JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 10 Issue 02
2026

Research Article
https://doi.org/10.56979/1002/2026/1294

A Multi-Model Machine Learning Approach for Reliable Breast Cancer Classification

Chahira Lhioui[1*]

[1]Department of Computer Science and Artificial Intelligence, College of Computing and Information Technology, University of Bisha, Bisha, Saudi Arabia.
[*]Corresponding Author: Chahira Lhioui , Email: shahira@ub.edu.sa

Received: December 11, 2025 **Accepted:** February 19, 2026

Abstract: Breast cancer is a critical health issue that requires early diagnosis in order to enhance clinical outcomes. This research paper is a proposal of a detailed machine learning system to the classification of breast tumors based on the morphological characteristics of the fine-needle aspiration photos. The workflow combines exploratory data analysis, feature standardization, Principal Component Analysis (PCA), and the relative comparison of seven kinds of supervised classifiers, including Logistic Regression, SVM, KNN, Decision Tree, Random Forest, Gradient Boosting and XGBoost. Exploratory data mining showed that there are strong correlations between geometric features and there are evident distributional differences between benign and malignant tumors. PCA affirmed that the attributes that prevail in the structure of variances are tumor size and concavity-related attributes. The comparative results showed that the model performance was always very high and the best results were obtained by the boosting-based ensemble methods in the accuracy, ROC-AUC and Precision-Recall results. The importance of features analysis revealed consistent key morphological predictors in models. The results show that linear models have a competitiveness, but ensemble models have better robustness and reliability in the classification of breast cancer.

Keywords: Breast Cancer Classification; Machine Learning; Ensemble Learning; XGBoost; Random Forest; Principal Component Analysis; ROC-AUC; Feature Importance; Medical Diagnosis.

1. Introduction

Breast cancer is one of the most major issues of the global public health, and is continuously reported to be among the most widespread cancers to have been diagnosed in the world. Recent world reporting and institutional overviews point out that timely identification and diagnosis is the key to decreasing mortality and enhancing long-term outcomes [1], [2]. Although clinical workflows apply to imaging, pathology, and expert interpretation, the systems of healthcare increasingly produce structured clinical and diagnostic datasets that can be used to make decisions. In this regard, machine learning (ML) has now become a significant aspect of the computer-aided diagnosis (CAD) studies, with the purpose of supporting clinicians through enhancing the predictability and promptness of risk stratification and classification.

The Wisconsin Breast Cancer datasets (both diagnostic and 569 samples and 30 calculated features as the diagnostic variant) remain one of the most popular structured benchmarks in the research of breast cancer ML as they are commonly used to compare the methods, study feature engineering, and perform reproducible experiments [3], [4], [5], [6]. Recent research on these datasets with supervised learning has shown high performance with both traditional models (e.g., Logistic Regression, k-Nearest Neighbors, Support Vector Machines) and more recent ensemble models (e.g., Random Forests, Gradient Boosting, and XGBoost) and high diagnostic accuracy on standardized preprocessing [7]. The fact that this benchmark is still being used in the literature today serves as evidence of its use in assessing the behavior

of the model, sensitivity to preprocessing, and comparative results when placed in uniform experimental conditions [8].

Besides, recent achievements in the sphere of breast cancer detection have incorporated more and more deep learning, quantum-inspired optimization and hybrid modeling methods to improve the diagnostic ac-curacy. Bilal et al. [9] Suggested a hybrid model of genomic and pathological data utilization with deep neural networks to detect breast cancer in a better way. Quantum optimization schemes are also investigated, such as SqueezeNet-SVM models [10], quantum-inspired grey wolf algorithms based SVM models (BC-QNet) [11] and quantum-inspired grey wolf algorithms (BC-QNet) [12]. Additionally, infusion methods in quantum computation have shown good potentials in early detection of multi-cancer systems. [13]. These papers point to the increasing development of having better ways of integrating advanced optimization and hybrid intelligence methods to enhance robustness and predictive capability of medical diagnosis systems.

Nevertheless, there are a number of constraints that are typical of recent studies. To begin with, the majority of studies give primary attention to reporting accuracy and a single confusion matrix but under-exploit threshold-independent evaluation, even though there is increasing interest that model evaluation in clinical settings should incorporate measures and curves that indicate decision thresholds and sensitivity to imbalance between classes [14][15]. Specifically, the analysis done recently shows that ROC-AUC and PR-AUC are not always the same in the case of imbalance in class and that PR-based assessment can yield performance on the positive class in some contexts [16]. Second, feature engineering and feature selection are commonly mentioned, but an overall exploration data analysis (EDA), such as visualization of missingness, feature distribution diagnostics, as well as reporting of the correlation structure are often not complete, diminishing interpretability and restricting reproducibility [6]. Third, visualization (e.g. PCA) and dimensionality reduction are sometimes not implemented as part of an entire pipeline to relate EDA, preprocessing, model training, and evaluation [17], [18], [19]. Such gaps complicate the comparison of outcomes across studies, in addition to the comprehension of the influence of the preprocessing and correlated features on the model performance.

To mitigate these limitations, this paper provides a reproducible end-to-end ML workflow to classify breast cancer using the Wisconsin diagnostic data that combines: (i) comprehensive EDA with missing value profiling, distribution plots and correlation analysis; (ii) standardized preprocessing using imputation and feature scaling; (iii) comparative testing of seven ubiquitous classifiers (Logistic Regression, KNN, SVM, Decision Tree, Random Forest, Gradient Boosting and XGBoost); (iv) multi-metric reporting of Accuracy, Precision The proposed study will supply a transparent and easily reproducible baseline of structured breast cancer diagnosis modeling through the combination of systematic visualization, effective evaluation metrics, and a consistent pipeline, consistent with the current evaluation practices that are increasingly encouraged in the AI literature in healthcare.

2. Materials and Methods

This paper takes a systematic guided machine learning approach to categorize breast tumors to be benign or malignant based on quantitative cellular characteristics of fine needle aspirate (FNA). The experimental design was formulated in such a way that there would be methodological rigor, reproducibility, and fair comparative assessment among the several classifiers. The general pipeline involves consecutive phases, such as the exploration of data, preprocessing, stratified sampling, multi-model training, dimensionality reduction, and the overall performance measures with the help of various statistical measures and graphical visualization tools that are introduced in Figure 1.

The current research employs the Breast Cancer Wisconsin Diagnostic data, comprising 569 cases and 30 continuous-valued variables obtained as a result of the analysis of digital images of fine needle aspirate (FNA) of a mass in the breast. The features are either benign (0) or malignant (1) [20] The features are used to describe morphological properties of cell nuclei such as, radius, texture, perimeter, area, smoothness, compactness, concavity, symmetry and fractal dimension.

These characteristics comprise radius, texture, perimeter, area, smoothness, compactness, concavity, symmetry as well as fractal dimension measurements. In both cases, a binary diagnostic outcome $y \in \{0,1\}$, where 0 denotes benign tumors and 1 denotes malignant tumors. The feature matrix is denoted as $X \in \mathbb{R}^{n \times d}$, where $n = 569$and $d = 30$.

Before the model developments, an extensive exploration data analysis (EDA) was performed, to gain insights into feature distributions, their relationship with each other, and possible data quality problems. All variables were calculated in terms of statistical summaries, such as mean, standard deviation, minimum, and maximum values. Numerical and visual analysis of missing values patterns were done.

