Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

Optimizing Breast Cancer Classification: SVM and Random Forest with Hybrid Hyperparameter Tuning and Feature Selection

Adil Setiawan[*1], Soeheri[2]

[1,2]Department of Computer Science, Universitas Potensi Utama, Indonesia

Email: [1]adio165@gmail.com

Received : Feb 7, 2026; Revised : Feb 9, 2026; Accepted : Feb 9, 2026; Published : Jun 15, 2026

Abstract

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide, underscoring the urgent need for early, accurate, and reliable diagnostic support systems. This study proposes an optimized breast cancer classification framework using Support Vector Machine (SVM) and Random Forest (RF) models enhanced through hybrid hyperparameter tuning and feature selection. The Breast Cancer Wisconsin (Diagnostic) dataset, comprising 569 samples with 30 numerical features derived from Fine Needle Aspirate (FNA) examinations, was utilized in this research. Feature selection was conducted using Random Forest feature importance to identify the most relevant diagnostic attributes and reduce dimensionality. Hybrid hyperparameter tuning was implemented using GridSearchCV combined with 5-fold cross-validation to obtain optimal model configurations. Model performance was evaluated using accuracy, malignant-class recall, confusion matrix analysis, and Receiver Operating Characteristic–Area Under the Curve (ROC–AUC). Experimental results show that the optimized SVM model achieved significant improvements in accuracy, recall, and ROC–AUC compared to baseline models, indicating enhanced sensitivity and discrimination capability, while the Random Forest model maintained stable performance with marginal gains after optimization. These findings highlight the critical importance of systematic optimization strategies in improving diagnostic safety and reducing false negatives, thereby contributing to the development of more reliable and clinically applicable machine learning-based medical decision support systems.

Keywords : *Breast Cancer Classification, Feature Selection, Hyperparameter Tuning, Random Forest, Support Vector Machine.*

1. INTRODUCTION

Cancer remains one of the most critical global health challenges and continues to be a leading cause of death worldwide. Among various cancer types, breast cancer is the most frequently diagnosed cancer in women, with incidence and mortality rates continuing to increase annually [1], [2]. Early and accurate diagnosis is essential to improve survival rates; however, detecting breast cancer at an early stage remains challenging because clinical symptoms are often subtle and not clearly observable.

Recent advances in Artificial Intelligence (AI), particularly in machine learning and deep learning, have significantly transformed medical diagnostics [3], [4]. AI-based approaches are capable of analyzing large and complex medical datasets and identifying hidden patterns that are difficult to detect using conventional methods [5], [6]. In oncology, these techniques have been applied to medical imaging, clinical records, and molecular data to support cancer detection, prognosis prediction, and clinical decision-making [7], [8]. Although deep learning methods dominate medical image analysis research [9], classical machine learning algorithms remain highly relevant for structured tabular datasets with numerical features and moderate sample sizes [10], [11].

Among classical approaches, Support Vector Machine (SVM) and Random Forest (RF) are widely adopted due to their robustness and strong classification capabilities. SVM is effective in

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

constructing non-linear decision boundaries through kernel functions, making it suitable for high-dimensional data [12]. Random Forest, as an ensemble learning method, enhances stability and reduces variance by aggregating multiple decision trees [13]. Both algorithms have been extensively applied in breast cancer classification and have demonstrated competitive performance across various datasets [14], [15]

However, model performance is not determined solely by algorithm selection. Hyperparameter configuration plays a critical role in achieving optimal generalization [16]. Inadequate tuning may result in overfitting or underfitting, thereby reducing reliability when models are applied to unseen clinical data. In the context of medical diagnosis [17], classification errors—especially false negatives in which malignant cases are misclassified as benign—can lead to delayed treatment and serious clinical consequences. Therefore, optimization strategies are essential not only to improve overall accuracy but also to enhance sensitivity and model discrimination capability [18].

A review of prior studies indicates that many works emphasize algorithm comparison or report high accuracy without thoroughly analyzing clinically significant metrics such as malignant-class recall and ROC–AUC [19], [20], [21]. Some studies evaluate breast cancer classification models on benchmark datasets [22], [23]. yet limited research systematically compares baseline and optimized configurations while explicitly examining the impact of hyperparameter tuning and feature selection on classification errors [24]. Furthermore, reproducible experimental designs that transparently evaluate optimization effects across multiple performance metrics remain relatively underreported [25], [26].

