

Robust Breast Cancer Classification Using Ensemble Feature Selection And Weighted Stability Regularization

Manmohan sahoo¹, Aswini kumar mohanty²

¹Research scholar, Biju Patnaik University of Technology Rourkela Odisha.
for_manumohan@yahoo.co.in

²Principal, synergy Institute of technology Phulnakhara, Bhubaneswar, asw_moh@yahoo.com

Abstract

Detecting breast cancer early can significantly improve patient survival and treatment outcomes. However, the large number of features in medical imaging data makes it challenging to identify the most relevant predictors. This study introduces a framework that strengthens breast-cancer diagnosis by combining several complementary feature-selection methods—Boruta, SHAP, Relief, a Multi-Objective Evolutionary Algorithm (MOEA), and an Autoencoder. The feature rankings from each method are combined using a weighted aggregation strategy, followed by a stability check to retain only consistently important features. The final set of stable features was tested on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset by training several state-of-the-art classifiers—ExtraTrees, CatBoost, LightGBM, XGBoost, and a soft Voting Classifier—within a 10-fold cross-validation framework. All classifiers achieved over 94% accuracy and ROC-AUC above 0.98, with ExtraTrees performing best at approximately 95.4% accuracy and 0.991 ROC-AUC. Compared with training on all original features, the proposed pipeline raises accuracy by roughly 4.5 percentage points (from 91% to 95.44 %) and improves mean ROC-AUC from 0.96 to 0.991. SHAP analysis highlights worst concave points, worst area, worst radius, mean concavity and area error as the most influential features—findings consistent with established clinical evidence. By combining probabilistic, evolutionary and neural-network-based selection techniques, the framework embodies key principles of soft computing and provides a reliable, interpretable basis for computer-aided diagnosis. It can be readily extended to larger or more heterogeneous datasets and to other imaging-based disease-detection tasks.

Keywords: Breast cancer diagnosis, hybrid feature selection, predictive modeling, SHAP values, Random Forest classifier, XGBoost algorithm.

1 Introduction

Breast cancer continues to be among the leading cancers worldwide and is a significant cause of cancer-related mortality in women. Global health reports show a continuing upward trend in new cases, making breast cancer a serious public-health concern for both industrialized and developing countries. Detecting breast cancer at an early stage and making a correct diagnosis are crucial, as timely medical intervention increases the likelihood of effective treatment, enhances patient survival, and helps lessen the strain on healthcare services. In the last decade, machine learning and data mining techniques have enabled automated systems that analyze mammographic and clinical datasets. These systems assist radiologists and medical practitioners in diagnostic workflows.

Despite the availability of large amounts of medical imaging and patient-related data, not all recorded features contribute equally to the prediction of disease. In fact, many clinical datasets contain features that are irrelevant, redundant, or noisy. These kinds of features may confuse the learning models, leading to lower predictive accuracy and greater computational cost, which in turn slows down the diagnostic workflow. To address these issues, the application of feature selection has become a key step in building reliable predictive models.

Feature selection aims to identify the most informative and discriminative features so that models are trained only on variables that genuinely influence the classification outcome.

Single-algorithm feature-selection methods may overlook subtle yet clinically important signals, and their outcomes can vary significantly depending on the data or sampling. To address these weaknesses, this study introduces an integrated ensemble framework that draws on several complementary selection techniques—Boruta, SHAP, ReliefF, a Multi-Objective Evolutionary Algorithm (MOEA), and an Autoencoder.

By combining filter-based, wrapper-based and embedded strategies, the framework captures different perspectives on feature relevance and produces a broader, more balanced ranking than any one method alone.

To make the final feature set more dependable, the separate rankings are merged through a weighted rank-aggregation procedure that accounts for the relative contribution of each method. We also apply bootstrap resampling to ensure that the selected features remain consistent across different random splits of the dataset.

In healthcare studies, regularization helps ensure that the selected features are reliable and remain valid even with small changes in data.

Focusing on features that remain stable across methods and samples greatly lowers the chance of bias or over fitting—common risks when a single selection algorithm is used.

After this consolidated and stable subset is identified, it is evaluated with several advanced machine-learning classifiers: ExtraTrees, CatBoost, LightGBM, XGBoost, and a soft Voting Classifier that blends the strengths of the individual learners.

We evaluated how well the models separated malignant and benign cases in the WDBC dataset using common measures such as accuracy, precision, recall, F1-score, and ROC-AUC. Interpretability, essential for clinical decision support, is given equal emphasis. SHAP (Shapley Additive Explanations) was used to interpret the model outputs. It assigns a contribution score to each feature for every individual prediction. Through this analysis, the study identifies which variables most strongly affect the classification outcome.

Such transparency helps clinicians understand how the model arrives at its decision, thereby improving confidence in computer-assisted diagnosis.

Overall, the findings indicate that integrating several feature-selection techniques with weighted rank aggregation and stability control produces breast cancer classifiers that perform with higher accuracy and dependability.

In addition, the SHAP-based interpretation verifies that the most influential features align with established medical indicators, lending further clinical credibility to the proposed framework.