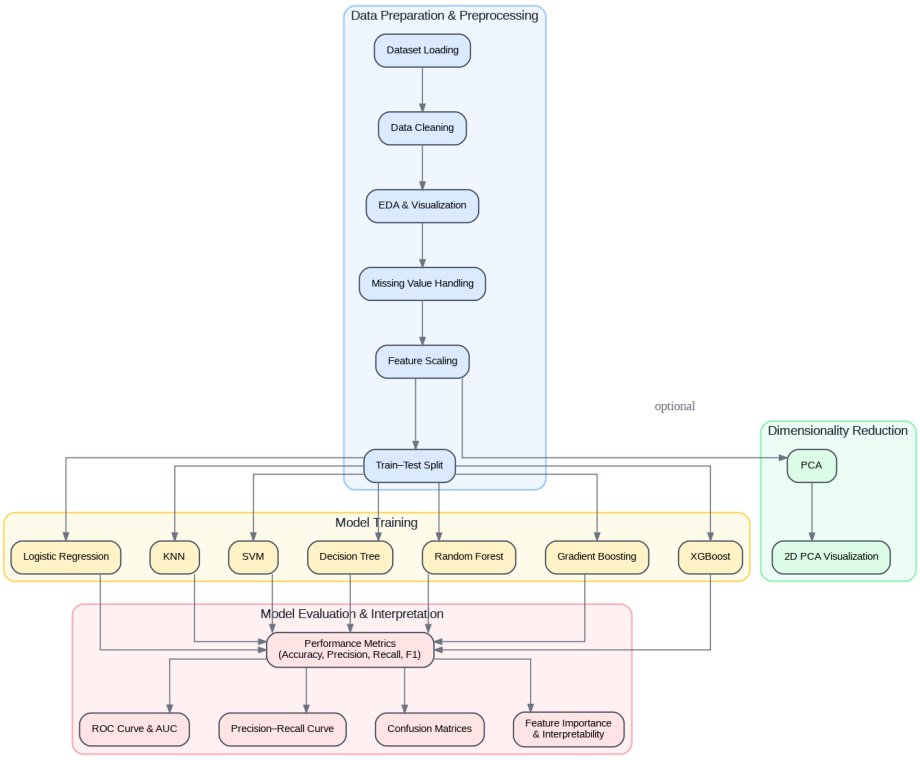

Figure 1. Methodology Flow Diagram

The number of missing values of the dataset is relatively small but the uniform imputation strategy was chosen to ensure methodological validity. The proportion of missing values per feature was given as:

$$\text{Missing Rate}_j = \frac{\text{Number of missing entries in feature } j}{n} \times 100 \quad (1)$$

Pearson correlation coefficient was used to analyze the relations between features and determine the linearity:

$$r_{ij} = \frac{\sum(x_i - \bar{x_i})(x_j - \bar{x_j})}{\sigma_i \sigma_j} \quad (2)$$

The correlation table that was obtained was graphically represented to identify multicollinearity and redundant features. Also, histograms and boxplots by diagnostic class were created to determine the separability of the classes and the possible presence of outlier behavior. Preprocessing steps were applied in a systematic manner to achieve numerical stability as well as the same training conditions across models. The imputed values were done by median imputation which is as follows:

$$x_{ij}^{*} = \begin{cases} x_{ij}, & \text{if observed} \\ \text{median}(X_j), & \text{if missing.} \end{cases} \quad (3)$$

The median imputation was chosen instead of the mean imputation because of its resistance to skewness and outliers in biomedical measurements. After the imputation, feature standardization was used since some of the models utilized in this research are distance-sensitive or scale-sensitive. The standardization was done as:

$$x'_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}, \quad (4)$$

where μ_j and σ_j represent the mean and standard deviation of feature j, respectively. Such a transformation guarantees zero mean and unit variance without overwhelming features with larger numerical values. Stratified sampling was then used to divide the dataset into training and testing subsets to maintain the ratio of malignant and benign cases in the two subsets. An 80:20 division was taken, where:

$$D = D_{train} \cup D_{test}, | D_{train} | = 0.8n, | D_{test} | = 0.2n. \quad (5)$$

The stratification will provide fair assessment and consistency in the balance of classes during training and evaluation phases. There were seven trained classification algorithms that were used under the same preprocessing conditions to facilitate a fair comparison. Logistic Regression was used as a probabilistic linear classifier which is used to model the conditional probability of malignancy by using the sigmoid function:

$$P(y = 1 \mid x) = \frac{1}{1+e^{-w^T x}}. \quad (6)$$

The parameters in the model were estimated using maximum likelihood. The use of Logistic Regression as an interpretable baseline model in medical classification problems is because it has a clear probability interpretation. K-Nearest Neighbors (KNN) was adopted as a non-parametric distance based classifier which classifies objects according to the majority decision of the closest neighbors in the Euclidean space:

$$\hat{y} = \arg\max_c \sum_{i \in N_k(x)} I(y_i = c), \quad (7)$$

where $N_k(x)$ represents the set of the k nearest samples to x. Since KNN depends directly on the distance between features, it is important that previous standardization take place in order to avoid bias by the features with higher scales. Support Vector Machine (SVM) was implemented by the use of a radial basis function (RBF) kernel to permit nonlinear decision boundaries. SVM tries to obtain a good hyperplane by solving:

$$\min_{w,b} \frac{1}{2} || w ||^2 + C \sum_{i=1}^{n} \xi_i, \quad (8)$$

Margin restricted. The parameter Ctrades off between the goal of maximizing margin and minimizing classification error. The implemented decision trees were recursive partitioning models which were implemented as splitting the feature space based on the impurity reduction. Entropy-based splitting was used in which the entropy can be defined as:

$$H(S) = -\sum_{i=1}^{k} p_i \log p_i. \quad (9)$$

Random Forest is an extension of Decision Trees through ensemble learning, which has the benefit of building multiple trees by bootstrap sampling and random feature selection. This last prediction is derived through majority voting:

$$\hat{y} = \text{mode}\big(T_1(x), T_2(x), \ldots, T_m(x)\big). \quad (10)$$

Gradient Boosting was employed to iteratively minimize residual error by sequentially adding weak learners:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x), \quad (11)$$

where $h_m(x)$ represents the weak learner in iteration m. This additive model is able to improve the predictive accuracy through the emphasis on the samples that had been previously misclassified. XGBoost which is an optimization of gradient boosting, uses regularization to avoid overfitting. Its objective functionality is as follows:

$$L = \sum l(y_i, \hat{y}_i) + \sum \Omega(f_k), \quad (12)$$

where $\Omega(f_k)$penalizes model complexity. XGBoost works well especially with structured tabular data. Principal Component Analysis (PCA) was used to examine feature redundancy as well as visualize class separability. PCA converts correlated variables into orthogonal principal components:

$$Z = XW, \quad (13)$$

where Wcontains eigenvectors of the covariance matrix. The explained variance ratio was computed as:

$$EVR_k = \frac{\lambda_k}{\sum \lambda_i}. \quad (14)$$

Visualizations were created based on the first two major components and offered the information on the behavior of class clustering. A combination of statistical measures was used to test model performance. Accuracy was computed as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}. \quad (15)$$

Precision and Recall were calculated as:

$$Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN}. \quad (16)$$

The F1-score, balancing Precision and Recall, was defined as:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}. \quad (17)$$

Threshold-independent evaluation was performed using Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC), where:

$$TPR = \frac{TP}{TP+FN}, FPR = \frac{FP}{FP+TN}. \quad (18)$$

Additionally, Precision–Recall curves were generated to better evaluate performance on the positive (malignant) class.

3. Results

In this section, the experimental results of the proposed multi-model comparative framework are given. Statistical, graphical and interpretability analysis is performed to analyze the results of the classifier behavior on the breast cancer dataset to have a complete picture of the behavior.

3.1. Exploratory Data Analysis Results

Exploratory Data Analysis (EDA) was also performed to investigate the statistical form, distribution characteristics, and relationship among features of the breast cancer data before the development of the models. Figure 2 shows the distribution of the classes of benign (0) and malignant (1) tumors. The con-data set has 569 samples consisting of 357 benign and 212 malignant. Despite the fact that benign samples are more common, the dataset is quite balanced, which eliminates the possibility of a serious classification bias. This distribution is optimal whenever there is strong supervised learning and the evaluation metrics are emphasized later in the sections.