This gap motivates the present study. Therefore, this research aims to optimize and comprehensively evaluate SVM and Random Forest models for breast cancer classification using the Breast Cancer Wisconsin (Diagnostic) dataset. Optimization is conducted through hybrid hyperparameter tuning and feature selection, with evaluation focusing on accuracy, malignant-class recall, confusion matrix analysis, and ROC–AUC. The originality of this study lies in (1) systematically comparing baseline and optimized models using clinically relevant metrics; (2) empirically analyzing the impact of optimization strategies on reducing false-negative classifications; and (3) providing a reproducible experimental framework that highlights the practical significance of hybrid optimization in improving diagnostic reliability for breast cancer classification.

2. METHOD

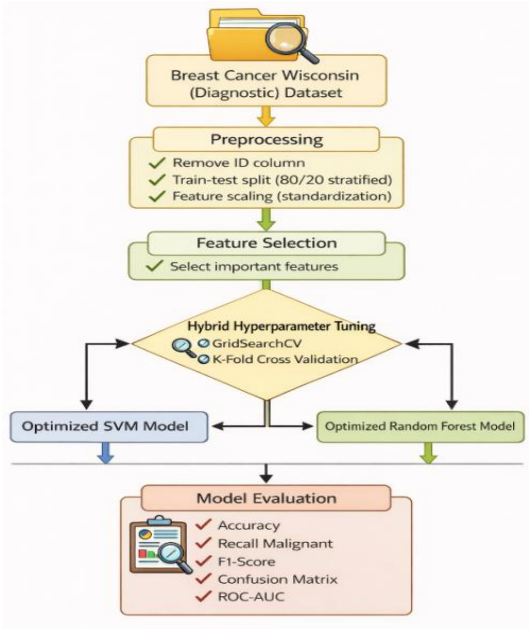

Figure 1. Research Methodology Flow Diagram

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

This study adopts a quantitative research approach using an experimental method to evaluate the performance of machine learning algorithms in breast cancer classification. The research methodology is systematically designed to ensure that each stage can be replicated and produces objective and reliable results. The primary focus of this study is to analyze the impact of hyperparameter tuning on the performance of Support Vector Machine (SVM) and Random Forest (RF) models when applied to a tabular breast cancer dataset. The overall stages of the research process are illustrated in Figure 1. which presents the Research Methodology Flow Diagram.

1. Feature Importance (Random Forest)

The dataset is first preprocessed through normalization to ensure that numerical features are scaled appropriately, particularly for SVM, which is sensitive to feature magnitude. Feature selection is then performed using Random Forest feature importance, where the importance score of each feature is computed based on the mean decrease in impurity (MDI). For a given feature f_j, its importance score is defined as:

$$FI(f_j) = \sum_{t=1}^{T} \sum_{n_\varepsilon N_t} \Delta I_{j,t} \quad (1)$$

2. Average Cross-Validation Score

Hyperparameter tuning is conducted using GridSearchCV combined with 5-fold cross-validation. The cross-validation process divides the dataset into five subsets, where four folds are used for training and one fold for validation iteratively. The average cross-validation score is computed as:

$$CV_{avg} = \frac{1}{k}\sum_{i=1}^{k} score_i \quad (2)$$

3. Accuracy

Model evaluation is performed on a separate test dataset that is not involved in training or tuning to ensure unbiased performance estimation. Several evaluation metrics are used due to their clinical relevance. Accuracy is calculated as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

4. Recall (Sensitivity)

Recall (Sensitivity), which is particularly critical in medical diagnosis for detecting malignant cases, is defined as:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

5. True Positive Rate (TPR)

where TPrepresents true positives and FN represents false negatives. In the context of breast cancer diagnosis, high recall for the malignant class is essential to minimize missed cancer cases.