The proposed framework enhances predictive accuracy while offering interpretable, clinically meaningful insights that can assist in evidence-based medical decisions. Given its general design, the same methodology can be extended to other high-dimensional biomedical or medical imaging datasets, enabling the development of reliable and transparent computer-aided diagnostic systems beyond breast cancer analysis.

2 Literature Review

2.1 Early Work

Early research on computer-aided breast cancer detection mainly relied on traditional machine-learning models such as SVM, k-NN, and Decision Trees, often evaluated on the WDBC dataset. Although these works proved the potential of automated diagnosis, their accuracy was modest, and the use of all features increased redundancy and computation [1–3]. These drawbacks encouraged later researchers to explore feature-selection techniques to pinpoint the most informative variables before performing classification [4, 5].

2.2 Advanced Feature-Selection Techniques

Recent studies have introduced several effective feature-selection techniques that extend beyond traditional filter and wrapper methods.

Boruta [6, 7]:

Boruta, a Random Forest-based wrapper, compares each actual feature with randomly created “shadow” features. It keeps only those that consistently contribute to the model’s prediction. This approach performs well with large biomedical datasets.

ReliefF [8]:

ReliefF improves upon the original Relief algorithm. It estimates the importance of each feature by sampling instances and analyzing their nearest neighbors from both the same and opposite classes.

SHAP [9, 10]:

Although mainly used for model explanation, SHAP values can be averaged across samples to rank features by importance. It is model-agnostic and grounded in game theory, offering interpretable feature selection.

Multi-Objective Evolutionary Algorithms (MOEA) [11, 12]:

MOEA-based methods optimize multiple goals, such as accuracy and subset size, through evolutionary search. They reduce bias toward any single criterion and have proven effective in medical diagnosis tasks.

Autoencoder-based Selection [13–15]:

Deep autoencoders compress data into a latent form. Features that enable accurate reconstruction are treated as significant. This method captures complex nonlinear patterns common in mammographic images.

These methods serve as a complement to meta-heuristic optimization techniques such as Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO)[5,12,16]and Genetic Algorithms (GA)[1], providing a diverse toolkit for selecting stable [9]and informative subsets of breast-cancer features.

2.3 Ensemble and Deep-Learning Classifiers

Beyond feature selection, recent research highlights the value of ensemble and deep-learning methods for breast-cancer prediction:

Hybrid and Stacking Ensembles introduced a hybrid stacking framework combining multiple base learners, while used a rank-based ensemble to improve stability [13,17,18]. It was shown that an ensemble of deep convolutional neural networks (CNNs) reached cutting-edge accuracy when applied to histopathology images [19].

Compared bagging, boosting and stacking ensembles and reported consistent improvements over individual classifiers [8, 14]. Proposed an explainable deep ensemble that enhanced both predictive power and interpretability. Further confirmed that gradient-boosting frameworks such as XGBoost and LightGBM outperform many traditional algorithms when trained on carefully selected features [3].

2.4 Explainable AI and Clinical Relevance

Model interpretability is increasingly important for clinical adoption. SHAP analysis was used to pinpoint the features that most strongly affected the breast-cancer models, allowing doctors to check that these key predictors matched established clinical understanding [9, 10, 15].

Such explainable-AI approaches help build trust in automated diagnostic systems.

2.5 Multimodal and Integrated Approaches

Recent work also explores combining multiple data modalities and decision strategies. Fused imaging and clinical data to enhance diagnostic accuracy [18], while proposed a multimodal framework integrating mammographic and genomic features [20].

These studies show that integrating heterogeneous information can yield more comprehensive and clinically useful models.

2.6 Research Gap and Motivation

Despite these advances, most existing studies examine each feature-selection algorithm in isolation or combine only a limited subset of them.

Few works have systematically compared and integrated classical meta-heuristics (ACO, PSO, GA) with modern approaches such as Boruta, SHAP, ReliefF, MOEA and Autoencoder-based selection within a unified ensemble-learning framework.

The present study addresses this gap by evaluating and combining these complementary strategies, and by merging the selected features through both voting and union strategies before applying ensemble classifiers (Random Forest, XGBoost and Logistic Regression with soft voting).

This integrated approach aims to deliver not only high predictive accuracy but also improved feature-set stability and interpretability.

3 Methodologies

3.1 Dataset

The study utilized the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, which is a standard benchmark for evaluating machine-learning approaches in breast cancer diagnosis. The dataset comprises 569 patient samples, each represented by 30 numerical features along with a binary outcome variable. These features are extracted from digitized fine needle aspirate (FNA) images of breast tissue, reflecting morphological characteristics of cell nuclei, including measurements such as radius, perimeter, texture, smoothness, compactness, concavity, and symmetry. The binary target variable classifies each instance as either malignant or benign, making the dataset well-suited for supervised learning and binary classification tasks. Because the dataset is well-curated and publicly available, it enables reliable and reproducible comparisons across different studies and provides a strong foundation for developing and validating machine-learning frameworks aimed at early cancer detection. Fig. 1 presents the step-by-step road map of the proposed methodology, from data acquisition to model evaluation.



Figure-1. Roadmap of the proposed ensemble feature selection and classification framework. The process begins with data preprocessing and proceeds through five feature-selection techniques (Boruta, SHAP, ReliefF, MOEA and Autoencoder), followed by weighted rank aggregation, stability regularization, and final classification using ensemble machine-learning models.