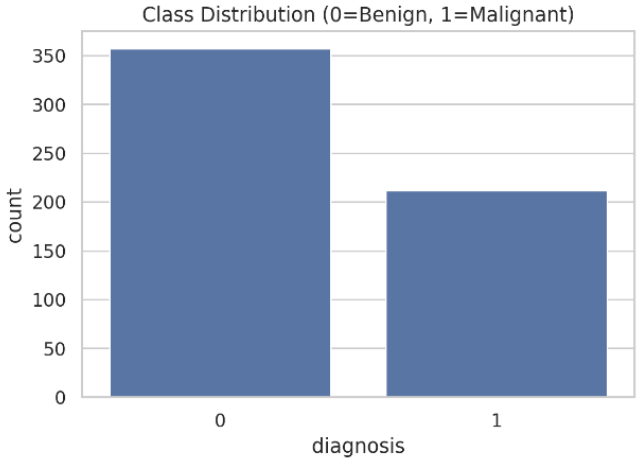

Figure 2. Distribution of benign and malignant samples.

The completeness of data was evaluated with the help of missing values bar plot and missing values matrix visualization presented in Figure 3 and Figure 4 respectively. In both of the visualizations, it is established that there are no missing values, as all the 30 numerical features have 569 non-null observations. The completeness of data used eliminates the possibility of statistical patterns being caused by the artifact of imputation instead of actual biological measurements.

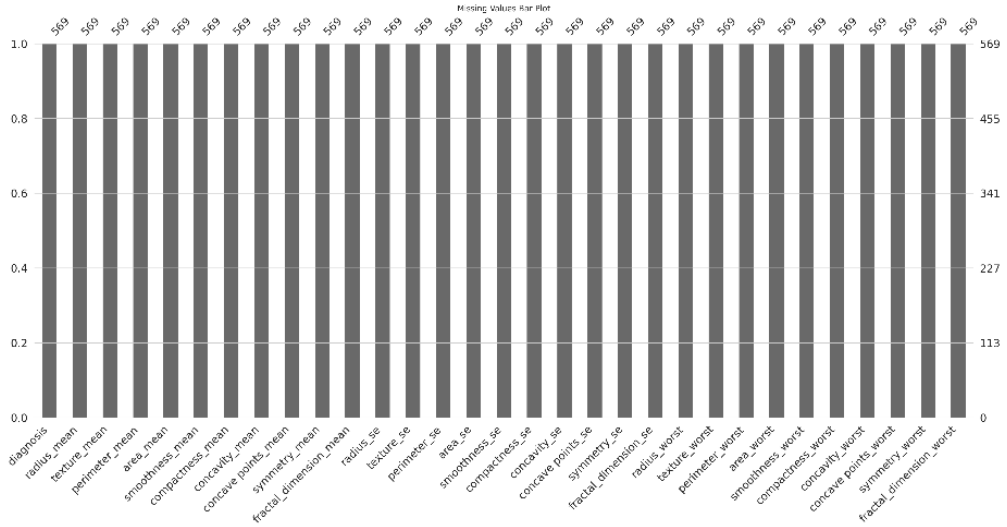

Figure 3. Feature-wise count of non-missing values.

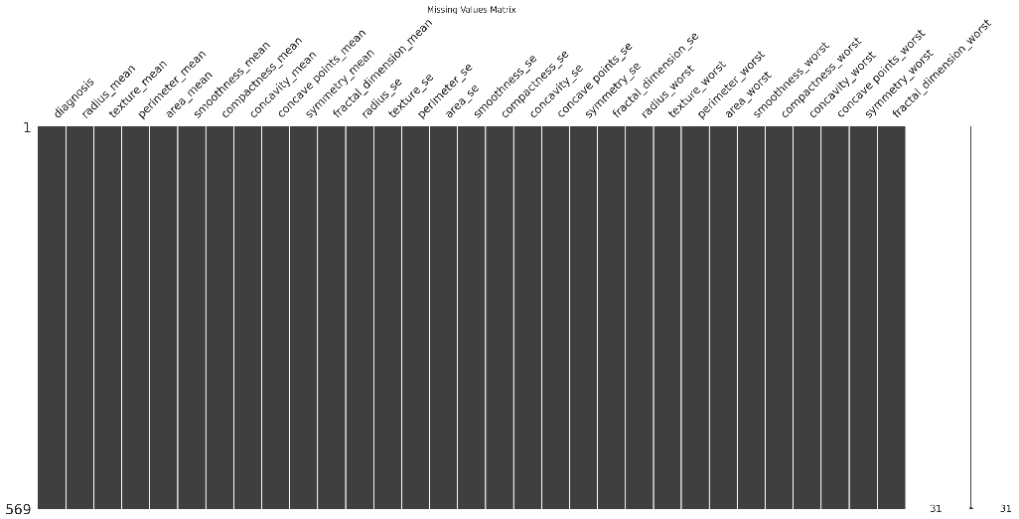

Figure 4. Matrix view of missing value structure.

The correlation structure of all the numerical features is shown in Figure 5. The heatmap once again shows a significant positive correlation between geometric size related characteristics (i.e., radius_mean, perimeter_mean, area_mean, and worst value of the characteristics) and the worst measurement of the characteristics. The existence of these near-linear relations is an indication of high levels of multicollinearity in the data set. Moreover, the feature of concavity, especially concave-points-mean and concave-points-worst, demonstrate positive correlate with the diagnosis variable significantly, which indicates that tumor boundary abnormality is a determinant factor of the malignancy. The patterns of clustered correlations indicate the redundancy of the geometric descriptors, and therefore, dimensionality reduction methods are applied in the next section.

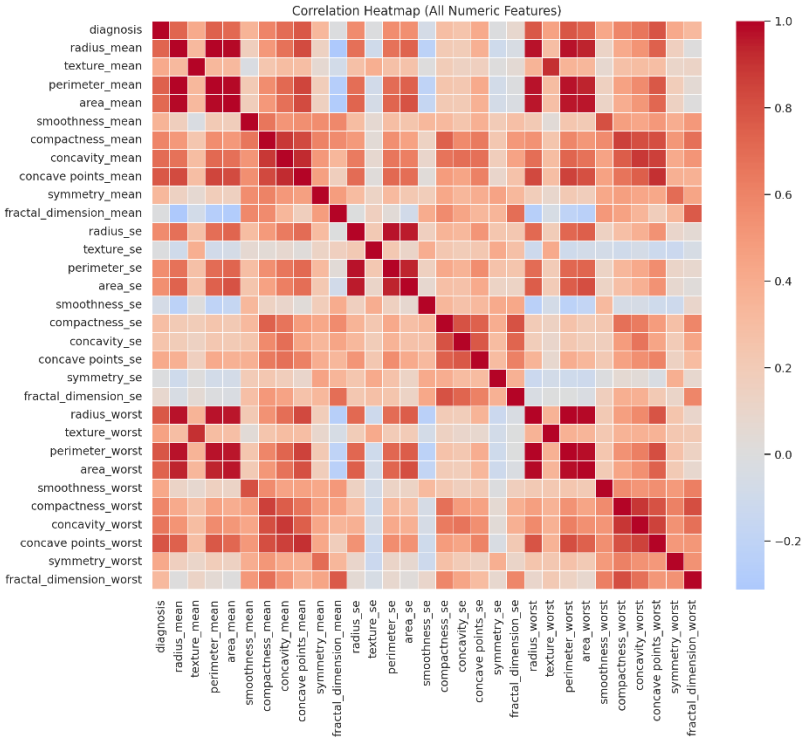

Figure 5. Correlation heatmap of all numerical features.

The distributional characteristic of the most influential features are shown in Figure 6 that demonstrates histograms of the top features with the highest relations with diagnosis. Geometric and concavity-based features always take up higher values of malignant tumors, especially perimeter worst, area worst, radius worst, and concave points worst. The apparent rightward movement of malignant distributions shows high discriminative ability.