The discrimination capability of the models is further assessed using the Area Under the Receiver Operating Characteristic Curve (ROC–AUC). The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR), where:

$$TPR = \frac{TP}{TP+FN} \quad (5)$$

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

6. False Positive Rate (FPR)

Equation (6) calculates the false positive rate, which measures the proportion of benign cases that are incorrectly classified as malignant. A low false positive rate indicates that the model is effective in avoiding incorrect cancer diagnoses, which is important to prevent unnecessary anxiety and medical procedures for patients.

$$FPR = \frac{FP}{FP+TN} \quad (6)$$

In The AUC value, ranging from 0 to 1, reflects the model's overall ability to distinguish between benign and malignant classes across all classification thresholds. In medical applications, a higher AUC indicates stronger diagnostic discrimination and improved reliability for clinical decision support. By integrating systematic experimentation, hybrid hyperparameter tuning, clinically relevant evaluation metrics, and a clearly defined workflow (Figure 1), this methodology ensures both technical rigor and practical significance for breast cancer diagnostic modeling. general, the proposed research methodology consists of five main stages, namely data acquisition, data preprocessing, machine learning modeling, hyperparameter optimization, and model performance evaluation. The conceptual workflow of the study is illustrated through a methodological flow diagram, as shown in Figure 1, which describes the relationship between each research stage from the initial data collection process to the final evaluation phase.

2.1 Dataset Acquisition

The dataset used in this study is the Breast Cancer Wisconsin (Diagnostic) dataset, which is a benchmark dataset widely used in breast cancer classification research. The dataset consists of 569 samples obtained from Fine Needle Aspirate (FNA) examinations of breast tissue, with two target classes, namely benign and malignant. Each sample is represented by 30 numerical features that describe the morphological characteristics of cell nuclei, including size, texture, shape, and surface complexity. A detailed summary of the dataset characteristics and attributes is presented in Table 1. entitled Description of the Breast Cancer Wisconsin (Diagnostic) Dataset.

Table 1. Description of the Breast Cancer Wisconsin (Diagnostic) Dataset

Parameter	Keterangan
Jumlah Sampel	569
Jumlah Fitur	30
Kelas	Benign, Malignant
Sumber Data	*Fine Needle Aspirate* (FNA)
Jenis Data	Numerik (Tabular)

The use of a benchmark dataset in medical classification research aims to ensure the comparability of experimental results with previous studies and to enhance research reproducibility. This dataset was selected because it has a clean data structure, contains no missing values in the main features, and exhibits a relatively balanced class distribution, making it suitable for evaluating the performance of classical machine learning algorithms.

2.2 Data Preprocessing

Data preprocessing is conducted to ensure data quality and to reduce potential bias prior to the model training process. At this stage, attributes that are not relevant to the classification task, such as identification columns and empty columns, are removed from the dataset. Subsequently, the diagnosis labels, which are originally categorical, are converted into numerical form so that they can be processed by machine learning algorithms.

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

The dataset is then separated into feature variables (X) and target labels (y), and divided into training and testing subsets with a proportion of 80% for training data and 20% for testing data. The data splitting process is carried out using stratified sampling to preserve the class distribution in both subsets, as recommended in tabular medical classification studies.

In addition, feature standardization is applied using z-score normalization, particularly to support the performance of algorithms that are sensitive to feature scale, such as Support Vector Machine. This standardization ensures that each feature contributes proportionally during the model learning process.

2.3 Feature Selection

Feature selection is performed to reduce data dimensionality and eliminate redundant or less informative features while preserving important diagnostic information. In this study, feature selection is conducted using a feature importance approach derived from the Random Forest model. Each feature is assigned an importance score based on its contribution to the classification process.

The top 15 features with the highest importance scores are selected and used as input for both SVM and Random Forest models. This selection is based on cumulative importance analysis to retain the most informative attributes while reducing model complexity. This step aims to improve computational efficiency, reduce overfitting risk, and enhance generalization performance without sacrificing classification accuracy.

$$FI(f_j) = \sum_{t=1}^{T} \Delta I_{j,t} \tag{7}$$

This formula is used to calculate the importance score of a feature f_j in the Random Forest model. The feature importance value is obtained by summing the decrease in impurity $\Delta I_{j,t}$ contributed by feature f_j across all nodes t and all decision trees in the forest. A higher feature importance score indicates a stronger contribution of the feature to the classification process. Features with the highest importance scores are selected during the feature selection stage.

2.4 Model Construction and Baseline Experiment

At the modeling stage, two machine learning algorithms are employed, namely Support Vector Machine (SVM) and Random Forest. SVM is selected due to its capability to construct non-linear decision boundaries through kernel functions, making it effective for high-dimensional data. In this study, the Radial Basis Function (RBF) kernel is used because of its flexibility in handling non-linear relationships among features.

