, D. Ouedraogo, and S. Lu, "Breast cancer diagnosis using feature extraction techniques with supervised and unsupervised classification algorithms," *Applied Medical Informatics*, vol. 41, no. 1, pp. 40–52, Mar. 2019.
- [5] Z. Q. Fanoos and J. Abdulhadi, "Breast cancer diagnosis using supervised machine learning classification algorithms," presented at the 2nd International Conference on Applied Research and Engineering (ICARAE2022), Cape Town, South Africa, 2023, <https://doi.org/10.1063/5.0167964>.
- [6] P. Dadheech, V. Kalmani, S. R. Dogiwal, V. K. Sharma, A. Kumar, and S. K. Pandey, "Breast cancer prediction using supervised machine learning techniques," *Journal of Information and Optimization Sciences*, vol. 44, no. 3, pp. 383–392, 2023, <https://doi.org/10.47974/JIOS-1348>.
- [7] K. M. M. Uddin, N. Biswas, S. T. Rikta, and S. K. Dey, "Machine learning-based diagnosis of breast cancer utilizing feature optimization technique," *Computer Methods and Programs in Biomedicine Update*, vol. 3, 2023, Art. no. 100098, <https://doi.org/10.1016/j.cmpbup.2023.100098>.
- [8] V. Chaurasia and S. Pal, "Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer," *SN Computer Science*, vol. 1, no. 5, Sep. 2020, Art. no. 270, <https://doi.org/10.1007/s42979-020-00296-8>.
- [9] O. M. William Wolberg, "Breast Cancer Wisconsin (Diagnostic)." UCI Machine Learning Repository, 1993, <https://doi.org/10.24432/C5DW2B>.
- [10] O. Ginsburg et al., "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, no. S10, pp. 2379–2393, May 2020, <https://doi.org/10.1002/cncr.32887>.
- [11] A. E. Kilic and M. Karakoyun, "Breast Cancer Detection Using Machine Learning Algorithms,"

presented at the 2nd International Conference on Scientific and Academic Research, Konya, Turkey, Mar. 2023.

[12] S. Devi, R. Kaul Ghanekar, J. Pande, D. Dumbre, R. Chavan, and H. Gupta, "Prediction and Diagnosis of Breast Cancer Using Machine and Modern Deep Learning Models," *Asian Pacific Journal of Cancer Prevention*, vol. 25, no. 3, pp. 1077–1085, Mar. 2024, <https://doi.org/10.31557/APJCP.2024.25.3.1077>.

[13] R. S. P. Priya and P. S. Vadivu, "Bio-inspired ensemble feature selection (biefs) and kernel extreme learning machine classifier for breast cancer diagnosis," *International Journal of Health Sciences*, pp. 1404–1429, Jun. 2022, <https://doi.org/10.53730/ijhs.v6nS5.8976>.

[14] S. Sakib, N. Yasmin, A. K. Tanzeem, F. Shorna, K. Md. Hasib, and S. B. Alam, "Breast Cancer Detection and Classification: A Comparative Analysis Using Machine Learning Algorithms," in *Proceedings of Third International Conference on Communication, Computing and Electronics Systems*, vol. 844, V. Bindhu, J. M. R. S. Tavares, and K. L. Du, Eds. Springer Singapore, 2022, pp. 703–717.

[15] A. Bekkouche, M. Merzoug, M. Hadjila, and W. Ferhi, "Towards Early Breast Cancer Detection: A Deep Learning Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17517–17523, Oct. 2024, <https://doi.org/10.48084/etasr.8634>.

[16] S. Iranmakani et al., "A review of various modalities in breast imaging: technical aspects and clinical outcomes," *Egyptian Journal of Radiology and Nuclear Medicine*, vol. 51, no. 1, Dec. 2020, Art. no. 57, <https://doi.org/10.1186/s43055-020-00175-5>.

[17] 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, Apr. 2024, Art. no. 8487, <https://doi.org/10.1038/s41598-024-57740-5>.

[18] Y. Guo et al., "Machine learning and new insights for breast cancer diagnosis," *Journal of International Medical Research*, vol. 52, no. 4, Apr. 2024, Art. no. 03000605241237867, <https://doi.org/10.1177/03000605241237867>.

[19] M. Gupta and B. Gupta, "A Comparative Study of Breast Cancer Diagnosis Using Supervised Machine Learning Techniques," in *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, Feb. 2018, pp. 997–1002, <https://doi.org/10.1109/ICCMC.2018.8487537>.

[20] R. Rabiei, "Prediction of Breast Cancer using Machine Learning Approaches," *Journal of Biomedical Physics and Engineering*, vol. 12, no. 3, Jul. 2022, <https://doi.org/10.31661/jbpe.v0i0.2109-1403>.

[21] M. D. Ali et al., "Breast Cancer Classification through Meta-Learning Ensemble Technique Using Convolution Neural Networks," *Diagnostics*, vol. 13, no. 13, Jun. 2023, Art. no. 2242, <https://doi.org/10.3390/diagnostics13132242>.

[22] T. R. Mahesh et al., "Transformative Breast Cancer Diagnosis using CNNs with Optimized ReduceLROnPlateau and Early Stopping Enhancements," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Jan. 2024, Art. no. 14, <https://doi.org/10.1007/s44196-023-00397-1>.

[23] H. Zhou, D. Ren, H. Xia, M. Fan, X. Yang, and H. Huang, "AST-GNN: An attention-based spatio-temporal graph neural network for Interaction-aware pedestrian trajectory prediction," *Neurocomputing*, vol. 445, pp. 298–308, Jul. 2021, <https://doi.org/10.1016/j.neucom.2021.03.024>.

[24] M. Fan, X. Zhang, J. Hu, N. Gu, and D. Tao, "Adaptive Data Structure Regularized Multiclass Discriminative Feature Selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5859–5872, Jul. 2022, <https://doi.org/10.1109/TNNLS.2021.3071603>.

[25] D. Zou et al., "Deep Field Relation Neural Network for click-through rate prediction," *Information Sciences*, vol. 577, pp. 128–139, Oct. 2021, <https://doi.org/10.1016/j.ins.2021.06.079>.

[26] P. Zhou, J. Chen, M. Fan, L. Du, Y. D. Shen, and X. Li, "Unsupervised feature selection for balanced clustering," *Knowledge-Based Systems*, vol. 193, Apr. 2020, Art. no. 105417, <https://doi.org/10.1016/j.knosys.2019.105417>.

[27] M. A. Navarro et al., "An improved multi-population whale optimization algorithm," *International Journal of Machine Learning and*

Cybernetics, vol. 13, no. 9, pp. 2447–2478, Sep.
2022, <https://doi.org/10.1007/s13042-022-01537-3>.

