IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 10 **ISSUE : 12** **Print / Issue Publication Date: April 2026**

ISSN : 2455-2143

DOI : 10.33564/IJEAST.2026.v10i12.021

Indexed In

Google Scholar OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2026
Vol. 10, Issue 12, ISSN No. 2455-2143, Pages 144-153
Published Online April 2026 in IJEAST (http://www.ijeast.com)

A COMPREHENSIVE COMPARATIVE ANALYSIS OF MULTIPLE MACHINE LEARNING CLASSIFIERS FOR BREAST CANCER CLASSIFICATION USING CLINICAL AND DIAGNOSTIC FEATURES

Amritpal Singh Yadav
Research Scholar,
Department of Computer Science Engineering,
Bhagwant University, Ajmer, Rajasthan, India
Assistant Professor,
Department of Computer Engineering,
Mahila Engineering College,
Ajmer, Rajasthan, India

Virendra Kumar Sharma
Professor,
Department of Electrical and Electronics Engineering,
Bhagwant University, Ajmer, Rajasthan, India

Abstract— **One of the main causes of death for women globally is breast cancer, so early and precise detection is critical. By advancement of Machine Learning (ML) practices, automated diagnostic systems have shown promising results in assisting clinical decision-making. In this paper, a thorough relative investigation of seven supervised ML algorithms like Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), Gradient Boosting (GB), and XGBoost for breast cancer detection. In order to assess the models by training accuracy, testing accuracy, confusion matrices, precision, recall, F1-score, and ROC–AUC curves. Selection of feature is performed to address the curse of dimensionality by reducing the original 32 features to 22 highly informative attributes using correlation analysis. Experimental results demonstrate that although ensemble models achieve perfect training accuracy, Support Vector Machine provides the best balance between high testing accuracy (98.24%), minimal misclassification, and strong generalization. Hence, the ultimate model for predicting breast cancer is chosen to be SVM.**

Keywords— **Breast Cancer Detection, SVM, Machine Learning, Feature Selection, Curse of Dimensionality, Classification**

I. INTRODUCTION

One of the most common and deadly illnesses impacting women worldwide is breast cancer (BC). As per study from international cancer research organizations, every year, a sizable percentage of cancer-related deceases are caused by breast cancer. Many people globally lose their lives to, which is one of the terribly deadly as well as familiar malignancies in females [1][16][17]. Improving survival rates requires early discovery since prompt diagnosis enables efficient treatment planning and implementation. However, conventional techniques for diagnosis, such mammography, biopsy, as well as histopathological examination are frequently time-consuming, subjective, dependent on professional interpretation.

With the fast evolution of health check records and computational power, machine learning has appeared as a powerful method for developing intelligent diagnostic systems. ML models are efficient of learning complex patterns from medical datasets and can assist healthcare professionals by providing fast, objective, and accurate predictions. In breast cancer diagnosis, ML techniques are commonly applied to catalogue tumors as benign or malignant built on numerical characteristics obtained since medical images or clinical reports. Gayathri et al. presented a broad survey of ML algorithms used in breast cancer diagnosis, highlighting their strengths and limitations [2].

Despite the availability of numerous ML algorithms, choosing appropriate model for breast cancer detection remains a challenge. Different algorithms exhibit varying performance depending on data distribution, feature dimensionality, noise, and class imbalance. Moreover, medical datasets frequently restrain a enormous figure of correlated features, that can degrade model performance due to the curse of dimensionality. Therefore, it is essential to conduct a systematic comparison of multiple algorithms using proper preprocessing and feature selection techniques. Khalid et al. explored the application of ML practices for both detection and prevention of BC, emphasizing early-stage diagnosis[3].

This research aims to perform a relative investigation of seven ML algorithms for breast cancer detection. The study emphasizes feature selection, model evaluation, and performance comparison to identify the most reliable classifier. The SVM model is further analyzed as final predictive model due to its superior performance on the given dataset.

This paper is arranged like Section II gives a detailed review of existing literature associated to ML approaches for breast cancer diagnosis. Section III expresses dataset applied in this study along with its key characteristics. Section IV discusses exploratory data analysis, including missing value inspection and feature distribution analysis. Section V explains the feature selection strategy adopted to mitigate the curse of dimensionality. Section VI outlines the proposed methodology, including data preprocessing, model implementation, and evaluation procedures. Section VII presents experimental results and comparative analysis of all classifiers. Section VIII discusses the ROC–AUC analysis, while Section IX provides a comparative performance evaluation. Section X justifies the selection of the final model. Section XI concludes paper, and Section XII highlights directions for future research. data.

II. LITERATURE REVIEW

Over the past decade, numerous learning has investigated application of ML techniques for BC diagnosis. Soon research mainly driven on statistical and rule-based methods, but with advancements in data-driven approaches, ML algorithms have gained significant attention. Sakib et al. shown a relative exploration of multiple machine learning classifiers to evaluate their effectiveness in breast cancer detection and classification [4].

Logistic Regression has been widely used like baseline classifier for its simplicity as well as interpretability. Numerous considerations narrated reasonable accuracy using LR, However, its capacity to identify intricate links in high-dimensional medical data is limited by its linear decision boundary. Kaur and Madaan provided a systematic review of machine learning and statistical methods for predicting breast cancer recurrence as well as metastasis [5].

k-Nearest Neighbors (KNN) algorithm has also been applied for breast cancer classification. KNN is a non-parametric method that classifies samples based on similarity measures. While KNN can achieve good accuracy, its performance is highly sensitive to select of k and distance metric. Additionally, KNN suffers from high computational complexity and degraded performance in high-dimensional spaces. Joshi and Patel discussed the role of artificial intelligence technologies in enhancing breast cancer diagnosis and improving clinical decision-making systems [6][18].