Random Forest is used as a comparative model because it is an ensemble learning algorithm that combines multiple decision trees to enhance model stability and robustness against noise. Random Forest is also relatively insensitive to overfitting and can automatically handle feature interactions.

In the initial phase, both models are trained using default hyperparameter configurations to obtain baseline models. These baseline models serve as reference points before the application of feature selection and hybrid hyperparameter tuning.

2.5 Hybrid Hyperparameter Tuning

This optimization approach is considered hybrid because it integrates feature selection and systematic hyperparameter tuning within a unified learning framework. Feature selection reduces dimensionality and removes irrelevant attributes prior to model training, while hyperparameter tuning optimizes model configurations using cross-validated grid search. The combination of these two processes enables more robust and reliable classification performance compared to applying either technique independently.

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

$$CV_{avg} = \frac{1}{k}\sum_{i=1}^{k} score_i \quad (8)$$

This formula computes the average performance score obtained from k-fold cross-validation. The dataset is divided into k subsets, where each subset is used once as validation data while the remaining subsets are used for training. The average cross-validation score provides a more reliable and stable estimate of model performance by reducing the bias caused by a single train–test split. In this study, k = 5.

2.6 Model Evaluation

Model performance evaluation is conducted to assess the ability of the proposed models to classify breast cancer data accurately and consistently. The evaluation process is performed using the testing dataset, which is not involved in the training phase, in order to provide an objective assessment of model generalization capability.

Several evaluation metrics are employed to provide a comprehensive assessment of model performance. Accuracy is used to measure the overall proportion of correct predictions. However, given the medical diagnostic context, this study places particular emphasis on recall for the malignant class, which reflects the model's ability to correctly identify cancer cases.

In addition, the confusion matrix is used to present the distribution of prediction outcomes in terms of true positives, true negatives, false positives, and false negatives. To evaluate the overall discrimination capability of the models, the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) are utilized, illustrating the relationship between true positive rate and false positive rate across different decision thresholds. Visualization of model evaluation results is presented in the form of confusion matrices and ROC curves, while a summary of evaluation metrics is provided through comparative performance tables.

2.7 Comparative Discussion with Previous Studies

Several previous studies on the Breast Cancer Wisconsin (Diagnostic) dataset report high classification accuracy using classical machine learning models such as SVM, Random Forest, K-Nearest Neighbor, and Logistic Regression, typically ranging between 95% and 98%. However, many of these studies primarily emphasize overall accuracy as the main evaluation metric, with limited discussion on malignant-class recall and false negative rates, which are clinically more critical in cancer diagnosis. In contrast to prior works that focus predominantly on accuracy improvement, this study explicitly prioritizes malignant-class recall and false negative reduction as central evaluation criteria. Although the baseline models already achieved 97.36% accuracy, the recall value of 0.928571 indicates that approximately 7.14% of malignant cases were misclassified. Through hybrid hyperparameter tuning and feature selection, the optimized SVM improved recall to 0.952381, reducing the proportion of undetected malignant cases. While the numerical improvement (+2.38%) may appear modest, in a clinical context this represents a meaningful reduction in diagnostic risk. Compared to studies that report marginal accuracy gains without analyzing error distribution, this research provides a structured evaluation using confusion matrices and ROC–AUC analysis to assess discrimination capability across decision thresholds. The optimized SVM achieved an AUC of 0.996032, which is slightly higher than many previously reported SVM implementations on the same dataset. More importantly, the improvement is validated using cross-validation during hyperparameter tuning, reducing the likelihood of overfitting—an aspect not always clearly addressed in earlier studies. Another distinction of this study lies in the comparative behavior between SVM and Random Forest after optimization. While some previous works report substantial improvements for ensemble methods after tuning, the results here show that Random Forest performance remains relatively stable (accuracy 0.973684 before and after optimization). This suggests that for well-structured tabular medical datasets, margin-based classifiers

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

like SVM may benefit more from systematic hyperparameter adjustment than ensemble-based methods that already exhibit strong baseline robustness.

