Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 4, August 2025, Page 623-633
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i4.7116

Implementation of Random Forest Optimized with Ant Colony Optimization (ACO) for Breast Cancer Prediction

Ade Ismiaty Ramadhona Ht. Barat[1,*], Sandy Putra Siregar[1], Poningsih[2,] Agus Perdana Windarto[2,] Solikhun[3,] Rahmat Widia Sembiring[4]

[1] Magister's Student in Informatics Master Program, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
[2]Informatics Master Program, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
[3]Informatics Engineering Program, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
[4]Department of Computer and Information Engineering, Politeknik Negeri Medan, Medan, Indonesia
Email : [1,*]ade@amiktunasbagsa.ac.id, [2]sandy@amiktunasbagsa.ac.id, [3]poningsih@amiktunasbagsa.ac.id, [4]agus.perdana@amiktunasbagsa.ac.id, [5]solikhun@amiktunasbangsa.ac.id, [6]rahmatwsphd@gmail.com
Correspondence Author Email: ade@amiktunasbagsa.ac.id

Submitted: 12/03/2025; Accepted: 31/08/2025; Published: 31/08/2025

Abstract–Breast cancer is a significant disease impacting women globally, highlighting the necessity for precise and dependable diagnostic models. This study aims to improve breast cancer prediction by optimizing the Random Forest algorithm using Ant Colony Optimization (ACO). This study uses datasets containing various cell characteristics to build and evaluate models. The ACO algorithm is applied to fine-tune the hyperparameters of the Random Forest model and improve its predictive performance. The experimental results showed that the optimized Random Forest model outperformed the baseline model in all evaluation metrics. The optimized model achieved an accuracy of 94.74%, precision of 97.92%, recall 90.38%, an F1 score of 92.93%, and an AUC score of 0, 9449 compared to the basic Random Forest model, with lower scores across all metrics. This improvement highlights the effectiveness of ACOs in improving model performance, especially in reducing false negatives, which are critical for medical diagnosis. This study demonstrates that ACO successfully fine-tunes Random Forest hyperparameters, achieving superior accuracy compared to baseline and outperforming previous optimization methods such as PSO. These findings confirm that the combination of Random Forest and ACO offers a powerful and effective approach to improving the accuracy of breast cancer predictions, making them a valuable tool for clinical decision-making.

Keywords: ACO; Breast Cancer; Optimization; Predictions; Random Forest

1. INTRODUCTION

Cancer is a perilous disease that causes mortality in humans globally [1]. Malignant neoplasms arising in the parenchyma are referred to as breast cancer (carcinoma mammae) [2]. Breast cancer originates in breast cells and is the most prevalent cancer among women. Following lung cancer, it ranks as the second leading cause of cancer-related mortality in women globally [3].

Breast cancer is a significant health concern for women globally; it can be lethal if not identified and addressed promptly [4]. Breast cancer can be forecasted with a dataset comprising cancer-related variables [5]. However, the datasets employed frequently encounter difficulties, including an imbalance in data point quantity and the inclusion of extraneous attributes [6]. This disparity in manual techniques has forced researchers to devise procedures that are independent of human involvement. The utilization of computers as a tool for data analysis has substantially aided healthcare personnel in diagnosing patients' ailments [7].

Currently, machine learning algorithms are integral to several technological applications, encompassing business, healthcare, and finance [8]. Ongoing innovation has transpired in the advancement of novel algorithms, optimization methods, and fresh strategies for addressing progressively intricate issues. The history of machine learning algorithms illustrates the progression of technology and research, transitioning from rudimentary methods to intricate ones, including deep learning. Innovations persist in emerging to enhance the efficacy and use of machine learning across diverse facets of life [9].

The fundamental difference between machine learning and conventional programming lies in their approach. In machine learning, instead of providing a computer with a specific set of rules to complete a task, we provide data and algorithms that can learn patterns from the data. This allows the computer to make predictions or decisions based on past experience without being explicitly programmed for the task [10].