Support Vector Machines have been extensively studied for medical diagnosis tasks. SVMs are especially valuable in handling high dimensional datasets and can construct optimal decision boundaries using kernel functions. Several researchers have demonstrated that SVM outperforms other classifiers in breast cancer detection, especially when combined with proper feature selection and normalization techniques. Jain et al. performed a comparative study of ensemble and multi-classifier models, demonstrating improved classification performance in breast cancer prediction [7].

Decision Tree-based models offer interpretability and ease of visualization. However, single DT are likely to overfitting. To address this problem, collaborative approaches such as RF and Gradient Boosting have been proposed. RF relates several DT to enhance generalization, while Gradient Boosting builds models sequentially to minimize prediction errors. Recent studies have shown that ensemble models often achieve higher accuracy than individual classifiers but require careful tuning to avoid overfitting.

XGBoost, an optimized gradient boosting procedure, has obtained popularity due to its scalability as well as superior execution in many classification problems. Researchers have reported high accuracy using XGBoost for breast cancer prediction, although its complexity and sensitivity to hyperparameters can pose challenges.

While existing studies demonstrate the effectiveness of various ML models, many lack a unified experimental framework that includes feature selection, curse of dimensionality analysis, and comprehensive evaluation metrics. This study addresses these gaps by systematically comparing seven ML algorithms using consistent preprocessing, feature reduction, and evaluation strategies.

III. DATASET DESCRIPTION

The present study employs the Breast Cancer Wisconsin (Diagnostic) dataset for analysis, which is openly offered on Kaggle platform. This collection of diagnostic records is regularly exploited used in ML research for binary classification tasks involving breast tumor detection. It comprises 569 patient scenarios, each represented by 32 attributes derived from fine needle aspiration (FNA) images of breast masses, containing an identification number, a diagnostic label, with 30 quantitative features that characterize the morphology of cell nuclei. Amethiya et al. analyzed breast cancer detection methods by integrating machine learning techniques with biosensor-based approaches for improved diagnostic accuracy [8].

For each tumor sample, ten fundamental cytological measurements are computed from digitized FNA images, capturing key geometric and textural properties of the cell nuclei. These measurements include radius (mean distance from the centre of the cell to points on the perimeter), texture (standard deviation of Gray-level intensities), perimeter, area, smoothness (local variation in radius length), compactness (related to shape regularity), concavity (severity of concave portions), concave points (number of concave features), symmetry, and fractal dimension (a numerical index of shape complexity). Das et al. conducted a comprehensive statistical evaluation of various machine learning classifiers to assess their performance in breast cancer diagnosis [9][19]. Each of these ten features is summarized in three statistical forms like mean, standard error, and the “worst” value (defined as the mean of three largest observations) resulting in a 30-dimensional numerical feature space for each sample. The diagnostic label indicates whether the tumor is malignant (cancerous) or benign (non-cancerous), encoded as 1 and 0, respectively, for supervised learning.

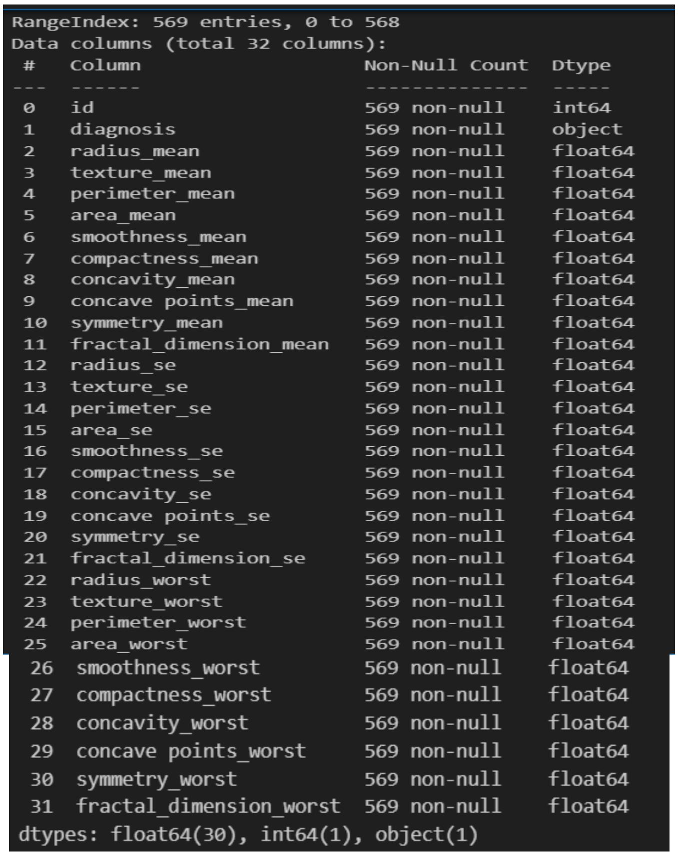
The dataset is complete by no missing values, and class distribution consists of a larger proportion of benign cases relative to malignant ones. This inherent class imbalance is accounted for during model evaluation using precision, recall, and ROC-AUC metrics in addition to accuracy.