3. RESULT

This section presents the experimental results and discussion regarding the performance of the Support Vector Machine (SVM) and Random Forest (RF) models for breast cancer classification using the Breast Cancer Wisconsin (Diagnostic) dataset. All results are obtained from evaluations conducted on the test dataset, which was not involved in the training process, thereby reflecting the generalization capability of the models. The analysis focuses on baseline model performance, the impact of feature selection and hybrid hyperparameter tuning, classification error distribution, and model discrimination capability. A comprehensive summary of the comparative evaluation results is presented in Table 2. entitled Performance Comparison of Optimized Models.

Table 2. Performance Comparison of Optimized Models

Model	Accuracy	Recall	ROC–AUC
baseline SVM	0.973684	0.928571	0.994709
baseline RF	0.973684	0.928571	0.992890
Optimized SVM	0.982456	0.952381	0.996032
Optimized RF	0.973684	0.928571	0.994378

3.1 Baseline Model Performance

Initial experiments were conducted by training the SVM and Random Forest models using default hyperparameter configurations to establish baseline performance. The purpose of this stage was to obtain an initial reference for model performance prior to applying feature selection and hyperparameter optimization.

The evaluation results indicate that both baseline models achieved relatively high classification accuracy on the test dataset. This suggests that the Breast Cancer Wisconsin (Diagnostic) dataset contains informative features that can be effectively learned by classical machine learning algorithms. However, despite the high accuracy values, further analysis revealed that the baseline models still produced misclassifications, particularly false negative predictions in malignant cases.The presence of false negatives in the baseline models is a critical concern in the context of medical diagnosis, as misclassifying malignant cases as benign may delay appropriate medical treatment. These findings highlight that accuracy alone is insufficient to assess model reliability in breast cancer classification and reinforce the need for further optimization. Both models achieved identical performance Accuracy: 0.973684, Recall (Malignant class): 0.928571, ROC–AUC:SVM: 0.994709, RF: 0.992890. Although the accuracy of 97.36% appears high, the recall value of 0.928571 indicates that approximately 7.14% of malignant cases were not correctly detected. In medical diagnosis, this level of false negatives remains clinically significant and requires improvement.

3.2 Impact of Feature Selection and Hybrid Hyperparameter Tuning

Feature selection and systematic hyperparameter tuning were applied to enhance model performance, stability, and generalization capability. Feature selection reduced data dimensionality by retaining only the most relevant diagnostic features based on importance scores derived from the Random Forest model. This reduction helps eliminate redundant and less informative attributes, decreases computational complexity, and minimizes the risk of overfitting. Meanwhile, hyperparameter tuning was conducted using GridSearchCV with cross-validation to identify the optimal parameter combinations for each algorithm, ensuring that the models achieve balanced bias–variance trade-offs and improved robustness on unseen data.

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

The experimental results demonstrate that the impact of these optimization techniques differs between the two algorithms. The SVM model shows a consistent and measurable improvement across multiple evaluation metrics after optimization. Notably, malignant-class recall and ROC–AUC values increase, indicating enhanced sensitivity in detecting cancer cases and stronger discrimination capability between benign and malignant samples. This suggests that SVM is more sensitive to hyperparameter configuration, particularly in adjusting the regularization strength and kernel parameters that shape the decision boundary. By optimizing these parameters and reducing irrelevant features, the SVM model achieves a more precise separation of classes and reduces false negative predictions.

In contrast, the Random Forest model exhibits relatively stable performance before and after optimization. Although minor improvements are observed, the overall performance gain is not as significant as that of the SVM model. This behavior indicates that Random Forest, as an ensemble-based algorithm, is inherently robust and less sensitive to moderate variations in hyperparameter settings. Its bagging mechanism and aggregation of multiple decision trees already provide strong baseline performance, making additional tuning yield limited incremental benefits. Therefore, while feature selection and tuning contribute to performance refinement, the magnitude of improvement is more pronounced in margin-based classifiers such as SVM than in ensemble methods like Random Forest for this particular dataset. The detailed classification results of the optimized SVM model are illustrated in Figure 2. which presents the confusion matrix of the optimized SVM model.