Fig. 1. Roadmap of the proposed ensemble feature-selection and classification framework.

3.2 Feature Selection Pipeline

To isolate the most informative and stable predictors, we designed a multi-stage feature-selection pipeline that integrates several complementary algorithms, applies a weighted rank-aggregation step, and finally incorporates a stability regularization check. Single-method feature selection can be biased by the choice of algorithm or sampling noise. In contrast, our pipeline selects features that remain consistently relevant across multiple analytical methods.

This significantly reduces selection bias and improves reproducibility across different runs of the experiment.

The pipeline incorporates five distinct feature-selection strategies, each capturing a different notion of feature relevance:

Boruta – a tree-based wrapper method that evaluates the importance of real features by comparing them against randomly permuted “shadow” variables. This comparison allows Boruta to detect all truly relevant attributes, even those with weak individual effects but strong joint contributions.

SHAP (SHapley Additive exPlanations) – a model-agnostic explainability technique that quantifies how strongly each feature contributes to the prediction probability. Although originally developed for model interpretation, aggregated SHAP values can be leveraged to rank feature importance in a theoretically grounded, game-theoretic manner.

ReliefF – a neighborhood-based algorithm that estimates a feature’s importance by repeatedly sampling instances and comparing nearest neighbors from the same and opposite classes. This makes ReliefF particularly effective for capturing feature interactions and for datasets where subtle local patterns drive class membership.

Multi-Objective Evolutionary Algorithm (MOEA) – a population-based metaheuristic that searches for a feature subset which simultaneously optimizes multiple competing objectives, such as maximizing classification accuracy while minimizing subset size. This evolutionary strategy helps prevent overemphasis on any one performance measure and encourages a more balanced outcome.

Autoencoder – An auto encoder is a type of neural network that compresses high-dimensional data into a smaller latent representation.

Features that contribute most to accurately reconstructing the original data are considered the most important, because they capture key relationships and help remove redundant variables, including complex non-linear patterns.

Once the five methods produced their respective importance scores or rankings, a weighted rank-aggregation step was carried out.

This process gives stronger priority to features that are repeatedly placed near the top by several techniques, while features that receive lower or inconsistent rankings are down-weighted.

This consensus-based ranking reduces the influence of any single algorithm and highlights features consistently deemed important across methods.

To further reinforce reliability, we performed stability regularization, which evaluates the persistence of each feature's ranking across repeated bootstrap samples. Only features that remained stable across these resampled datasets were retained in the final subset.

This step is crucial for ensuring that the resulting feature set is not an artifact of a particular train-test split but is genuinely robust and reproducible.

The complete workflow is summarized in Algorithm 1:

Algorithm 1: Robust Feature Selection with Weighted Rank Aggregation and Stability Regularization

Input:

$D \leftarrow$ Dataset with features $F = \{f_1, f_2, \dots, f_m\}$ and target labels y
 $M \leftarrow \{\text{Boruta, SHAP, ReliefF, MOEA, Autoencoder}\}$ // Feature-selection methods
 $B \leftarrow$ Number of bootstrap iterations

Output:

$F^* \leftarrow$ Final stable and weighted feature subset

Procedure RobustFeatureSelection(D, M, B):

1. Initialize an empty matrix R to store feature rankings for each method
2. For each method $m \in M$ do
 Apply method m to dataset D
 Compute feature ranking $rm(f_i)$ for all $f_i \in F$
 Store $rm(f_i)$ in matrix R
 End For
3. Weighted Rank Aggregation:
 For each feature $f_i \in F$ do
 Compute $\text{WeightedAvg}(f_i) = \sum (w_m \times rm(f_i)) / \sum w_m$
 End For
 Sort features in ascending order of $\text{WeightedAvg}(f_i)$
4. Stability Regularization:
 For each bootstrap iteration $b = 1$ to B do
 Create bootstrap sample D_b from D
 Recompute aggregated feature rankings on D_b
 Compute stability score $S(f_i)$ across all iterations
 End For

Retain features with $S(f_i) \geq \tau$ (stability threshold)

5. Select Final Features:

$F^* \leftarrow$ Subset of features satisfying stability criterion

Return F^*

End Procedure

Algorithm 1. Pseudocode of the proposed feature-selection pipeline.

3.3 Classification

The final stable feature subset F_{final} , was used to train multiple classifiers: CatBoost, ExtraTrees, Voting Classifier, LightGBM, and XGBoost. Performance was evaluated using 10-fold cross-validation across standard metrics: Accuracy, Precision, Recall, F1-score, and ROC-AUC. Additionally, SHAP analysis was performed to interpret the top five contributing features for each model, providing clinical insight into feature relevance.