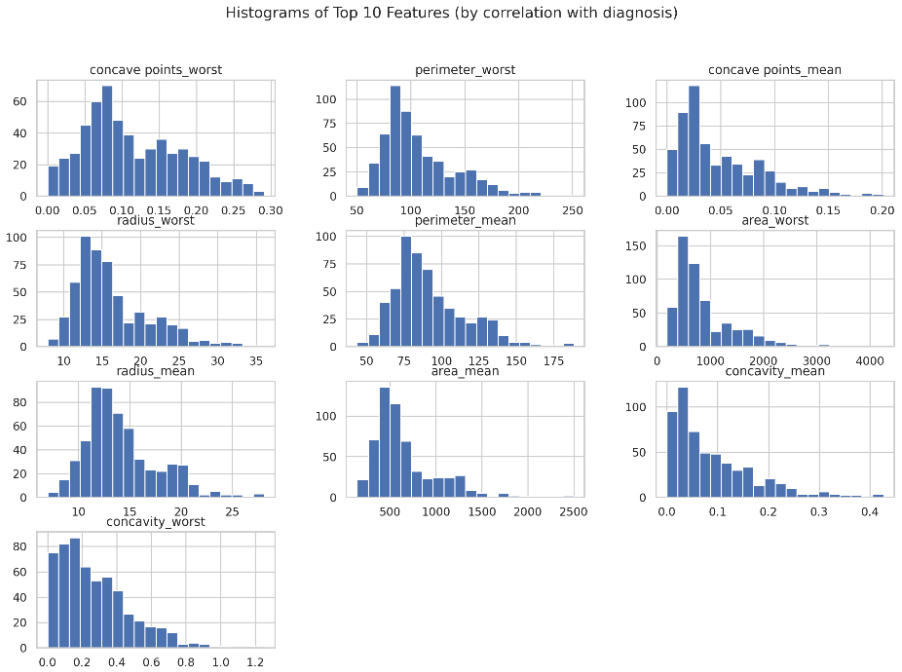

Figure 6. Histograms of top features correlated with diagnosis.

Additional separation of classes evidence is presented in Figure 7 that shows boxplots of the most discriminative features divided by diagnosis. The median values and interquartile ranges of malignant tumors are greater and wider in almost all significant characteristics. The separation of concave points worst and perimeter worst is particularly strong, with the least overlap between the classes observed. These results support the significance of morphology abnormality and tumor size in the detection of malignancy.

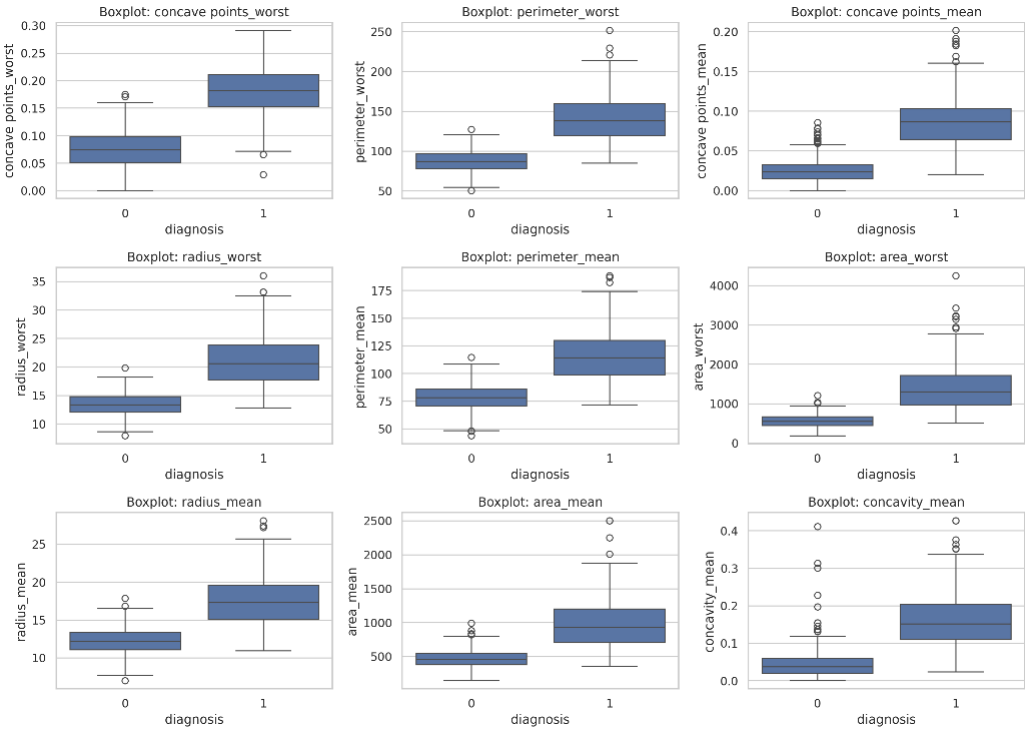

Figure 7. Boxplots of key features by diagnosis class.

Figure 8 illustrates pairwise relationships between the most significant features. The scatter-plots indicate close linear relationships among the geometric measurements of perimeter and radius. Besides, malignant samples are concentrated in areas with large magnitudes of features whilst benign samples are concentrated in the areas with low values. Poor as it may be, the general mode of clustering would indicate that there is a structural separability inherent in the dataset to be supervised classified.

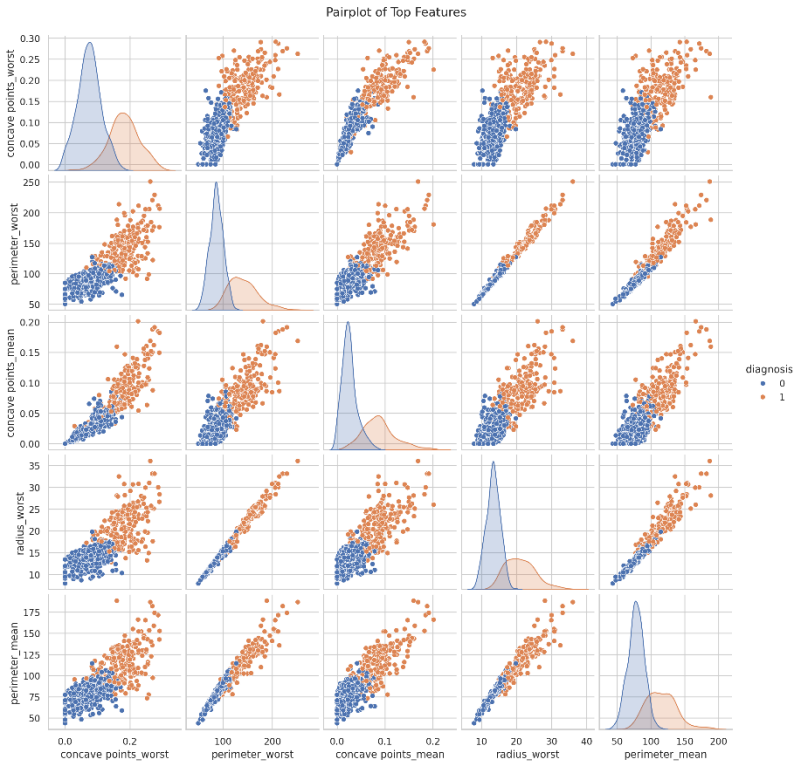

Figure 8. Pairwise feature relationships for top variables.

Finally, the exploratory analysis proves that the dataset is not incomplete, statistically consistent and that it has meaningful separability between benign and malignant tumors. Multicollinearity between geometric descriptors is also very strong, which makes dimensionality reduction justifiable, whereas distributional shifts on key features also suggest high predictive viability of machine learning models.

3.2. PCA-Based Dimensionality Reduction

Principal Component Analysis (PCA) was carried out to project the standardized features into a lower dimensional space, at the maximal level of variance. It was found that the first two major dimensions explained a significant amount of variability, and thus it was feasible to visualize the dataset structure on a two-dimensional scale.