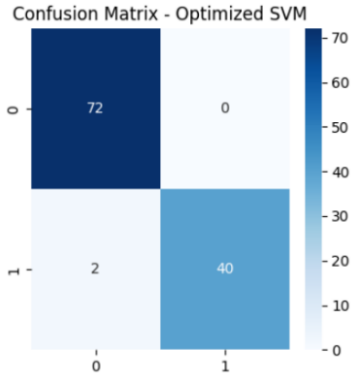

Figure 2. Optimized SVM model confusion matrix

In the optimized SVM model, the confusion matrix shows that most samples are correctly classified. The number of false negatives in the malignant class is very low, indicating that the model has strong sensitivity in detecting breast cancer cases. Additionally, the absence of false positives demonstrates that the model effectively avoids misclassifying benign patients as malignant. For comparison, the classification performance of the optimized Random Forest model is illustrated in Figure 3. which presents the confusion matrix of the optimized Random Forest model.

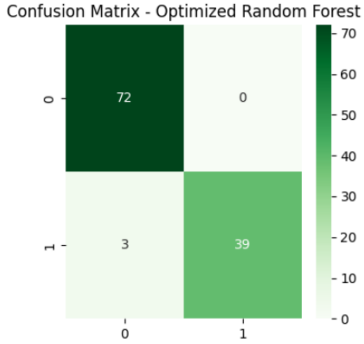

Figure 3. Optimized Random Forest model confusion matrix

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

In the optimized Random Forest model, the confusion matrix indicates that the majority of samples are classified accurately, with most benign and malignant cases correctly identified. Although the number of false negatives is slightly higher compared to the SVM model, the overall sensitivity in detecting malignant cases remains high, demonstrating reliable cancer detection capability. Importantly, no false positive predictions are observed, meaning that benign patients are not incorrectly classified as malignant. This balance between high sensitivity and perfect specificity indicates that the optimized Random Forest model provides stable and clinically acceptable classification performance.

3.3 Classification Error Analysis Using Confusion Matrix

To gain deeper insight into model behavior, classification error distribution was analyzed using confusion matrices. The confusion matrix provides a detailed breakdown of prediction outcomes in terms of true positives, true negatives, false positives, and false negatives.

The optimized SVM model demonstrates a reduction in false negative predictions compared to the baseline model. This result indicates that the optimized SVM is more effective in identifying malignant cases, thereby improving diagnostic sensitivity. Moreover, the number of false positive predictions remains low, suggesting that the model maintains good specificity.

For the optimized Random Forest model, the confusion matrix shows a similar pattern to the baseline configuration, with a slightly higher number of false negatives compared to the optimized SVM. Although the numerical difference is not large, it becomes significant in medical applications where even a small number of undetected cancer cases may have serious clinical implications.

3.4 Model Discrimination Evaluation Using ROC–AUC

The discrimination capability of the models was further evaluated using the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC). The ROC curve illustrates the trade-off between the true positive rate and false positive rate across different classification thresholds. A higher AUC value indicates better overall classification performance and stronger discriminative power in distinguishing between benign and malignant cases. By analyzing the ROC curves of both models, it is possible to compare their ability to maintain high sensitivity while minimizing false alarms across various decision thresholds. In addition to the ROC analysis, the detailed classification distribution of the optimized Random Forest model is presented in Figure 3. which shows the confusion matrix of the optimized Random Forest model.

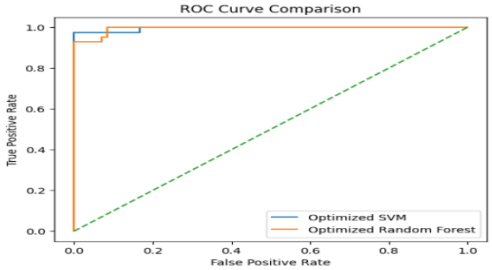

Figure 4. ROC Curve Comparison of SVM and Random Forest models

Both SVM and Random Forest models produce ROC curves that are close to the upper-left corner of the plot, indicating strong discrimination performance. However, the optimized SVM model achieves a slightly higher AUC value than the Random Forest model. This result confirms that the optimized SVM has a more stable and consistent ability to distinguish between benign and malignant classes across various threshold settings. The ROC–AUC analysis supports the findings from the confusion matrix and recall evaluation, reinforcing the conclusion that the optimized SVM model offers superior diagnostic sensitivity.

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

3.5 Comparative Analysis of Baseline and Optimized Models

A comprehensive comparison between baseline and optimized models is summarized through the evaluation metrics obtained from the test dataset. The results indicate that hybrid hyperparameter tuning and feature selection provide a noticeable improvement for the SVM model, particularly in malignant-class recall and AUC values.