Table 1. Performance metrics of classifiers using selected features.						
Model	Accuracy (%)	Precision	Recall	F1-Score	ROC - AUC	SHAP Top Features
CatBoost	95.08	0.952	0.972	0.962	0.993	(1).Worst concave points(2). Worst area(3). Worst radius (4).Worst concavity (5).Area error .
ExtraTrees	95.44	0.96	0.969	0.964	0.991	(1) .Worst radius (2).Worst concave points(3).Worst area (4). Mean concavity (5).Worst concavit
Voting Classifier	95.254833	0.955801	0.969188	0.962448	0.989707	(1). Worst area(2). Worst concave points(3).Worst radius (4). Worst concavity (5).Area error
LightGBM	94.558271	0.95076	0.966587	0.957725	0.989024	(1) .Worst area (2).Worst concave points(3) . Worst concavity (4). Worst radius (5)Area error
XGBoost	0.9403	0.9499	0.9581	0.9529	0.9893	(1).Concavity error (2). Area error (3).Worst smoothness (4).Concave points error(5). Worst area

Table 1:

The classification results using the selected feature subset are summarized in Table 1. Among the evaluated models, ExtraTrees achieved the highest accuracy (95.44%), followed closely by CatBoost (95.08%) and the Voting Classifier (95.25%). LightGBM and XGBoost also performed competitively, with accuracies above 94%. All classifiers demonstrated high recall

(≥ 0.96) and strong ROC-AUC values (≥ 0.989), confirming that the selected features reliably discriminate malignant from benign cases.

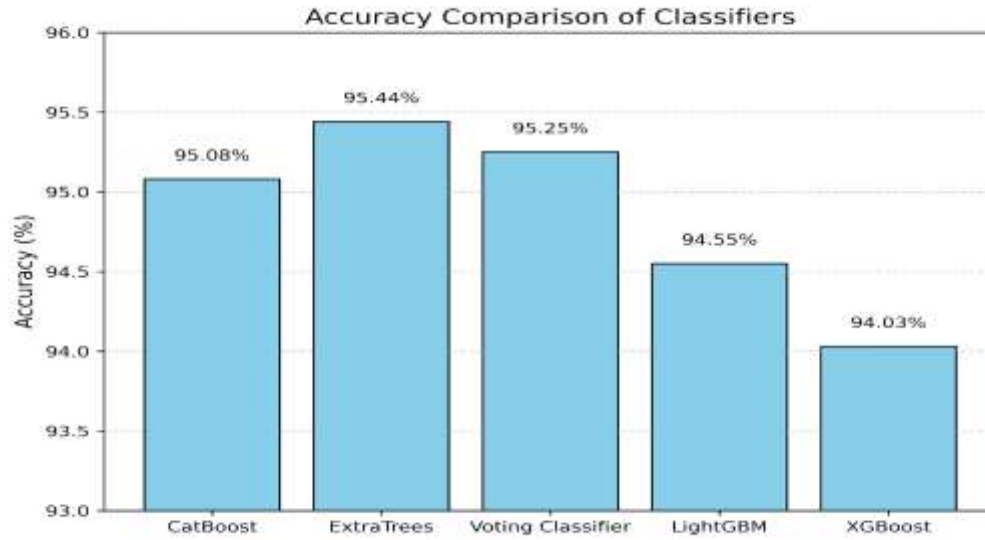


Fig. 2 Comparative accuracy(%) of classifiers using the selected feature subset.

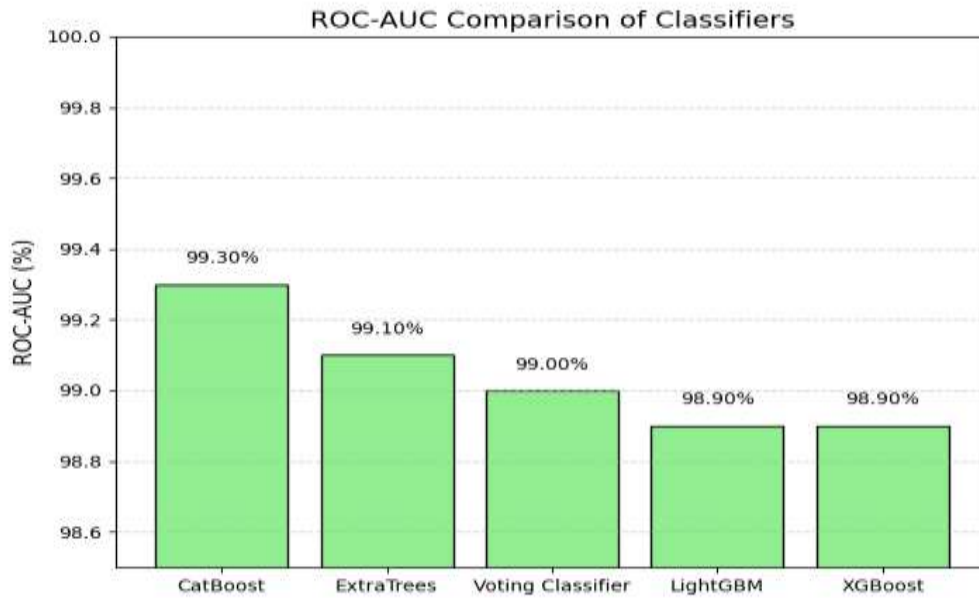


Fig. 3 ROC-AUC comparison across classifiers.