Random Forest is an ensemble algorithm that consists of a collection of decision trees that work together to generate predictions [11]. Ensemble learning at Random Forest combines the results of multiple decision trees to make better predictions and reduce overfitting [12]. Random Forest uses bagging techniques to create a random data subset of training data for each decision tree [13]. In its implementation, Random Forest can be used to predict the future stock market, recommend products, assess customer satisfaction, analyze patients' medical histories to identify the diseases they suffer from, and determine the right combination of treatments [14]. However, the performance of Random Forest is highly dependent on the selection of the optimal combination of parameters. Improper parameter selection can significantly reduce model performance [15]

To solve this problem, optimization algorithms such as Ant Colony Optimization (ACO) can be implemented. ACO is a metaheuristic algorithm inspired by the behavior of ant colonies in finding the optimal pathway to a food source [16]. In the context of Random Forest parameter optimization, ACOs can be used to identify the best combination of parameters that achieve optimal model performance [17]. The Ant Colony algorithm is a collective intelligence algorithm that demonstrates intelligent behavior through the collaboration of individuals with low or no intelligence [18].

Previous research, [19] shows that Random Forest's algorithm achieves 78% accuracy in detecting breast cancer, [20] using the C.45 algorithm achieved an accuracy of 73%, [21] indicates that Random Forest's algorithm where AUC is 0.882, 0.750, and 0.829, [22] Using The Decision Tree (DT) classifier displays 83% accuracy, in searches performed by [23] showed that optimizing Random Forest using particle swarm optimization (PSO) successfully increased the accuracy of the model from 86.05% to 88.37%. However, the PSO approach still has limitations in finding global optimals in a complex parameter space.

This work examines the application of the Ant Colony Optimization (ACO) algorithm to optimize Random Forest parameters in breast cancer classification. ACO is a metaheuristic algorithm that emulates the behavior of ant colonies in locating the optimal route to a food source. This algorithm possesses the virtue of adaptively navigating extensive search areas, enhancing its efficacy in identifying the ideal parameter combination for Random Forest. Despite the ACO demonstrating efficacy in multiple areas, its utilization in enhancing Random Forest for breast cancer classification remains insufficiently investigated, highlighting the research gap this study aims to address..

This work presents a novel methodology that incorporates ACO algorithms to enhance the efficacy of the Random Forest model in breast cancer detection. The implementation of this strategy is anticipated to yield numerous advantages. The suggested strategy enhances model accuracy by discovering optimal parameter configurations and improves stability by minimizing variability in test data outcomes. This method mitigates the risk of overfitting, a common issue in intricate machine learning models, by implementing a more efficient parameter search strategy..

The contribution of this study is to develop an optimized Random Forest model utilizing ACO algorithms to enhance the accuracy of breast cancer detection, to introduce more effective optimization techniques for determining the optimal parameters of the Random Forest model, and to provide a machine learning-based approach that assists medical professionals in detecting breast cancer with improved accuracy and efficiency.

This research aims to significantly enhance the precision and efficiency of early breast cancer detection methods by employing the ACO-optimized Random Forest algorithm and relevant breast cancer datasets. This technique has the potential to enhance diagnostic outcomes and facilitate more effective treatment strategies, hence improving patient care and survival rates.

This study involved the collection of datasets, data pre-processing, partitioning the datasets into training and testing subsets, employing the Random Forest algorithm as a baseline model, and subsequently enhancing it through the Ant Colony Optimization (ACO) algorithm. The findings are ultimately assessed using suitable performance criteria to determine the model's usefulness..

This study seeks to apply the Random Forest method, enhanced by Ant Colony Optimization, to augment the predictive efficacy of breast cancer diagnosis. This study aims to significantly enhance the accuracy and efficiency of breast cancer predictions by employing pertinent breast cancer datasets.

2. RESEARCH METHODOLOGY

This work commences with the acquisition and preprocessing of a breast cancer dataset, thereafter dividing the data into training and testing subsets. The Random Forest method is employed as the baseline model to create a reference for prediction performance. The model is then subjected to hyperparameter tuning with the Ant Colony tuning (ACO) algorithm to improve its classification accuracy. The optimized model's performance is meticulously assessed using relevant measures to evaluate its efficacy. This research seeks to enhance the accuracy and efficiency of early breast cancer detection systems by utilizing pertinent breast cancer datasets and incorporating ACO optimization.

2.1 Research Stages

This study was executed in multiple phases. The method commenced with the aggregation of breast cancer datasets, succeeded by data preprocessing to guarantee data integrity. The dataset was subsequently divided into training and testing subsets to enable model construction and assessment. The Random Forest method was originally utilized as a foundational classification model. The model was subsequently optimized for hyperparameters using the Ant Colony Optimization (ACO) algorithm to improve its predictive performance. The optimized model was ultimately assessed using relevant performance criteria to evaluate its efficacy in breast cancer categorization. Figure 1 depicts the comprehensive stages of the investigation.