On the other hand, the Random Forest model demonstrates relatively stable performance with minimal changes after optimization. This outcome suggests that margin-based classifiers such as SVM benefit more from parameter tuning and feature reduction on tabular datasets with well-separated class boundaries, whereas ensemble-based methods like Random Forest tend to reach near-optimal performance more quickly.

3.6 Overall Findings

From a methodological perspective, the results emphasize the importance of evaluating machine learning models using multiple performance metrics rather than relying solely on accuracy. In breast cancer diagnosis, malignant-class recall and false negative analysis are critical due to their direct impact on patient safety.

Furthermore, this study demonstrates that hybrid hyperparameter tuning and feature selection are essential components in developing reliable machine learning-based diagnostic systems. While optimization does not always yield substantial numerical improvements on benchmark datasets, it plays a crucial role in improving model stability, sensitivity, and trustworthiness in medical applications. The confusion matrix analysis (Figures 2 and 3) strengthens this interpretation by clearly presenting the distribution of true positives, true negatives, false positives, and false negatives. The very low number of false negatives in the optimized SVM model indicates strong effectiveness in detecting malignant cases. This finding is particularly important in medical decision-making, as previous studies emphasize that false negatives are the most critical type of error due to their potential to delay treatment and increase patient mortality risk. Therefore, the improvement in malignant-class recall represents not only statistical enhancement but also a clinically meaningful contribution aligned with established medical AI evaluation principles. Meanwhile, the Random Forest model demonstrates stable classification behavior, confirming its robustness as an ensemble-based approach despite relatively moderate optimization gains.

The ROC curve and AUC analysis further demonstrate the models' discrimination capability across different decision thresholds. A higher ROC–AUC value indicates consistent separation between benign and malignant cases, which is essential in clinical environments where threshold adjustments may be required based on risk tolerance. From a clinical perspective, improvements in recall and ROC–AUC directly support patient safety and early detection, as accurate identification of malignant cases significantly increases survival rates and improves treatment outcomes. Overall, these findings confirm that hybrid hyperparameter tuning and feature selection are not merely technical refinements but clinically relevant strategies that enhance sensitivity, reduce diagnostic risk, and strengthen the reliability of breast cancer classification systems.

4. CONCLUSION

This study demonstrates that the integration of feature selection and hybrid hyperparameter tuning effectively improves the performance of SVM and Random Forest models for breast cancer classification. The optimized SVM model achieves higher performance compared to its baseline configuration, with accuracy improving from 0.973684 to 0.982456, malignant-class recall increasing from 0.928571 to 0.952381, and ROC–AUC rising from 0.994709 to 0.996032. Meanwhile, the Random Forest model maintains stable and robust performance, with accuracy remaining at 0.973684, recall at

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

0.928571, and ROC–AUC slightly improving from 0.992890 to 0.994378 after optimization. These findings confirm that hybrid optimization strategies play a crucial role in enhancing diagnostic reliability, particularly in reducing false negatives and improving model discrimination in machine learning-based medical decision-support systems. Overall, the results highlight that hybrid hyperparameter tuning combined with feature selection is essential for strengthening model reliability, sensitivity, and stability in breast cancer classification tasks. Beyond this specific case study, the proposed hybrid optimization framework offers a systematic and reproducible approach that can potentially be adapted to other medical diagnostic applications involving structured clinical data.

Furthermore, this research contributes significantly to the field of Medical Informatics by demonstrating that hybrid feature selection not only improves predictive performance but also enhances computational efficiency through dimensionality reduction. By removing redundant and less informative features, the proposed framework reduces model complexity, optimizes computational resources, and shortens training time while maintaining high diagnostic accuracy. This balance between efficiency and reliability strengthens the practical applicability of machine learning models in real-world clinical environments, where scalable, fast, and accurate decision-support systems are critically needed.

REFERENCES

[1] A. S. Boddu and J. A. Jan, "A systematic review of machine learning algorithms for breast cancer detection," Tissue and Cell, vol. 95, Aug. 2025, Art. no. 102929. doi: 10.1016/j.tice.2025.102929.