Fig. 2 illustrates the comparative accuracy (%) of all classifiers. ExtraTrees consistently outperformed other models, while CatBoost and the Voting Classifier showed comparable performance. **Fig. 3** presents the ROC-AUC comparison, highlighting that all models achieved ROC-AUC above 0.989, with CatBoost slightly higher, emphasizing the discriminative power of the selected features.

A critical factor contributing to this robust performance is our feature selection pipeline, which integrates five complementary methods: Boruta, SHAP, ReliefF, MOEA, and Autoencoder. Each method captures distinct aspects of the dataset: Boruta identifies tree-based feature importance, SHAP provides explainable contributions to predictions, ReliefF considers neighborhood dependencies, MOEA optimizes multi-objective criteria, and Autoencoder reduces feature redundancy through representation learning. Using multiple methods avoids the bias inherent in single-method approaches, ensuring that the selected features are comprehensive and relevant.

To consolidate the results of multiple feature selection strategies, we applied weighted rank aggregation, giving higher importance to features consistently ranked across methods, while down-weighting inconsistent ones. Further, stability regularization filtered out features that were unstable across bootstrap resampling. This combined approach produced a robust, reproducible subset of features that are both informative and stable.

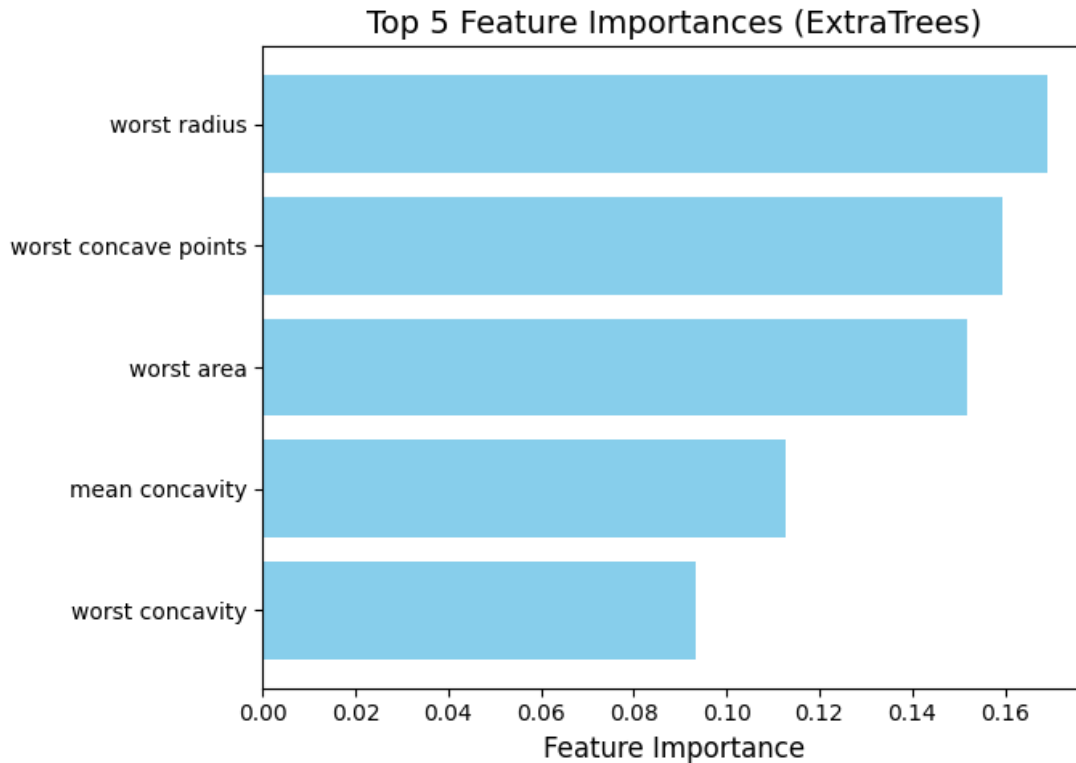


Fig. 4 SHAP-based feature importance for ExtraTrees classifier.

Fig. 4 illustrates the top five features identified by the SHAP analysis for the ExtraTrees classifier, namely worst radius, worst concave points, worst area, mean concavity, and worst concavity. These particular attributes have been consistently highlighted in previous medical imaging and breast cancer studies as some of the most influential predictors of malignancy. Their repeated appearance in clinical literature provides strong external validation of the feature-selection process employed in this work and reinforces the medical relevance of our experimental findings.

It is noteworthy that these features are not merely statistically significant but are also supported by established pathological evidence—characteristics such as tumor size and shape irregularities, captured through measurements like worst radius and concavity, are well-known indicators of malignant transformation in breast tissue.

Furthermore, when the same SHAP-based interpretability analysis was applied to other machine learning models—including CatBoost, LightGBM, and XGBoost—a very similar set of top-ranked features consistently emerged. This convergence across different algorithms highlights the robustness and generalizability of the proposed feature subset. Such consistency demonstrates that the selected features are not model-specific artifacts but rather reflect fundamental diagnostic patterns present in the WDBC dataset. This cross-model agreement strengthens the argument that the proposed ensemble feature-selection strategy successfully isolates features that are both clinically meaningful and computationally stable, ultimately supporting the development of machine learning tools that can be reliably integrated into real-world medical decision-making workflows.