In this study, the original 30 quantitative features were further processed using correlation-based feature selection to mitigate the curse of dimensionality and enhance model performance. Highly correlated attributes were identified and removed, yielding a reduced set of 22 features that retain significant discriminatory power for tumor classification. Rajpoot et al. emphasized the importance of feature selection techniques in enhancing the predictive performance of machine learning models for breast cancer detection [10][20]

IV. EXPLORATORY DATA ANALYSIS (EDA)

A. Missing Value Analysis

The dataset was assessed for mislaid values with missing no visualization library. The bar plot confirms that the dataset contains no missing values, ensuring data completeness and reliability.



```
RangeIndex: 569 entries, 0 to 568
Data columns (total 32 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   id                       569 non-null    int64  
 1   diagnosis                569 non-null    object 
 2   radius_mean              569 non-null    float64
 3   texture_mean             569 non-null    float64
 4   perimeter_mean           569 non-null    float64
 5   area_mean                569 non-null    float64
 6   smoothness_mean          569 non-null    float64
 7   compactness_mean         569 non-null    float64
 8   concavity_mean           569 non-null    float64
 9   concave points_mean      569 non-null    float64
 10  symmetry_mean            569 non-null    float64
 11  fractal_dimension_mean   569 non-null    float64
 12  radius_se                569 non-null    float64
 13  texture_se               569 non-null    float64
 14  perimeter_se             569 non-null    float64
 15  area_se                  569 non-null    float64
 16  smoothness_se            569 non-null    float64
 17  compactness_se           569 non-null    float64
 18  concavity_se             569 non-null    float64
 19  concave points_se        569 non-null    float64
 20  symmetry_se              569 non-null    float64
 21  fractal_dimension_se     569 non-null    float64
 22  radius_worst             569 non-null    float64
 23  texture_worst            569 non-null    float64
 24  perimeter_worst          569 non-null    float64
 25  area_worst               569 non-null    float64
 26  smoothness_worst         569 non-null    float64
 27  compactness_worst        569 non-null    float64
 28  concavity_worst          569 non-null    float64
 29  concave points_worst     569 non-null    float64
 30  symmetry_worst           569 non-null    float64
 31  fractal_dimension_worst  569 non-null    float64
dtypes: float64(30), int64(1), object(1)
```

Fig. 1. Overview of the feature composition and data structure of breast cancer wisconsin diagnostic dataset.

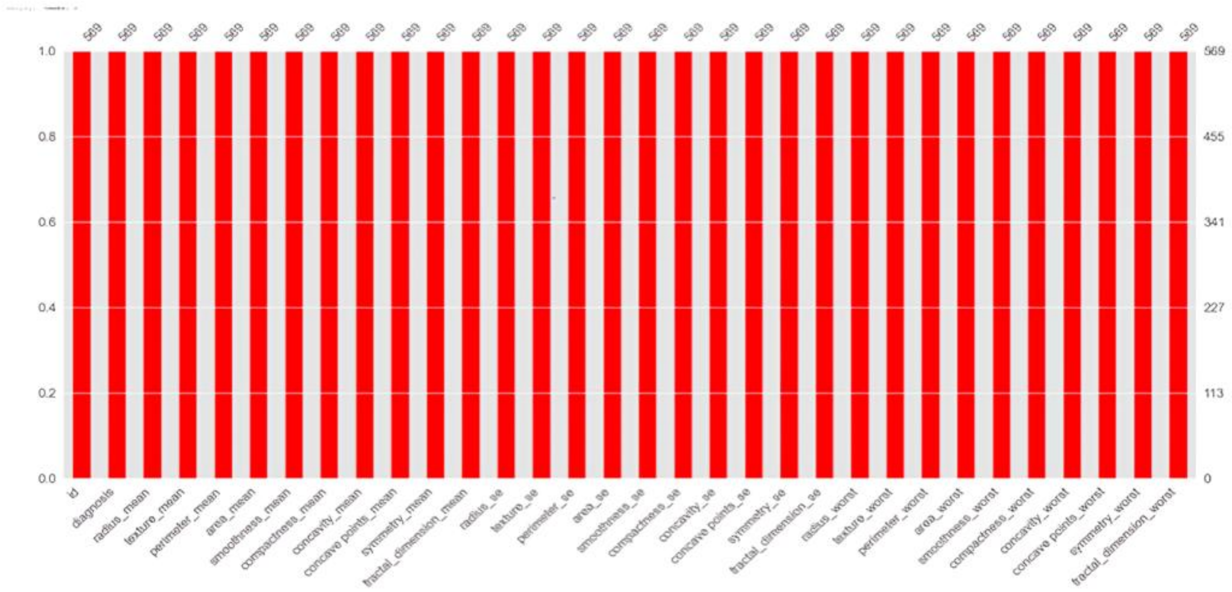


Fig. 2. Missing value bar plot

B. Target Variable Distribution

A histogram was used to examine the distribution of the target variable and evaluate the balance between benign and malignant cases. The visualization reveals that benign samples slightly outnumber malignant samples in the dataset. Although this imbalance is not severe, it can still influence model learning by biasing predictions toward majority category. To address this concern, performance evaluation was not limited to accuracy alone; additional metrics like precision, recall, F1-score, and ROC–AUC were employed to ensure a reliable and unbiased assessment of model performance.

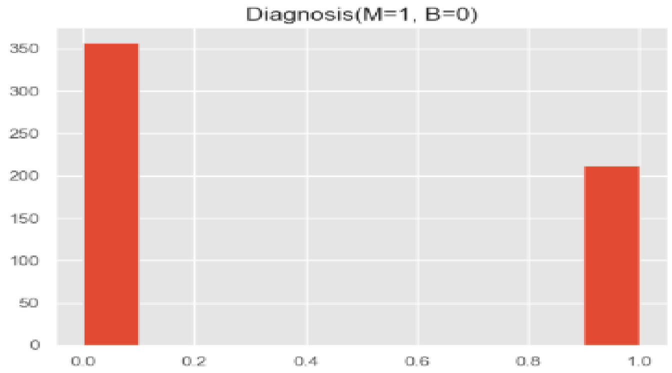


Fig. 3. Distribution of diagnosis classes.