Figure 9 shows projection of the training data using PCA. The scatter plot shows that there is noticeable clustering of benign and malignant samples along the PC1 axis. The malignant cases are more inclined to the high values of PC1 with the benign ones clustering at the lower levels. Though some small overlap is observed in the areas along class boundaries, the general separation of data ensures that there is high intrinsic discriminative structure in the data. But existence of overlap implies that we are not able to achieve perfect linear separability hence we may apply nonlinear classifiers in later modeling.

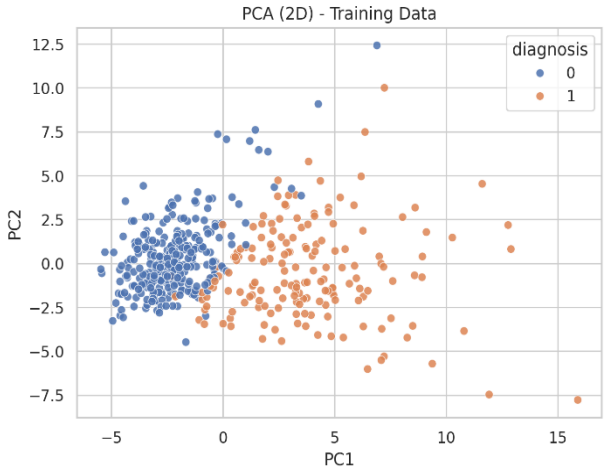

Figure 9. Two-dimensional PCA projection of training samples.

Figure 10 shows the contribution of original features to each of the main components. The loadings indicate that the size characteristics (geometric size, radius, perimeter, area) and the measures of concavity are the strongest contributors to PC1. These results support the patterns of correlation that were detected in Section 3.1 and point to tumor morphology as the predominant source of variance.

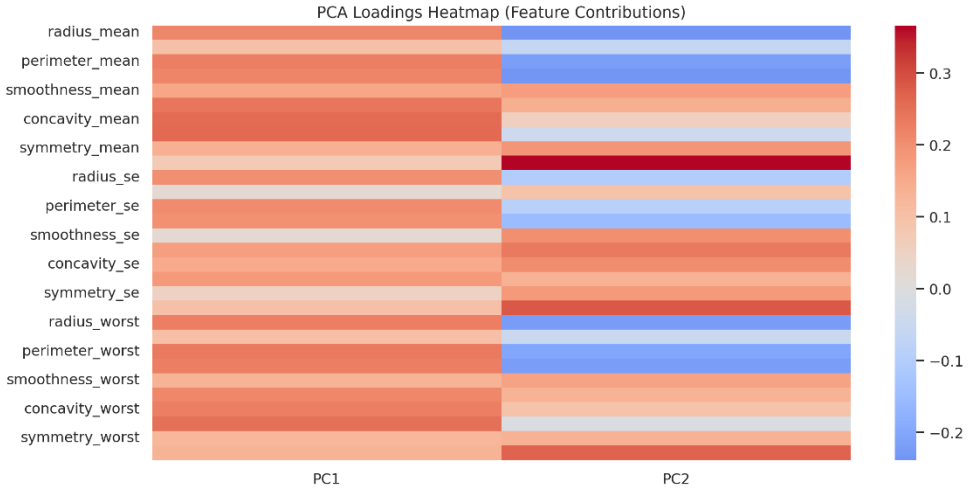

Figure 10. Heatmap of feature contributions to principal components.

Overall, PCA simply affirms that the two variables of significant separation in the classes are the tumor size and the abnormality of the boundaries, along with the necessity of using sophisticated classification schemes to represent the remaining non-linear variation.

3.3. Comparative Model Performance

The resultant quantitative comparison of all of the assessed classifiers appears in Table 1. To have balanced and threshold-independent evaluation, the models are evaluated based on the accuracy, precision, recall, F1-score, ROC-AUC and Precision Recall AUC.

Table 1. Comparative performance metrics of all evaluated machine learning models.

Model	Accuracy	Precision	Recall	F1	ROC_AUC	PR_AUC
Logistic Regression	0.974	0.976	0.952	0.964	0.993	0.991
Random Forest	0.974	0.953	0.976	0.965	0.980	0.983
Gradient Boosting	0.965	0.952	0.952	0.952	0.990	0.989
SVM	0.956	0.930	0.952	0.941	0.996	0.994
XGBoost	0.947	0.909	0.952	0.930	0.989	0.988
KNN	0.939	0.889	0.952	0.920	0.980	0.974
Decision Tree	0.930	0.854	0.976	0.911	0.939	0.843

Also the comparative analysis of the seven classifiers showed that all models had a high predictive performance. As it can be seen in Figure 11, ensemble-based methods, especially Gradient Boosting and XGBoost, made the most accurate predictions with a close second place by Random Forest. This further proves the high discriminative power of boosting based models with the ROC-AUC comparison in Figure 12.

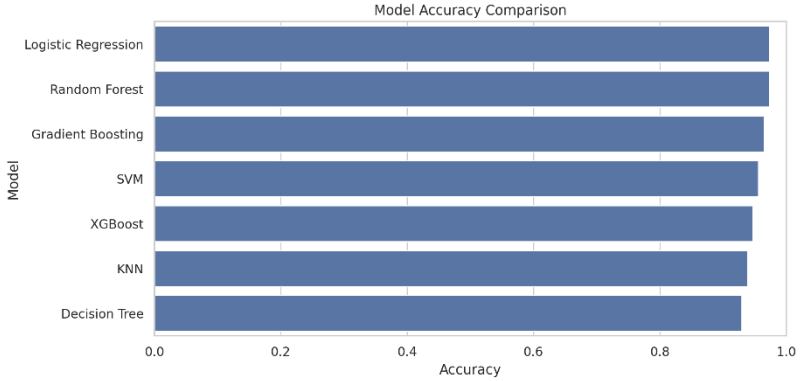

Figure 11. Accuracy comparison across classifiers.

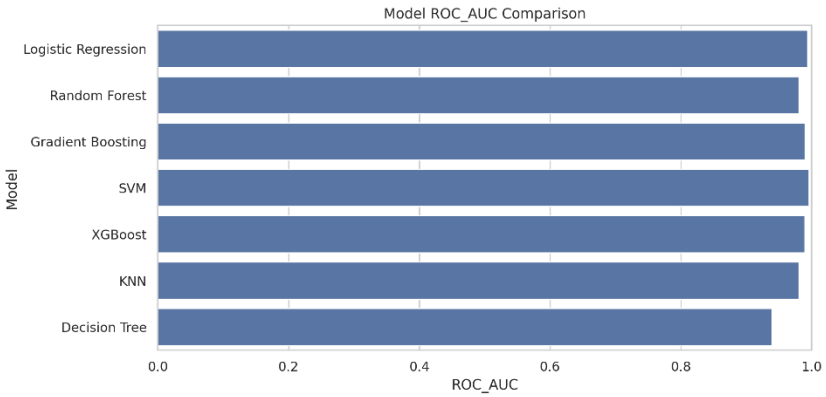

Figure 12. ROC-AUC comparison across classifiers.

The Logistic Regression also did well, which means that there is a lot of linear separability in the standardized feature space. Nevertheless, nonlinear models like SVM and boosting models were slightly better than their linear counterparts indicating the existence of higher-order interactions between features. KNN was found to perform competitively, albeit a little bit worse in recalling some boundary cases, where the standalone Decision Tree was underperforming compared to random Forest and it will be seen that ensemble based aggregation of classifiers is beneficial. Figure 13 and Figure 14 show that the ROC and the PrecisionRecall curves exhibit high classification power in all the models, and ensemble methods are more stable in all thresholds.

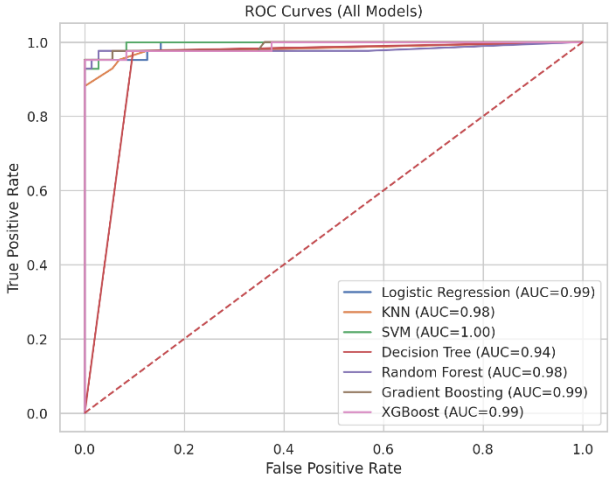

Figure 13. ROC curves for all evaluated models.