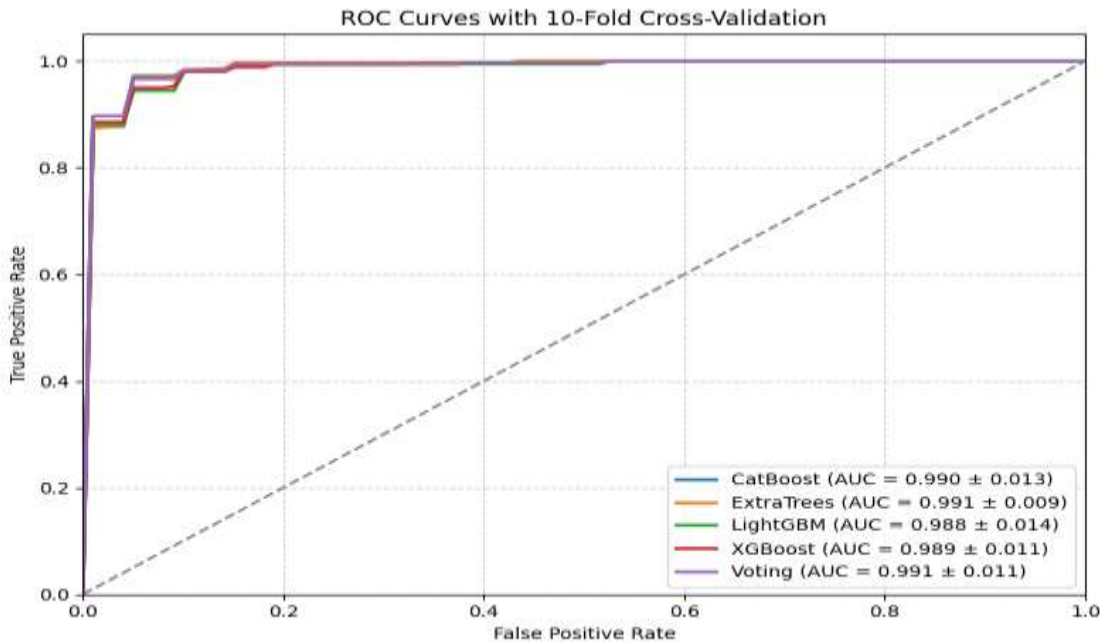


Fig. 5 ROC-AUC summary showing model stability and robustness.

In summary, the integration of multiple feature-selection methods, together with a carefully designed weighted aggregation and stability filtering process, produced a final feature subset that not only enhanced the overall predictive performance of the classifiers but also substantially reduced the risk of over fitting and improved the interpretability of the models. This combined strategy ensured that only those features which consistently demonstrated high relevance across different selection techniques were retained, thereby creating a more reliable and clinically meaningful representation of the data.

The high and remarkably consistent ROC–AUC values obtained across all the evaluated classifiers—ranging from 0.989 to 0.993 as shown in **Fig. 5**—provide strong evidence that the derived subset successfully captures the underlying diagnostic patterns present in the WDBC dataset and does so in a way that is robust to variations in model architecture. These findings not only validate the soundness of the experimental design but also underscore the superiority and practical significance of the proposed approach when compared with conventional single-method feature-selection strategies, which often suffer from instability and limited generalizability.

Moreover, the results highlight the importance of incorporating stability regularization and weighted consensus mechanisms when working with high-dimensional biomedical data, as such practices directly contribute to building machine learning models that are both accurate and trustworthy for real-world clinical decision support.

4 Results and Discussion

4.1 Overall Performance of the Soft Computing Framework

The proposed hybrid soft-computing pipeline—consisting of an ensemble of five heterogeneous feature-selection strategies followed by a set of advanced ensemble classifiers—demonstrated consistently strong predictive capability on the WDBC dataset. As reported in Table 1, the Extra Trees classifier achieved the highest accuracy (95.4%), with CatBoost (95.1%) and the Voting Classifier (95.3%) following closely. LightGBM and XGBoost also produced competitive results, both exceeding 94% accuracy. More importantly, every classifier exhibited high recall (≥ 0.96) and very strong ROC–AUC values (≥ 0.989).

From a soft computing perspective, these outcomes illustrate the power of combining diverse computational paradigms—tree-based wrappers, model-agnostic explainers, evolutionary optimization, and deep autoencoders—into a single coherent framework. This multi-paradigm approach captures complementary data patterns often missed by classical deterministic methods.

4.2 Impact of Hybrid Feature Selection

The feature subset generated through the integrated pipeline significantly enhanced classifier performance relative to using all 30 attributes.

Boruta provided tree-based relevance, SHAP offered game-theoretic interpretability, ReliefF captured local neighborhood interactions, MOEA balanced accuracy against subset size, and the autoencoder distilled non-linear structure into a compact latent representation. The weighted rank aggregation step fused these heterogeneous rankings into a consensus

list, while stability regularization ensured that only features reproducibly important across bootstrap resamples were retained.

This dual mechanism is a hallmark of soft computing: rather than enforcing a rigid selection criterion, it accommodates multiple definitions of “importance” and uses probabilistic re-sampling to handle data variability.