C. Feature Distribution Analysis

Density plots were generated for all numerical features to analyze their distribution patterns, skewness, and variability. The plots show that many features deviate from a normal distribution and exhibit noticeable skewness, particularly in area- and perimeter-related attributes. Variations in spread across features indicate differences in scale and magnitude. Such non-uniform distributions can negatively affect model learning. Therefore, feature scaling is necessary to ensure balanced and unbiased model performance.

V. FEATURE SELECTION AND CURSE OF DIMENSIONALITY

High-dimensional datasets often face the curse of dimensionality, a phenomenon where having too many features can lead to sparse data representation, increased computational complexity, and degraded model generalization. To mitigate these challenges, a correlation-based feature selection approach was employed, which focuses on identifying and retaining the most informative features while eliminating redundancy.

A Pearson correlation matrix was calculated to quantify the linear relationships between every pair of features, and those with correlation coefficients greater than 0.92 were deemed highly correlated and redundant. By applying an upper triangular correlation mask, only one feature from each correlated pair was kept, ensuring that duplicate information did not inflate the model. As a result, the feature set was reduced from 32 to 22 key attributes and these are shown in Fig.6. .

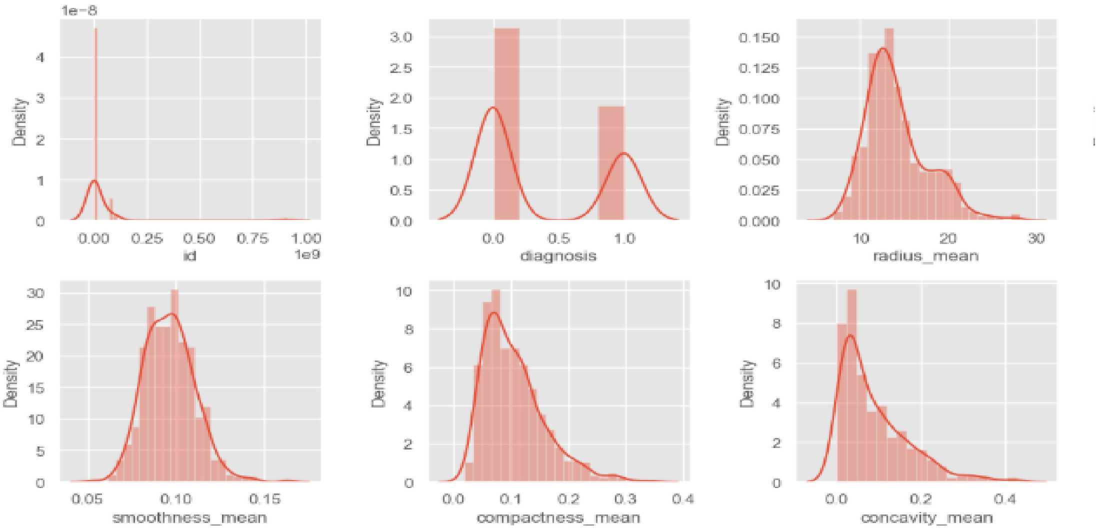


Fig. 4. Density curves of some features.

This reduction in dimensionality not only improves computational efficiency but also minimizes overfitting, stabilizes model training, and improves overall predictive accuracy. Additionally, by focusing on the most significant characteristics, interpretability of model is improved, allowing for clearer perceptions into which variables drive predictions.

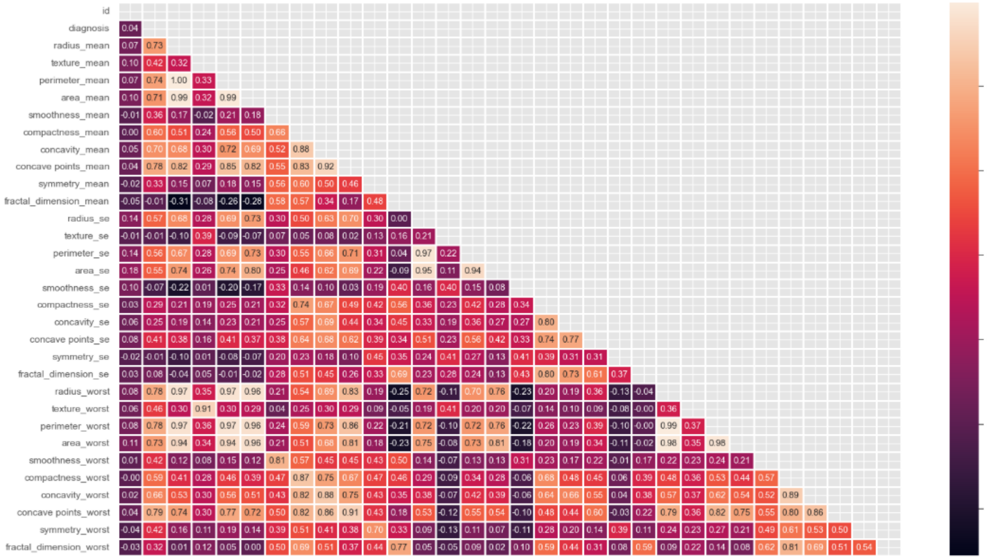


Fig. 5. Correlation heatmap.

VI. METHODOLOGY

A. Data Splitting and Scaling

The dataset has been divided into training (80%) and testing (20%) sets. Feature scaling has been achieved using StandardScaler, ensuring assuming that all features contribute equally to distance-based measures and margin-based classifiers. Priyadharsini and Sathiaseelan compared machine learning and deep learning models, highlighting their effectiveness in breast cancer classification tasks [11][21][22].

B. Machine Learning Models Implemented

The dataset was fragmented into training (80%) and testing (20%) sets. Feature scaling was performed using StandardScaler, ensuring that all features contribute equally to distance-based and margin-based classifiers.