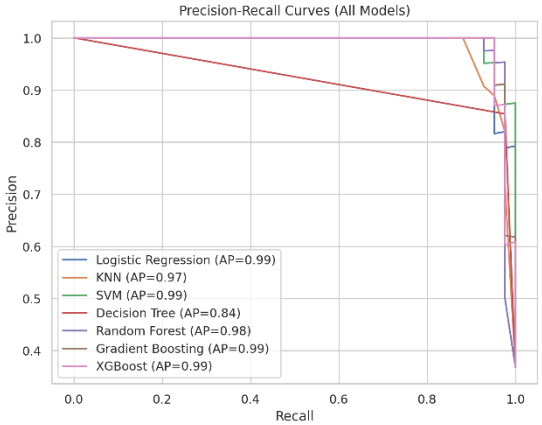

Figure 14. Precision–Recall curves for all evaluated models.

Overall performance ranking followed: Boosting-based models ≥ Random Forest ≥ SVM ≥ Logistic Regression ≥ KNN ≥ Decision Tree.

3.4. Confusion Matrix Analysis

Figure 15 (a-g) provides the confusion matrices of all the classifiers. False positives and false negatives were not high in any of the models indicating robust classification.

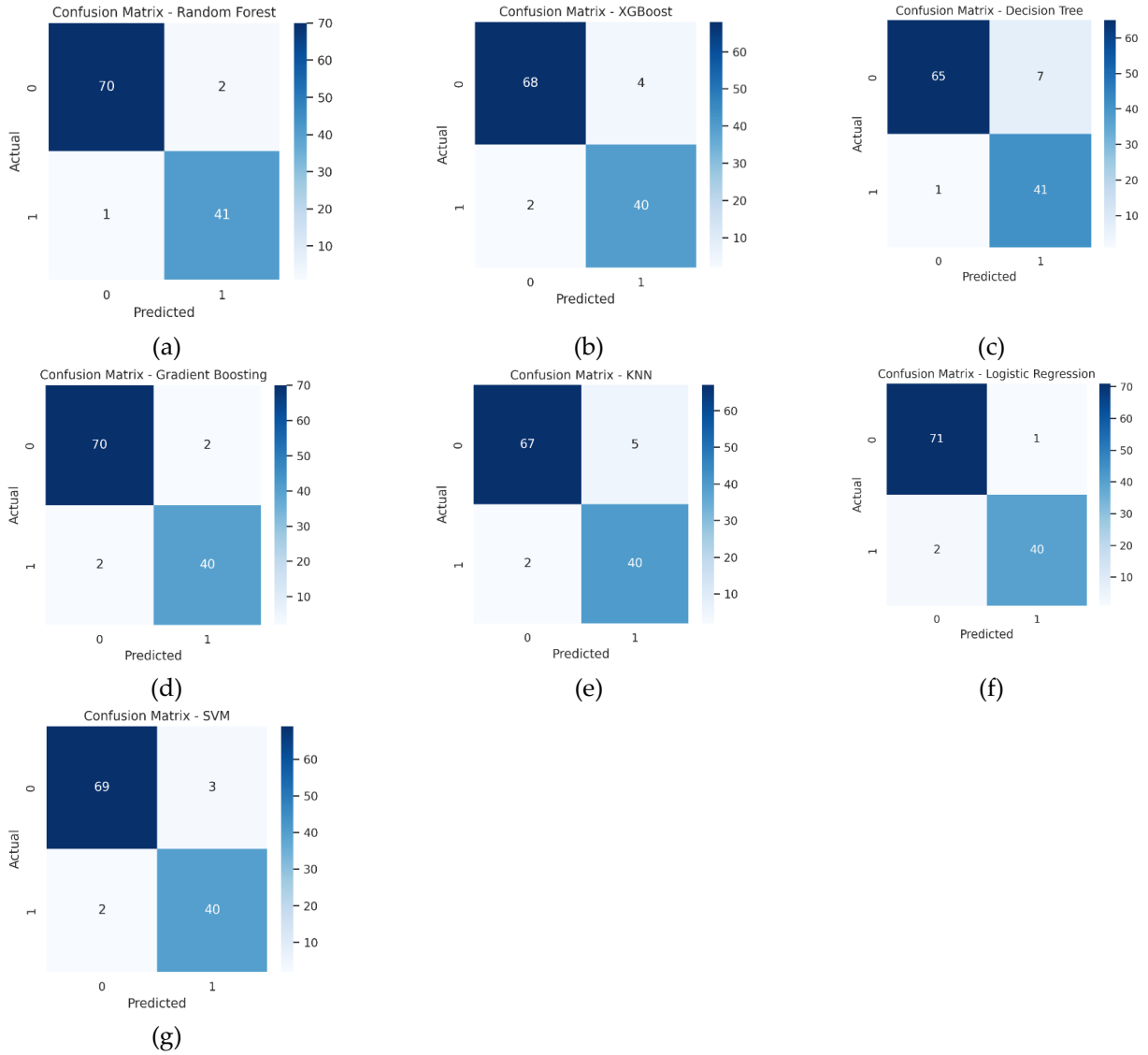

Figure 15. Confusion Matrix

When it comes to medical diagnosis, it is especially important to minimize the rate of false negatives that is, the false identifications of malignant tumours as benign. Ensemble methods particularly Random Forest, Gradient Boosting and XGBoost, showed the lowest false negativity resulting in an increased recall rate with the malignant group. This increased sensitivity renders boosting-based methods especially appropriate to the diagnostic decision-support system where missing malignant cases is clinically very dangerous. Logistic Regression and SVM also had a good proportion of sensitivity and specificity, with good and consistent classification outcomes. Conversely, the isolated Decision Tree was marginally more erratic in misclassification and this supports the benefits of ensemble aggregation in enhancing generalization and minimizing variance.

3.5. ROC Curve Analysis

Figure 16 illustrates the ROC curves of all the classifiers. All the models are highly convex to-ward towards the upper-left corner, which implies excellent discriminative capacity. Ensemble approaches, especially Gradient Boosting and XGBoost, had the greatest ROC-AUC values with a very close second place going to SVM. The proximity of the ROC curves is the fact that the dataset is organized in a highly informative way. However, the marginal improvements in the average of AUC of boosting-based models signify more robustness in boosting model at the classification threshold and improved processing of borderline cases.