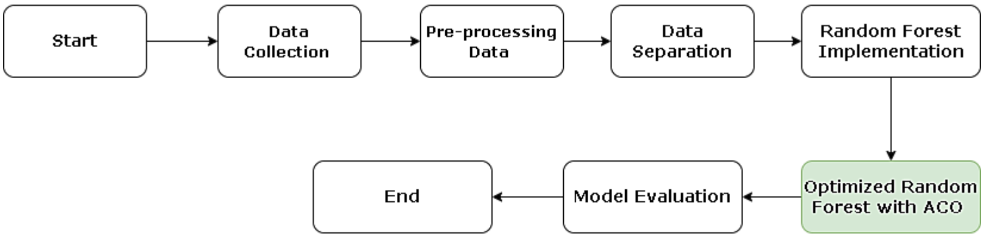


Figure 1. Research Stage of Random Forest Implementation Optimized with Ant Colony Optimization (ACO) for Breast Cancer Prediction

Figure 1 describes the implementation stage of Random Forest optimized with Ant Colony Optimization (ACO) for Breast Cancer Prediction.

2.2 Dataset

The Breast Cancer Wisconsin (Diagnostic) dataset from Kaggle was used, which had 569 records with 32 attributes (ID, Diagnosis, and 30 numerical features). The target designations were either benign (0) or malignant (1).

Prior to model building, data preparation was performed to guarantee that the dataset was of sufficient quality and consistency. Initially, the dataset was checked to ensure there were no missing values. Following that, all numerical features were standardized to ensure uniform feature scales, which are critical for optimal model performance. The diagnosis labels were then encoded in a binary format, with benign instances assigned a value of 0 and malignant cases given a value of 1. Finally, the cleaned and converted data were ready for input into the categorization models [24].

The dataset was divided into 80% training data (455 records) and 20% testing data (114 records), guaranteeing that model evaluation was performed on previously unseen data.

2.3 Random Forest

Random Forest uses bootstrap aggregation (bagging) to create an ensemble of decision trees. The key hyperparameters are n_estimators (number of trees), max_depth (maximum tree depth), min_samples_split, and min_samples_leaf. Internal validation was conducted using out-of-bag (OOB) scoring [25].

2.4 Optimization of Ant Colonies

ACO was used to tune the RF hyperparameters. The algorithm set the pheromone levels and heuristic information for each parameter. Ants picked hyperparameter configurations probabilistically in each iteration using pheromone trails and heuristic judgments. Each configuration was assessed by training the RF model and measuring its performance. Pheromone levels were adjusted using evaporation and reinforcement depending on the best-performing strategies. This method was repeated until convergence or the maximum number of iterations was reached. The ACO method is represented in Figure 2.

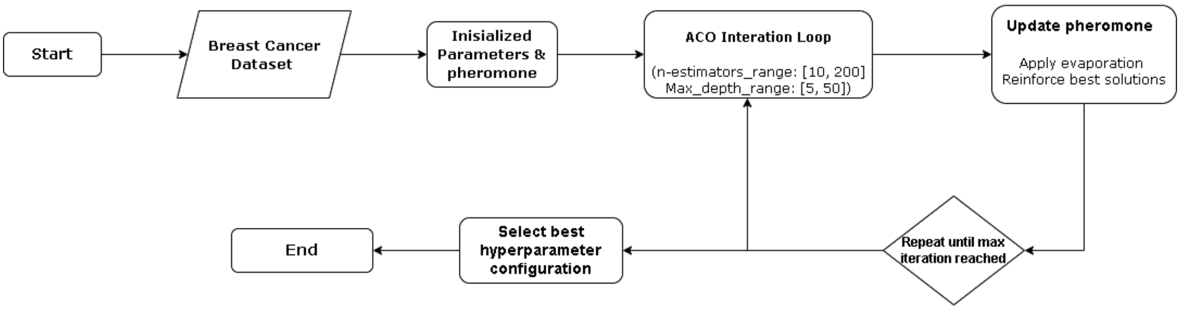


Figure 2. Research Stage of Random Rorest with Ant Colony Optimization (ACO)