The following seven classifiers were implemented:

1. Logistic Regression
2. K-Nearest Neighbors
3. Support Vector Machine
4. Decision Tree Classifier
5. Random Forest Classifier
6. Gradient Boosting Classifier
7. XGBoost Classifier

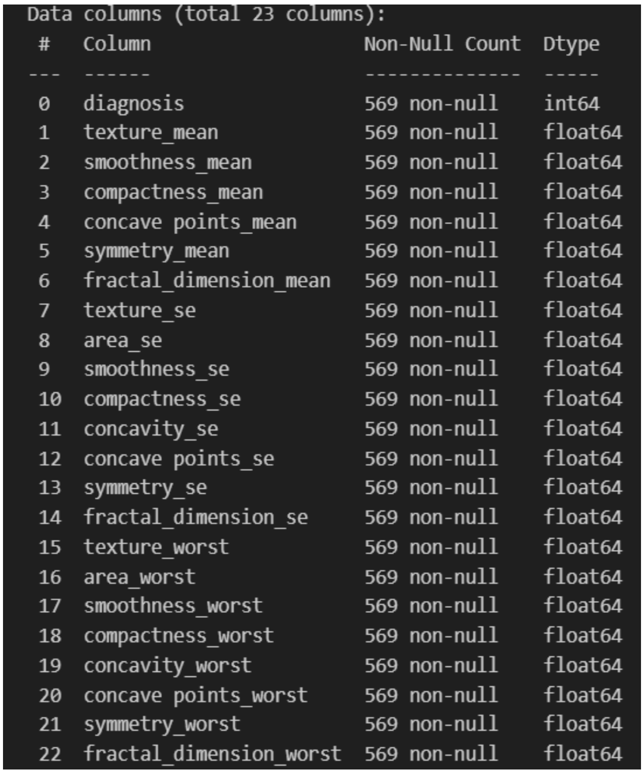
Each model was trained using the same training facts and evaluated spending identical metrics for fair comparison. Preetam et al. highlighted recent advancements in AI-powered imaging technologies for improving breast cancer diagnosis and treatment planning [12][23].

VII. EXPERIMENTAL RESULTS AND ANALYSIS

A. Logistic Regression

The Training Accuracy: 98.90%
Testing Accuracy: 96.49%

The confusion matrix shows only 4 misclassifications, indicating stable performance with minimal overfitting. Logistic Regression demonstrates strong baseline performance but a slight decrease in recall for malignant instances. Khan et al. reviewed recent advances in segmentation and classification techniques for accurate and efficient breast cancer detection [13][24][25].



```
Data columns (total 23 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   diagnosis                569 non-null    int64
 1   texture_mean             569 non-null    float64
 2   smoothness_mean          569 non-null    float64
 3   compactness_mean         569 non-null    float64
 4   concave points_mean      569 non-null    float64
 5   symmetry_mean            569 non-null    float64
 6   fractal_dimension_mean   569 non-null    float64
 7   texture_se               569 non-null    float64
 8   area_se                  569 non-null    float64
 9   smoothness_se            569 non-null    float64
 10  compactness_se           569 non-null    float64
 11  concavity_se             569 non-null    float64
 12  concave points_se        569 non-null    float64
 13  symmetry_se              569 non-null    float64
 14  fractal_dimension_se     569 non-null    float64
 15  texture_worst            569 non-null    float64
 16  area_worst               569 non-null    float64
 17  smoothness_worst         569 non-null    float64
 18  compactness_worst        569 non-null    float64
 19  concavity_worst          569 non-null    float64
 20  concave points_worst     569 non-null    float64
 21  symmetry_worst           569 non-null    float64
 22  fractal_dimension_worst  569 non-null    float64
```

Fig. 6. Selected 22 features after correlation-based reduction

B. K-Nearest Neighbors

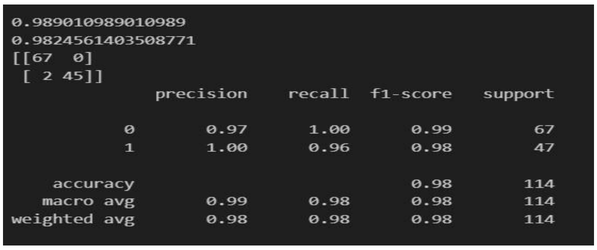
- The Training Accuracy: 96.70%
- Testing Accuracy: 95.61%

KNN shows reasonable accuracy but lower recall for malignant cases, reflecting sensitivity to feature distribution and neighborhood selection.

C. Support Vector Machine (Proposed Model)

- Training Accuracy: 98.90%
- Testing Accuracy: 98.24%

The SVM model achieves the highest testing accuracy, correctly classifying most samples with only 2 misclassifications, indicating superior predictive performance. The confusion matrix shows perfect recall (1.00) for benign cases, ensuring that all benign instances are accurately classified. For malignant cases, the model attains near-perfect precision (1.00) along by a high recall of 0.96, reflecting its strong ability to detect malignant instances while minimizing false positives. Overall, results emphasize the robustness as well as consistency of SVM in handling classification tasks with high accuracy and balanced performance across classes. Bandi et al. proposed methods to enhance the efficiency as well as accuracy of machine learning algorithms for breast cancer prediction and classification [14][26][27].



```
0.989010989010989
0.9824561403508771
[[67  0]
 [ 2 45]]
              precision    recall  f1-score   support

           0       0.97      1.00      0.99        67
           1       1.00      0.96      0.98        47

    accuracy                           0.98       114
   macro avg       0.99      0.98      0.98       114
weighted avg       0.98      0.98      0.98       114
```

Fig. 7. Training accuracy, testing accuracy, confusion matrix as well as classification describe for support vector machine.

D. Decision Tree Classifier

- Training Accuracy: 97.36%
- Testing Accuracy: 92.10%

Decision Tree suffers from overfitting, as evidenced by reduced testing accuracy and higher misclassification count.