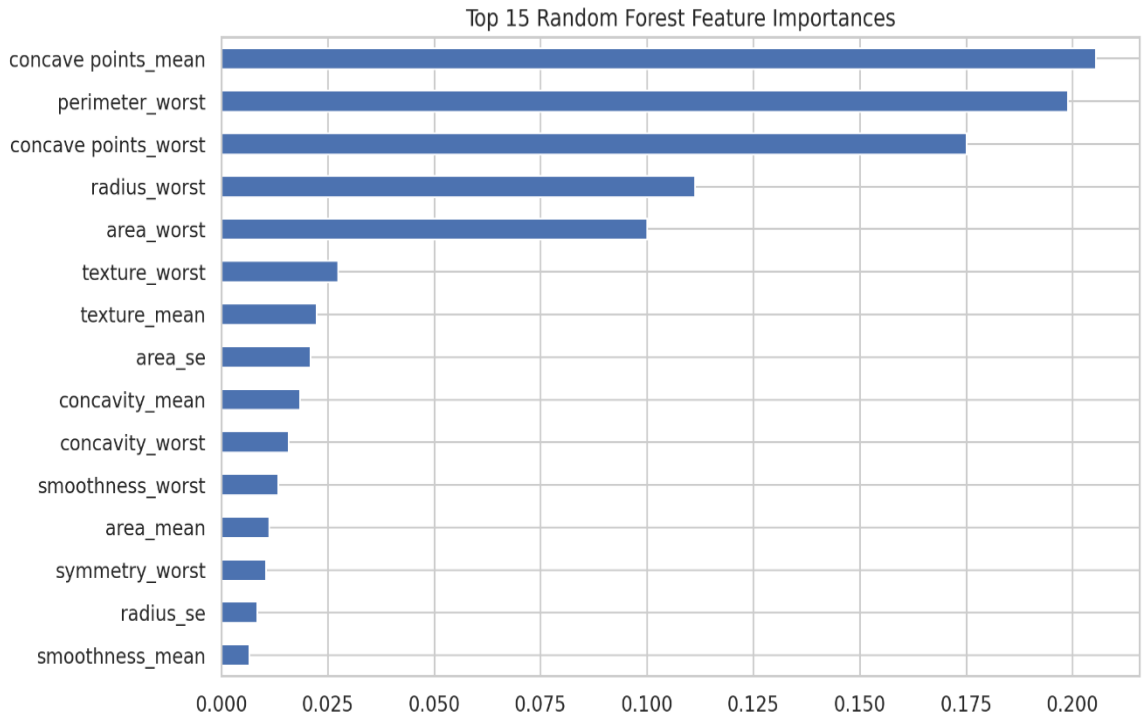

Figure 16. Top 15 feature importances from Random Forest.

3.7. Feature Importance and Interpretability

In the model, the importance of each feature is displayed and explained. Random Forest and XGBoost feature importance analysis is shown in Figure 17 and Figure 18 respectively. Geometric and concavity-related characteristics, including radius, perimeter, area, and concave points, are regular and strong predictors in both models.

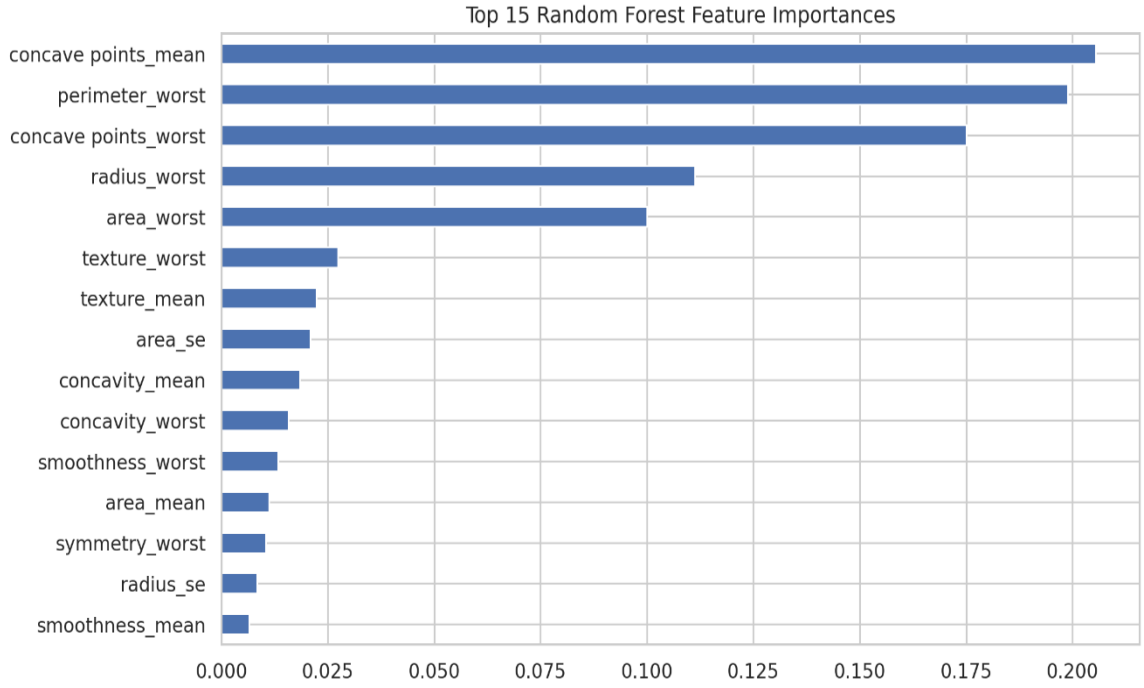

Figure 17. Top 15 feature importance from Random Forest

On the same note, it can be seen that similar trends are observed in the analysis of the Logistic Regression coefficient in Figure 19, which confirms the consistency between linear and nonlinear modes of modeling. Such consistency contributes to better interpretability and to the reliability of the identified predictive features.

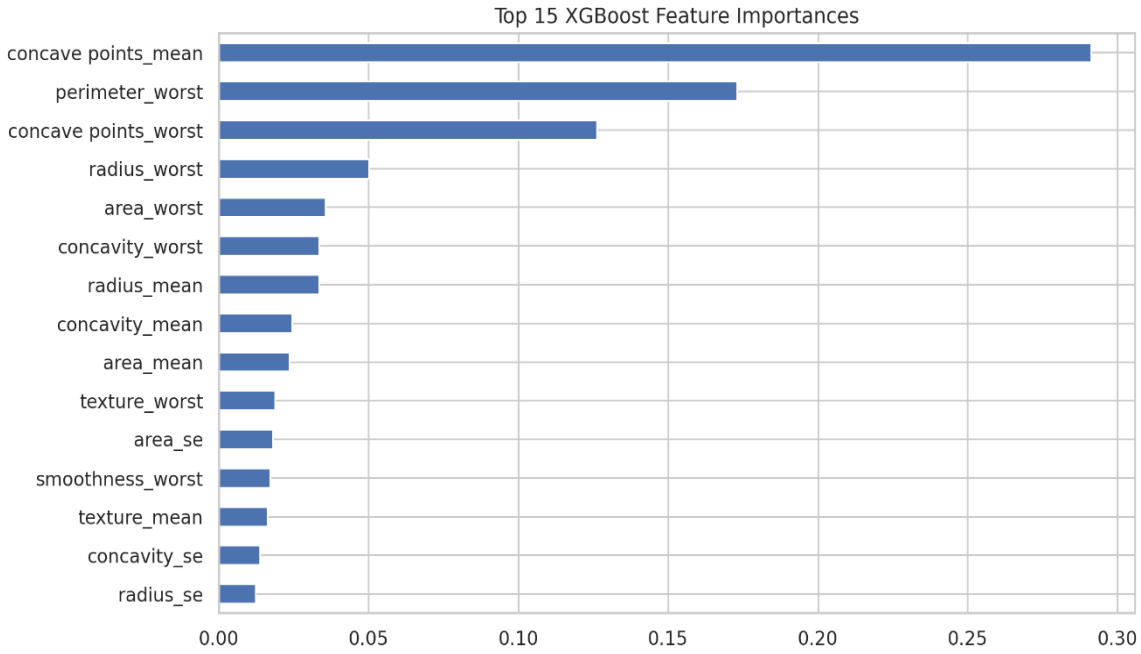

Figure 18. Top 15 feature importances from XGBoost.