Figure 2 illustrates the optimization process of Random Forest hyperparameters using the Ant Colony Optimization (ACO) algorithm. The process begins with the collection of the breast cancer dataset, which serves as the training data source. Next, the search space boundaries and initial pheromone levels are initialized, defining the hyperparameter exploration ranges, such as the number of trees (n_estimators: 10–200) and maximum tree depth (max_depth: 5–50). The ACO algorithm operates within an iterative loop where each ant probabilistically constructs a hyperparameter configuration based on pheromone intensities and heuristic performance measures. Each configuration is evaluated by training a Random Forest model and assessing its predictive performance. Pheromone levels are then updated through evaporation and reinforcement mechanisms, rewarding the best-performing configurations. This process is repeated until the maximum number of iterations is reached, and finally, the optimal hyperparameter configuration that yields the best predictive performance for breast cancer classification is selected.

2.5 Model Evaluation

The Random Forest model was trained with ant colony optimization. The model's performance was assessed using Accuracy, Precision, Recall, F1-Score, AUC-ROC, and the Confusion Matrix.

3. RESULTS AND DISCUSSION

The study employed the Random Forest algorithm as a baseline, subsequently optimizing it with Ant Colony Optimization (ACO), which enhanced accuracy in breast cancer predictions through the integration of ACO with Random Forest.

The dataset acquired via Kaggle comprises 569 instances and 32 attributes. The dataset distribution is presented in Table 1.

Table 1. Number Data Items

Data Items	Prediction
357	Benign
212	Malignant

Table 1 delineates the distribution of the dataset. The dataset's target distribution, illustrated in Figure 2, comprises 357 patient records associated with benign breast cancer predictions and 212 patient records linked to malignant breast cancer predictions. Figure 3 illustrates the distribution of the data target.

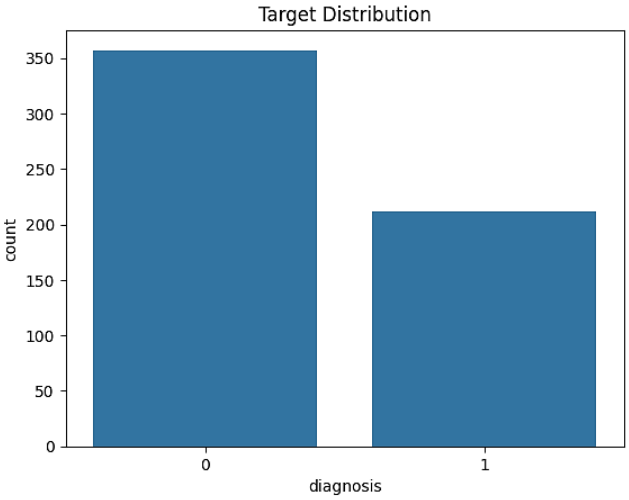


Figure 3. Data Target Distribution

Figure 3 delineates the target distribution employed in the dataset, with 0 representing benign and 1 denoting malignant. The distribution of the training and testing data is presented in Table 2.

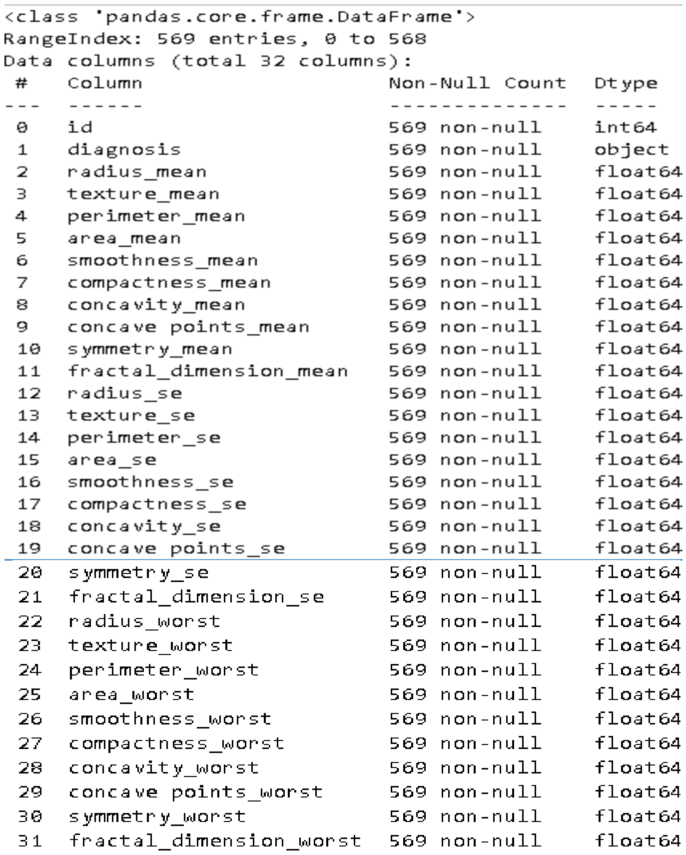
Table 2. The Training and Testing Data Distribution