E. Random Forest Classifier

- Training Accuracy: 99%
- Testing Accuracy: 98.10%

Random Forest shows excellent performance; however, perfect training accuracy indicates memorization, increasing the risk of overfitting compared to SVM.

F. Gradient Boosting Classifier

- Training Accuracy: 99%
- Testing Accuracy: 97.36%

Gradient Boosting performs well but exhibits a noticeable gap amongst training and testing accuracy, representing moderate overfitting.

G. XGBoostClassifier

- Training Accuracy: 99.10%
- Testing Accuracy: 94.73%

XGBoost shows strong training performance but reduced generalization on test data, likely due to dataset size.

VIII. ROC-AUC ANALYSIS

The Receiver Operating Characteristic (ROC) curve was employed to estimate and calculate discriminative capability of the implemented ML models. ROC curve illustrates trade-off concerning True Positive Rate (Sensitivity) as well as False Positive Rate among different classification thresholds.

A diagonal baseline reflects random classification, whereas proximity of the curve to the top-left corner signifies stronger model performance.

As shown in Fig. 8. all classifiers demonstrate strong predictive ability, with curves significantly above the diagonal baseline. Among them, SVM and RF exhibit utmost prominent ROC curves, reflecting their superior ability to recognize concerning benevolent and malicious cases. Logistic Regression and XGBoost also show competitive performance, while Decision Tree and KNN display comparatively lower, yet acceptable, discrimination power.

The consistently high ROC performance confirms that the selected features and preprocessing steps effectively capture meaningful patterns in the dataset, enabling robust Breast Cancer Classification.

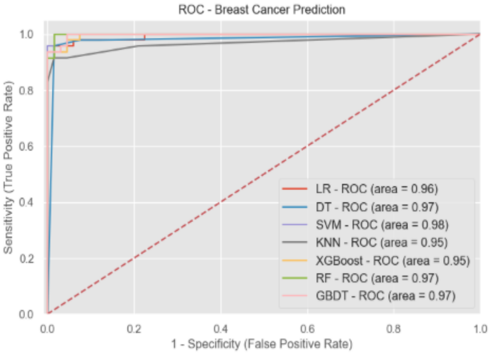


Fig. 8. ROC curves of all classifiers for breast cancer classification.

IX. COMPARATIVE PERFORMANCE EVALUATION

To provide a holistic comparison, model performance was further analyzed using a bar chart illustrating testing accuracy and ROC–AUC scores for all classifiers. This visualization enables a clear comparison of both classification correctness and class separation capability. Takele and Chen presented a comparative analysis of data-driven models for predicting breast cancer survival outcomes [15][28][29].

As depicted in Fig. 9. Support Vector Machine achieves one of the highest testing accuracies while maintaining an excellent ROC–AUC score. Although Random Forest also performs comparably, its perfect training accuracy suggests a higher tendency toward memorization. Gradient Boosting and Logistic Regression show balanced performance, whereas Decision Tree and KNN lag slightly due to lower generalization capability.

This combined evaluation highlights SVM as the most consistent and reliable model across multiple metrics.

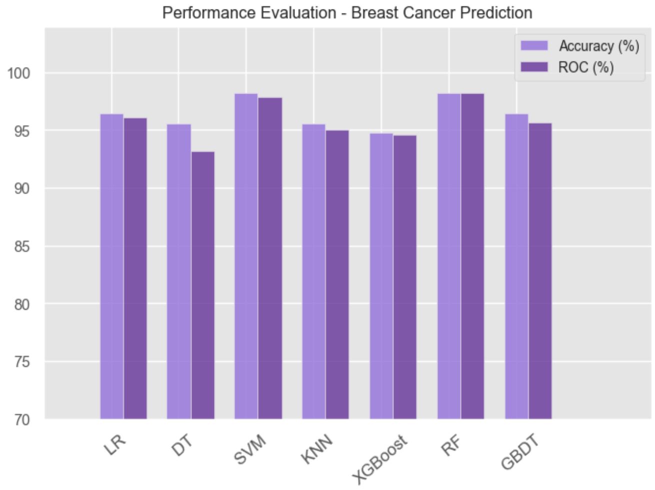


Fig. 9. Accuracy and ROC–AUC comparison of all classifiers for breast cancer classification.

TABLE- 1 COMPARATIVE PERFORMANCE METRICS

Classifier	**Training Accuracy (%)**	**Testing Accuracy (%)**
LogisticRegression	**98.90**	**96.49**
K-Nearest Neighbors	**96.70**	**95.61**
Support Vector Machine	**98.90**	**98.24**
Decision Tree Classifier	**97.36**	**92.10**
Random Forest Classifier	**99.00**	**98.10**
Gradient Boosting Classifier	**99.00**	**97.36**
XGBoost Classifier	**99.10**	**94.73**

X. MODEL SELECTION JUSTIFICATION

The choice of a suitable classification model constitutes a critical stage in development of a reliable breast cancer prediction system, as incorrect classification can lead to severe clinical consequences. In this study, multiple machine learning classifiers—including LR, DT, K-Nearest Neighbors, RF, Gradient Boosting, XGBoost, and SVM were assessed by comprehensive evaluation metrics such as accuracy as well as ROC–AUC.

Although ensemble-based models like Random Forest and Gradient Boosting achieved exceptionally high training accuracy, their performance exhibited a marginal decline on the testing dataset, indicating a tendency toward overfitting. In contrast, SVM classifier proved a strong balance between training and testing performance, reflecting superior generalization capability. This behaviour is particularly important in medical applications, where models must perform consistently on unseen patient data.