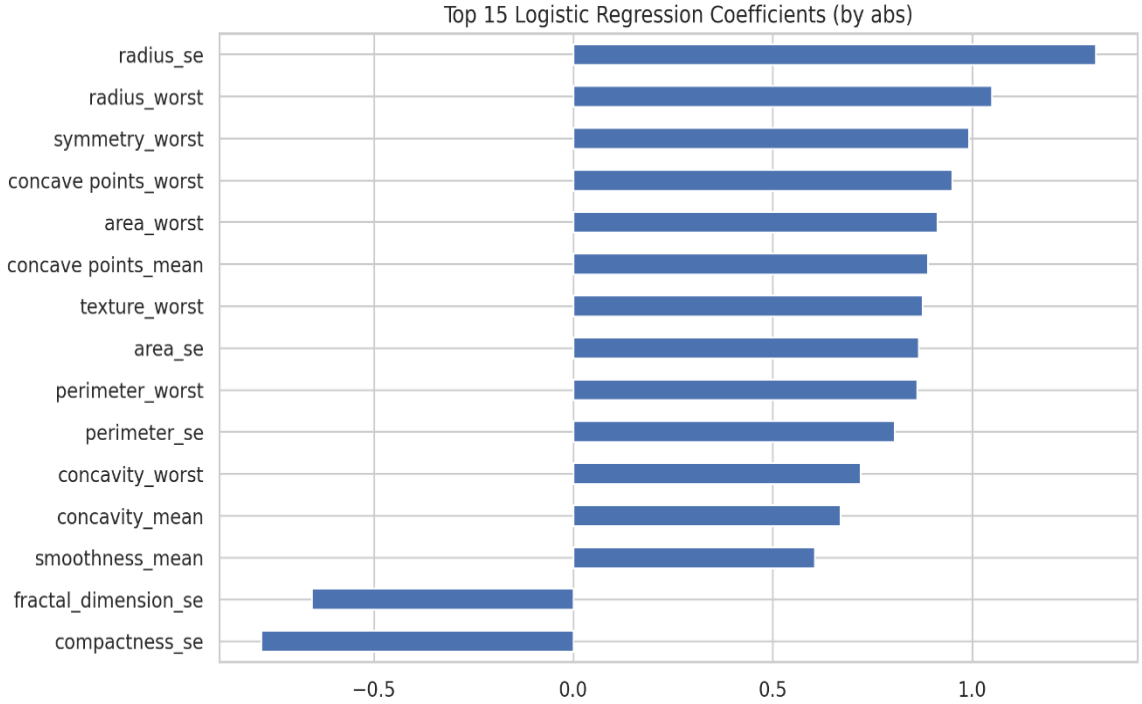

Figure 19. Top 15 coefficients from Logistic Regression.

4. Discussion

The findings of the experiment indicate that the dataset of breast cancer is very separable after standardized preprocessing. Linear models are highly performing, which implies that there exist robust discriminative signals in the feature space. Nevertheless, ensemble-based approaches always provide slight but significant increases in accuracy, ROC-AUC, and PR-AUC, which points to the capability of ensemble-based approach to understand nonlinear interactions and mitigate overfitting by aggregation. The combination of the complete EDA, dimensionality reduction, the multi-model comparison, the multimetric evaluation, and all of these features in one framework is methodologically transparent and reproducible. PCA verified the large variance capture at initial components, whilst ensemble models had a better robustness in hard boundary cases. Comprehensively, the results indicate that boosting-based

ensemble models would make the most dependable performance in the structured classification of breast cancer though linear models are also viable, interpretable, and computationally efficient choices.

5. Conclusions

This paper has designed a reproducible and well-organized machine learning system to classify breast cancer based on morphological features. Exploratory analysis affirmed that geometry attributes had high levels of feature separability and multicollinearity. PCA indicated tumor size and concavity as dominant sources of variance. Comparison of seven classifiers revealed uniformly high predictive per-performance, and boosting-based ensemble approaches showed a higher performance than linear and single-tree models in a variety of measures. Thresholds and confusion matrix analysis ensured that low false negative rates of the best models were achieved, which is important in medical diagnosis. In general, ensemble methods offer the most stable performance when it comes to structured tasks of breast cancer classification, whereas linear models are still viable and understandable competitors. The suggested framework provides an effective and clear background of decision-support systems in clinical applications.

Funding: This research received no external funding.
Data Availability Statement: The dataset used in this study is publicly available without any restrictions and can be downloaded using the following link: https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic
Conflicts of Interest: The authors declare no conflict of interest
Acknowledgement: The authors are thankful to the Deanship of Graduate Studies and Scientific Research at University of Bisha for supporting this work through the Fast-Track Research Support Program.

References

1. World Health Organization, "Breast cancer," 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer
2. American Cancer Society, "Breast Cancer Facts & Figures 2024–2025," 2024. [Online]. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/2024/breast-cancer-facts-and-figures-2024.pdf
3. S. Aamir et al., "Predicting breast cancer leveraging supervised machine learning techniques," Comput. Math. Methods Med., vol. 2022, p. 5869529, 2022.
4. I. Ozcan, H. Aydin, and A. Cetinkaya, "Comparison of classification success rates of different machine learning algorithms in the diagnosis of breast cancer," Asian Pacific Journal of Cancer Prevention, vol. 23, no. 10, p. 3287, 2022.
5. 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.
6. V. Jaiswal, P. Saurabh, U. K. Lilhore, M. Pathak, S. Simaiya, and S. Dalal, "A breast cancer risk prediction and classification model with ensemble learning and big data fusion," Decision Analytics Journal, vol. 8, p. 100298, 2023.
7. D. Sharda, R. Bansal, P. Gumber, S. Kandari, and I. Arora, "Comparative evaluation of support vector classifier, random forest, and XGBoost for early breast cancer prediction with feature importance and class balancing," Cureus Journal of Computer Science, vol. 2, no. 1, 2025.
8. R. K. Yadav, P. Singh, and P. Kashtriya, "Diagnosis of breast cancer using machine learning techniques—A survey," Procedia Comput. Sci., vol. 218, pp. 1434–1443, 2023.
9. A. Bilal et al., "Fusion of genomic and pathological data for breast cancer detection using BCDNN," Front. Med. (Lausanne)., vol. 13, p. 1726223, 2026.
10. A. Bilal, A. Alkhathlan, F. A. Kateb, A. Tahir, M. Shafiq, and H. Long, "A quantum-optimized approach for breast cancer detection using SqueezeNet-SVM," Sci. Rep., vol. 15, no. 1, p. 3254, 2025.
11. A. Bilal et al., "BC-QNet: A quantum-infused ELM model for breast cancer diagnosis," Comput. Biol. Med., vol. 175, p. 108483, 2024.
12. A. Bilal, A. Imran, T. I. Baig, X. Liu, E. Abouel Nasr, and H. Long, "Breast cancer diagnosis using SVM optimized by quantum-inspired grey wolf optimization," Sci. Rep., vol. 14, no. 1, p. 10714, 2024.
13. A. Bilal, M. Shafiq, W. J. Obidallah, Y. A. Alduraywish, and H. Long, "Quantum computational infusion in extreme learning machines for early multi-cancer detection," J. Big Data, vol. 12, no. 1, p. 27, 2025.
14. E. Richardson, R. Trevizani, J. A. Greenbaum, H. Carter, M. Nielsen, and B. Peters, "The receiver operating characteristic curve accurately assesses imbalanced datasets," Patterns, vol. 5, no. 6, 2024.
15. M. Owusu-Adjei, J. Ben Hayfron-Acquah, T. Frimpong, and G. Abdul-Salaam, "Imbalanced class distribution and performance evaluation metrics: A systematic review," PLOS Digital Health, vol. 2, no. 11, p. e0000290, 2023.
16. S. Liu et al., "Comparison of evaluation metrics of deep learning for imbalanced imaging data in osteoarthritis studies," Osteoarthritis Cartilage, vol. 31, no. 9, pp. 1242–1248, 2023.
17. B. A. Akinnuwesi, B. O. Macaulay, and B. S. Aribisala, "Breast cancer risk assessment and early diagnosis using PCA and SVM techniques," Inform. Med. Unlocked, vol. 21, p. 100459, 2020.
18. K. Bian, M. Zhou, F. Hu, and W. Lai, "RF-PCA: A new solution for rapid identification of breast cancer categorical data," Front. Genet., vol. 11, p. 566057, 2020.
19. P. S. Murty et al., "Integrative hybrid deep learning for enhanced breast cancer diagnosis," Sci. Rep., vol. 14, no. 1, p. 26287, 2024.
20. UCI Machine Learning Repository, "Breast Cancer Wisconsin (Diagnostic)," 1995.