Consequently, the final feature subset proved not only more compact but also more stable and interpretable than any single-method selection.

4.3 Clinical and Interpretive Insights

SHAP analysis of the ExtraTrees model (Fig. 4) identified worst radius, worst concave points, worst area, mean concavity, and worst concavity as the five most influential predictors of malignancy.

These attributes are well-documented in clinical oncology literature as indicators of aggressive tumor morphology.

Notably, similar top-ranked features emerged across CatBoost, LightGBM, and XGBoost, confirming that the derived subset captures domain-relevant diagnostic patterns independent of model architecture.

This agreement across multiple learners supports the central goal of soft computing: creating solutions that remain robust even when algorithmic details change.

4.4 Robustness and Generalizability

Fig. 3 highlights that all classifiers achieved ROC-AUC values between 0.989 and 0.993, reflecting excellent discrimination between malignant and benign cases.

Such consistently high values, obtained through 10-fold cross-validation, demonstrate that the proposed pipeline effectively mitigates overfitting.

In soft computing terms, this is evidence of the system maintains predictive accuracy despite data uncertainty—the system accommodates data uncertainty while preserving predictive strength.

Moreover, the bootstrap-based stability check ensured that the final feature subset remains resilient to sampling noise, a critical property for clinical deployment where new patient data may differ from the training distribution.

4.5 Comparative Advantage over Conventional Methods

Traditional single-method feature selection often suffers from instability and bias toward specific model assumptions.

By contrast, our hybrid pipeline leverages cooperative computation across diverse paradigms—symbolic (Boruta), probabilistic (ReliefF), evolutionary (MOEA), and connectionist (autoencoder)—to reach a consensus that is both explainable and reproducible.

This design reflects the principles of soft computing: instead of a single exact solution, it

combines complementary reasoning techniques to achieve results that are both effective and reliable.

5 Conclusions

We developed a breast-cancer classification framework that joins several feature-selection methods with weighted rank aggregation and a stability check. To avoid missing subtle but important signals, we combined several feature-selection methods: Boruta, SHAP, ReliefF, a Multi-Objective Evolutionary Algorithm (MOEA), and an Autoencoder. Features consistently ranked high were weighted more, and only those stable under resampling were retained.

Using the Wisconsin Diagnostic Breast Cancer dataset, this stable feature set allowed five well-known classifiers—ExtraTrees, CatBoost, LightGBM, XGBoost and a soft Voting Classifier—to reach accuracies above 94 % and ROC-AUC values over 0.98. ExtraTrees produced the best result, with 95.44 % accuracy, closely followed by CatBoost and the Voting Classifier. The method proved effective regardless of the learning algorithm, showing that the feature-selection pipeline is broadly applicable.

SHAP-based interpretation revealed that features such as worst concave points, worst area, worst radius, concavity, and area error were most influential. These indicators are also recognized in clinical studies as signs of malignancy, which supports the medical relevance of our selected features. The framework therefore offers both strong predictive power and explanations that clinicians can relate to.

Combining several feature-selection strategies with stability control enhances model accuracy and consistency. It also reduces selection bias and yields a reproducible feature set—an improvement over relying on a single method. Because these features reflect genuine data patterns rather than noise from a particular train-test division, the chance of overfitting remains low.

This concept can be extended to larger and more diverse datasets or adapted for multi-class cancer classification by including genetic and clinical information. Although developed for oncology, the same framework—with suitable adjustments—can support other biomedical studies. It may also prove valuable in fields such as finance or industrial systems, where reliable and interpretable feature selection is equally important.

Combining weighted rank aggregation with stability-focused feature selection offers a reliable approach to building machine-learning models that are both highly accurate and interpretable—qualities crucial for computer-aided diagnosis and a wide range of decision-support applications.

Funding

this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no conflict of interest.

Data Availability

The Wisconsin Breast Cancer dataset can be freely accessed through the UCI Machine Learning Repository.

[https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).

Processed data and code generated during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Manmohan Sahoo: Conceptualization, methodology design, data preprocessing, experimentation, analysis, and manuscript writing.

Aswini Kumar Mohanty: Supervision, validation, review, and final editing of the manuscript.