Furthermore, SVM achieved one of the highest ROC–AUC values among all evaluated models, highlighting its effectiveness in separating benign and malignant breast cancer cases across different classification thresholds. The ability to maximize the margin between classes allows SVM to address complex classification boundaries even in large-scale feature spaces commonly found in biomedical datasets. This makes SVM especially suitable for breast cancer classification, where subtle feature variations can significantly impact diagnostic outcomes.

Another important consideration is the misclassification rate. SVM exhibited a lower rate of erroneous positive and negative classifications compared to other models, that is critical in reducing preventable medical interventions and minimizing risk of missed diagnoses. Additionally, SVM offers computational efficiency when compared to complex ensemble models, rendering it a practical solution for real-time applications in resource-limited clinical scenarios.

Considering its high predictive accuracy, robust generalization performance, strong ROC–AUC results, and suitability for high-dimensional data, SVM is selected as most suitable and reliable model to breast cancer classification in this study.

XI. CONCLUSION

This study provides a thorough evaluation of various machine learning classifiers for breast cancer classification, utilizing clinical and diagnostic features. Initial exploratory data analysis revealed non-normal feature distributions and mild class imbalance, which were addressed through feature scaling and appropriate evaluation metrics. Correlation-based feature selection successfully reduced redundancy by eliminating highly correlated attributes, thereby mitigating the curse of dimensionality and improving model efficiency.

Experimental results demonstrate that the Support Vector Machine outperforms additional classifiers with respect to accuracy, generalization capability, and stability. Here, ROC–AUC analysis extend confirms SVM’s strong class separation ability. Result indicated that SVM is a consistent and well-performing model intended for breast cancer classification and has strong applicability in computer-aided clinical decision-making systems.

XII. FUTURE SCOPE

While suggested breast cancer classification framework determines strong predictive performance, several avenues strive to further improve its effectiveness as well as clinical applicability. Forthcoming extensions of this work may involve incorporation of deep learning procedures like Convolutional Neural Networks (CNNs) and hybrid neural architectures, which can automatically learn complex feature representations and potentially improve diagnostic accuracy when larger datasets are available.

The model can also be extended by encompassing added medical and genomic data, like patient record, hormone receptor status, or imaging features, to develop a more comprehensive and personalized diagnostic system. This multimodal approach could significantly improve decision-making in early-stage breast cancer detection.

Advanced feature collection and dimensionality reduction procedures, including hybrid filter–wrapper methods or evolutionary algorithms, may further optimize model performance by identifying the most discriminative attributes. Additionally, implementing cross-institutional datasets would help evaluate durability and scalability of model throughout distinct populations.

From a deployment perspective, incorporation of proposed model hooked on real-time clinical decision confirmation approaches or web-based diagnostic platforms could assist healthcare professionals in making faster and more accurate assessments. With continued refinement and validation, such systems have the potential to contribute meaningfully to early breast cancer diagnosis and improved patient outcomes.

XIII. REFERENCES

[1]. Naseem, Usman, et al. "An automatic detection of breast cancer diagnosis and prognosis based on machine learning using ensemble of classifiers." Ieee Access 10 (2022): 78242-78252.

[2]. Gayathri, B. M., C. P. Sumathi, and T. Santhanam. "Breast cancer diagnosis using machine learning algorithms-a survey." International Journal of Distributed and Parallel Systems 4.3 (2013): 105.

[3]. Khalid, Arslan, et al. "Breast cancer detection and prevention using machine learning." Diagnostics 13.19 (2023): 3113.

[4]. Sakib, Shadman, et al. "Breast cancer detection and classification: A comparative analysis using machine learning algorithms." Proceedings of third international conference on communication, computing and electronics systems: ICCCES 2021. Singapore: Springer Singapore, 2022.

[5]. Kaur, Charanpreet, and Rosy Madaan. "Machine Learning and Statistical Approaches for Predicting Breast Cancer Recurrence and Metastasis: A Systematic Review." Advances in Artificial Intelligence and Machine Learning 6.1 (2026): 279.

[6]. Joshi, Sachi, and Upesh Patel. "Breast Cancer Diagnosis With Ai Technologies." Journal of Multiscale Modelling (2026).

[7]. Jain, Garima & Jain, Sanat & Phulre, Ajay & Sharma, Rahul & Joshi, Suneet. (2024). A Comparative Analysis of Breast Cancer Classification using Ensemble and Multi Classifier Algorithms. 1-6. 10.1109/ICDSCNC62492.2024.10939716.

[8]. Yash Amethiya, Prince Pipariya, Shlok Patel, Manan Shah,Comparative analysis of breast cancer detection using machine learning and biosensors, Intelligent Medicine,Volume 2, Issue 2,2022,Pages 69-81,ISSN 2667-1026,https://doi.org/10.1016/j.imed.2021.08.004.

[9]. Das, S.S.; Mahaprasad, A.; Padhy, N.; Misra, S.; Panigrahi, R.; Mahapatro, P.K.; Arangi, D. Comparative Evaluation of Machine Learning Classifiers for Breast Cancer Diagnosis: A Comprehensive Statistical Analysis. Eng.Proc. 2026, 124,35. https://doi.org/10.3390/engproc2026124035.

[10]. Rajpoot, C. S., Sharma, G., Gupta, P., Dadheech, P., Yahya, U., & Aneja, N. (2024). Feature Selection-

based Machine Learning Comparative Analysis for Predicting Breast Cancer. Applied Artificial Intelligence, 38(1). https://doi.org/10.1080/08839514.2024.2340386.

[11]. Priyadharsini, M.Suriya & Sathiaseelan, J.. (2021). comparative analysis of machine learning and deep learning prediction of breast cancer classification.

[12]. Preetam, Subham, et al. "Next-gen diagnostics: artificial intelligence-powered imaging in breast cancer care." Journal of Cancer Metastasis and Treatment 11 (2025): N-A.