[2] D. Añez et al., "Artificial intelligence pipeline for mammography-based breast cancer detection: An integrated systematic review and large-scale experimental validation," Medicina, vol. 61, no. 12, p. 2237, Dec. 2025. doi: 10.3390/medicina61122237.

[3] H. Chen et al., "Classification prediction of breast cancer based on machine learning," Computational Intelligence and Neuroscience, 2023, Art. no. 6530719. doi: 10.1155/2023/6530719.

[4] H. Qi et al., "Machine learning-based models for prediction of breast cancer recurrence risk," BMC Medical Informatics and Decision Making, vol. 23, 2023. doi: 10.1186/s12911-023-02377-z.

[5] S. Ayanouz et al., "Machine learning algorithms for breast cancer analysis," IAES International Journal of Artificial Intelligence, vol. 13, no. 4, pp. 4372–4379, 2024. doi: 10.11591/ijai.v13.i4.pp4372-4379.

[6] M. A. Elsadig et al., "Breast cancer detection using machine learning approaches," International Journal of Electrical and Computer Engineering, vol. 13, no. 1, pp. 736–745, 2022. doi: 10.11591/ijece.v13i1.pp736-745.

[7] R. Tachicart et al., "Comparative study of machine learning algorithms for breast cancer diagnosis," Journal of Medical Artificial Intelligence, 2025. doi: 10.21037/jmai-24-368.

[8] J. Wang and L. Li, "Hybrid deep learning and machine learning model for breast cancer detection," Computers in Biology and Medicine, vol. 158, 2023, Art. no. 106612. doi: 10.1016/j.compbiomed.2023.106612.

[9] Y. Lee et al., "CNN models for histopathological breast cancer classification," IEEE Access, vol. 11, 2023. doi: 10.1109/ACCESS.2023.3247211.

[10] A. F. Agarap, "On breast cancer detection using machine learning," arXiv preprint, 2017. doi: 10.48550/arXiv.1711.07831.

[11] F. J. Kaunang et al., "Breast cancer detection using decision tree and random forest," Journal of Applied Informatics and Computing, 2025. doi: 10.30871/jaic.v9i2.9073.

[12] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995. doi: 10.1007/BF00994018.

[13] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324.

[14] S. Singh, "Breast cancer prediction using machine learning," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2024. doi:

Vol. 7, No. 3, June 2026, Page. 2778-2789
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5720

10.32628/CSEIT206457.
[15] T. Sun, "Breast cancer prediction based on multiple ML algorithms," Highlights in Science, Engineering and Technology, 2024. doi: 10.54097/0yvhen56.
[16] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY, USA: Springer, 2009.
[17] M. Sugimoto et al., "Machine learning techniques for breast cancer diagnosis," Annals of Breast Surgery, 2023. doi: 10.21037/abs-21-63.
[18] A. Gupta and P. Singh, "XGBoost-based breast cancer classification," Journal of Medical Systems, 2023. doi: 10.1007/s10916-023-01948-0.
[19] B. Jain and N. Singla, "Breast cancer detection using ML algorithms," Journal of Computers, Mechanical and Management, 2023. doi: 10.57159/gadl.jcmm.2.6.230109.
[20] I. Buyung et al., "Effective breast cancer detection using deep learning," Jurnal Ilmu Pengetahuan dan Teknologi Komputer, vol. 8, no. 2, 2024. doi: 10.33480/jitk.v8i2.4077.
[21] G. Singh, "Breast cancer prediction using ML," International Journal of Engineering and Computer Science, vol. 13, no. 3, 2024. doi: 10.18535/ijecs/v13i03.4794.
[22] A. E. Kılıç and M. Karakoyun, "Breast cancer detection using ML algorithms," International Journal of Advanced Natural Sciences and Engineering Researches, 2024. doi: 10.59287/ijanser.401.
[23] H. Chen et al., "Breast cancer dataset modeling study," Computational Intelligence and Neuroscience, 2023. doi: 10.1155/2023/6530719.
[24] E. Khalil et al., "Ensemble machine learning for improved breast cancer prediction," BMC Bioinformatics, 2023. doi: 10.1186/s12859-023-05261-4.
[25] P. Patel and T. Shah, "Feature selection in breast cancer classification," Expert Systems with Applications, 2023. doi: 10.1016/j.eswa.2021.114741.
[26] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.