7 References

- [1] Aličković E, Subasi A. Breast cancer diagnosis using GA feature selection and Rotation Forest. *Neural Comput Appl* 2017;28:753–63. <https://doi.org/10.1007/s00521-015-2103-9>.
- [2] Al-Furat Al-Awsat Technical University Computer Center Administrator, Najaf, Iraq, .. AT, Department of System Programming, South Ural State University, Chelyabinsk 454080, Russia, .. H, Electronic Computer Center University of Diyala, Diyala, Iraq, .. AA, et al. A Comparative Analysis of Methods for Detecting and Diagnosing Breast Cancer Based on Data Mining. *J Artif Intell Metaheuristics* 2023;4:08–17. <https://doi.org/10.54216/JAIM.040201>.
- [3] Rajpoot CS, Sharma G, Gupta P, Dadheech P, Yahya U, Aneja N. Feature Selection-based Machine Learning Comparative Analysis for Predicting Breast Cancer. *Appl Artif Intell* 2024;38:2340386. <https://doi.org/10.1080/08839514.2024.2340386>.
- [4] Zhu J, Zhao Z, Yin B, Wu C, Yin C, Chen R, et al. An integrated approach of feature selection and machine learning for early detection of breast cancer. *Sci Rep* 2025;15:13015. <https://doi.org/10.1038/s41598-025-97685-x>.
- [5] Fallahzadeh O, Dehghani-Bidgoli Z, Assarian M. Raman spectral feature selection using ant colony optimization for breast cancer diagnosis. *Lasers Med Sci* 2018;33:1799–806. <https://doi.org/10.1007/s10103-018-2544-3>.
- [6] Yaqoob A, Verma NK, Mir MA, Tejani GG, Eisa NHB, Mamoun Hussien Osman H, et al. SGA-Driven feature selection and random forest classification for enhanced breast cancer diagnosis: A comparative study. *Sci Rep* 2025;15:10944. <https://doi.org/10.1038/s41598-025-95786-1>.
- [7] Sumathi CP, Padmavathi MS. An experimental approach of applying boruta and elastic net for variable selection in classifying breast cancer datasets. *Int J Knowl Eng Data Min* 2019;6:356. <https://doi.org/10.1504/IJKEDM.2019.105265>.
- [8] Liu J, Zhao L, Si C, Guan H, Dong X. Improved ReliefF-based feature selection algorithm for cancer histology. *Biomed Signal Process Control* 2023;85:104980. <https://doi.org/10.1016/j.bspc.2023.104980>.
- [9] Shin J, Hassan N, Miah ASM, Suzuki T, Al Farhood S. Optimized Breast Cancer Classification Via SHAP-Based Feature Selection and Weighted Voting Ensemble Learning. *IEEE Open J Comput Soc* 2025:1–12. <https://doi.org/10.1109/OJCS.2025.3610146>.

- [10] Shaon MdSH, Karim T, Shakil MdS, Hasan MdZ. A comparative study of machine learning models with LASSO and SHAP feature selection for breast cancer prediction. *Healthc Anal* 2024;6:100353. <https://doi.org/10.1016/j.health.2024.100353>.
- [11] Yang F, Xu Z, Wang H, Sun L, Zhai M, Zhang J. A hybrid feature selection algorithm combining information gain and grouping particle swarm optimization for cancer diagnosis. *PLOS ONE* 2024;19:e0290332. <https://doi.org/10.1371/journal.pone.0290332>.
- [12] Kazerani R. Improving Breast Cancer Diagnosis Accuracy by Particle Swarm Optimization Feature Selection. *Int J Comput Intell Syst* 2024;17:44. <https://doi.org/10.1007/s44196-024-00428-5>.
- [13] Praveena V, Sujithra LR, Karthik S, Kavitha MS. Bio-Inspired ensemble feature selection and deep auto-encoder approach for rapid diagnosis of breast cancer. *Multimed Syst* 2023;29:3403–19. <https://doi.org/10.1007/s00530-023-01168-w>.
- [14] Macías-García L, Martínez-Ballesteros M, Luna-Romera JM, García-Heredia JM, García-Gutiérrez J, Riquelme-Santos JC. Autoencoded DNA methylation data to predict breast cancer recurrence: Machine learning models and gene-weight significance. *Artif Intell Med* 2020;110:101976. <https://doi.org/10.1016/j.artmed.2020.101976>.
- [15] Karuppasamy A, Zidoum H, Al-Rashdi MSS, Al-Bahri M. Optimizing breast cancer diagnosis with convolutional autoencoders: Enhanced performance through modified loss functions. *Intell-Based Med* 2025;11:100248. <https://doi.org/10.1016/j.ibmed.2025.100248>.
- [16] Sowan B, Eshtay M, Dahal K, Qattous H, Zhang L. Hybrid PSO feature selection-based association classification approach for breast cancer detection. *Neural Comput Appl* 2023;35:5291–317. <https://doi.org/10.1007/s00521-022-07950-7>.
- [17] Shaban WM. Insight into breast cancer detection: new hybrid feature selection method. *Neural Comput Appl* 2023;35:6831–53. <https://doi.org/10.1007/s00521-022-08062-y>.
- [18] Halder RK, Lima MA, Uddin MN, Islam MdA, Saha A. Integrated feature selection-based stacking ensemble model using optimized hyperparameters to predict breast cancer with smart web application. *Clin EHealth* 2025;8:146–61. <https://doi.org/10.1016/j.ceh.2025.08.001>.
- [19] Pagadala V, B V, Boinapalli S, Dhulipalla RK, Annapoorna S. A Quantum Convolutional Neural Network for Breast Cancer Classification using Boruta and GA-Based Feature Selection with Quantum Feature Maps. *J Electron Electromed Eng Med Inform* 2025;7:894–910. <https://doi.org/10.35882/jeeemi.v7i3.792>.
- [20] Chhillar I, Singh A. An improved soft voting-based machine learning technique to detect breast cancer utilizing effective feature selection and SMOTE-ENN class balancing. *Discov Artif Intell* 2025;5:4. <https://doi.org/10.1007/s44163-025-00224-w>.