[13]. Khan, Muhammad Moosa Raza, et al. "Recent Advances in Breast Cancer Detection: A Review on Segmentation and Classification Techniques." ICCK Journal of Image Analysis and Processing 1.4 (2025): 147-161.

[14]. Bandi, Sriya, Jaya Prakash Vemuri, and Jyoti Kainthola. "Improving the Efficiency of Machine Learning Algorithms for Prediction and Classification of Breast Cancer." International Conference on Smart Computing and Informatics. Cham: Springer Nature Switzerland, 2025.

[15]. Takele, Kasahun, and Ding-Geng Chen. "A comparative analysis of data-driven models for breast cancer survival prediction." Scientific Reports (2026).

[16]. Ahmed, K.A., Humaira, I., Khan, A.R., Hasan, M.S., Islam, M., Roy, A., Karim, M., Uddin, M., Mohammad, A., and Xames, M.D. (2025). Advancing Breast Cancer Prediction: Comparative Analysis of Machine Learning Models and Deep Learning-Based Multi-Model Ensembles on Original and Synthetic Datasets, PLOS ONE, 20(6), e0326221.

[17]. Karamti, H., Alharthi, R., Umer, M., Shaiba, H., Ishaq, A., Abuzinadah, N., Alsubai, S., and Ashraf, I. (2024). Breast Cancer Detection Employing Stacked Ensemble Model with Convolutional Features, Computers in Biology and Medicine.

[18]. Singh, J., Alsaidan, O.A., Aodah, A., Alrobaian, M., Almalki, W.H., Almujri, S.S., Sahoo, A., Alam, K., Lal, J.A., and Barkat, M.A. (2026). Artificial Intelligence in Breast Cancer: Clinical Applications in Diagnosis, Prognosis, and Therapeutics, Expert Review of Anticancer Therapy, 22(2), pp.249–269.

[19]. Jain, P.; Aggarwal, S.; Adam, S.; Imam, M. (2024). Parametric Optimization and Comparative Study of Machine Learning and Deep Learning Algorithms for Breast Cancer Diagnosis, Breast Disease, 43, pp.257–270.

[20]. Kumar, A.; Saini, R.; Kumar, R. (2024). A Comparative Analysis of Machine Learning Algorithms for Breast Cancer Detection and Identification of Key Predictive Features, Traitement du Signal, 41(1), pp.127–140.

[21]. Yusoff, M.; Haryanto, T.; Suhartanto, H.; Mustafa, W.A.; Zain, J.M.; Kusmardi, K. (2023). Accuracy Analysis of Deep Learning Methods in Breast Cancer Classification: A Structured Review, Diagnostics, 13(4), pp.683.

[22]. Mikhailova, V.; Anbarjafari, G. (2022). Comparative Analysis of Classification Algorithms on the Breast Cancer Recurrence Using Machine Learning, Medical & Biological Engineering & Computing, 60(9), pp.2589–2600.

[23]. Al-Karawi, D.; Al-Zaidi, S.; Helael, K.A.; Obeidat, N.; Mouhsen, A.M.; Ajam, T.; Alshalabi, B.A.; Salman, M.; Ahmed, M.H. (2024). A Review of Artificial Intelligence in Breast Imaging, Tomography, 10(5), pp.705–726.

[24]. igueroa, R.L.; Zeng-Treitler, Q.; Ngo, L.H.; Goryachev, S.; Wiechmann, L. (2023). Deep Learning Models for Breast Cancer Detection in Mammography: A Review of Segmentation and Classification Techniques, Computers in Biology and Medicine, 158, pp.106870.

[25]. Khan, S.; Hussain, M.; Aboalsamh, H.; Bebis, G. (2023). A Survey on Deep Learning-Based Breast Cancer Detection Using Mammography: Segmentation, Classification, and Challenges, Artificial Intelligence in Medicine, 139, pp.102527.

[26]. El-Naqa, I.; Murphy, M.J. (2023). Machine Learning in Oncology: Methods, Applications, and Challenges in Cancer Prediction and Classification, Cancer Research, 83(6), pp.1012–1023.

[27]. Ullah, Z.; Khan, M.A.; Tariq, U.; Kadry, S.; Nam, Y. (2024). An Efficient Machine Learning Framework for Breast Cancer Classification Using Optimized Feature Selection and Hyperparameter Tuning, Computers in Biology and Medicine, 168, pp.107737.

[28]. Hamedi, S.Z.; Emami, H.; Khayamzadeh, M.; Rabiei, R.; Aria, M.; Akrami, M.; Zangouri, V. (2024). Application of Machine Learning in Breast Cancer Survival Prediction Using a Multimethod Approach, Scientific Reports, 14, pp.30147.

[29]. Noman, S.M.; Fadel, Y.M.; Henedak, M.T.; Attia, N.A.; Essam, M.; Elmaasarawii, S.; Fouad, F.A.; Eltasawi, E.G.; Al-Atabany, W. (2025). Leveraging Survival Analysis and Machine Learning for Accurate Prediction of Breast Cancer Recurrence and Metastasis, Scientific Reports, 15, pp.3728.

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE AND TECHNOLOGY

ABOUT IJEAST

International Journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high-quality research papers in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

FOCUS AREAS

- Engineering
- Applied Science
- Technology
- Innovation & Development
- Interdisciplinary Studies

PEER REVIEWED

All submissions are rigorously peer reviewed to ensure quality.

OPEN ACCESS

Free and unrestricted access to research for all.

GLOBAL REACH

Connecting researchers and professionals worldwide.

TIMELY PUBLICATION

We ensure a swift and efficient publication process.

For more information, visit our website

www.ijeast.com

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

editor@ijeast.com
www.ijeast.com
India

2455-2143