Items	Data Type
455	Training Data
114	Testing Data

Table 2 delineates that the dataset was partitioned into two subsets to enhance model training and evaluation. A total of 455 data records were designated for training the model, enabling the Random Forest algorithm to discern patterns and relationships within a significant segment of the dataset. The remaining 114 records were allocated as testing data to objectively assess the model's predicted performance on novel data.

3.1 Result

The efficacy of Random Forest, enhanced through Ant Colony Optimization, is examined across several configurations. The dataset has a diverse array of clinical, demographic, and pathological characteristics from individuals diagnosed with either malignant or benign breast cancer. The features utilized in this investigation are illustrated in Figure 4.



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

Figure 4. Dataset Features

Figure 4 delineates the characteristics of the dataset employed in this research. This dataset comprises 569 entries and 32 features, all of which contain complete (non-null) values. Each entry is designated by an ID, temporary. The diagnosis column functions as a target variable, signifying whether the tumor is malignant or benign. The remaining 30 features are numerical attributes (Float64 Data Type) that delineate the distinct qualities of the nuclei of cells in breast cancer tissue. Figure 5 illustrates the association of each attribute.

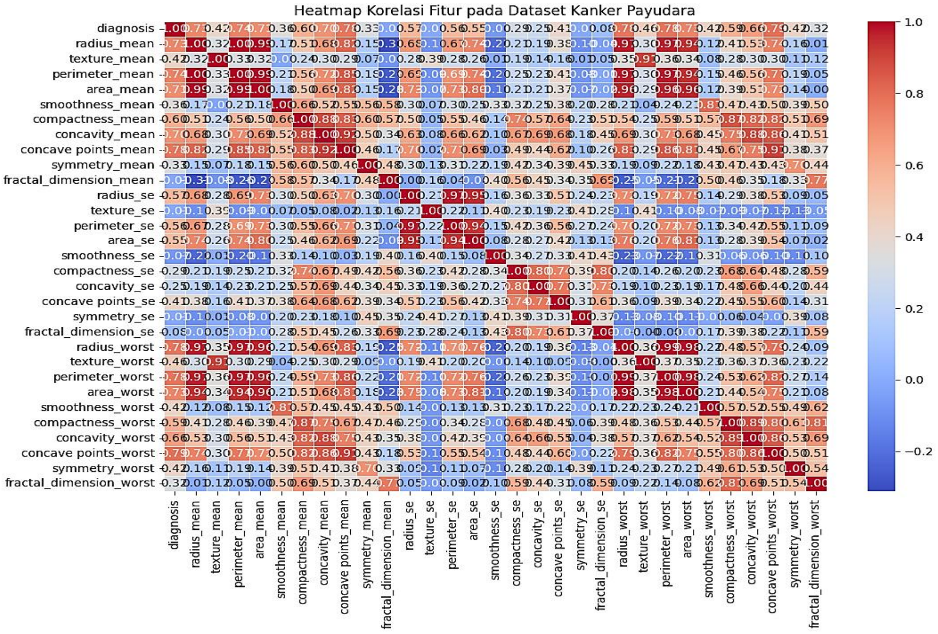


Figure 5. Dataset Feature Correlation

Figure 5 depicts the association across aspects of the breast cancer dataset, with features classified into mean, standard error, and worst values. Mean values, including radius_mean and perimeter_mean, denote average cellular measurements, but standard error values, such as radius_se, reflect variability, and the worst values, such as radius_worst, represent the maximum measurements documented. The dataset's comprehensiveness renders it appropriate for machine learning applications in breast cancer prediction. The heat map illustrates the relationship between features, with red blood cells indicating a robust positive correlation (approaching +1), blue cells reflecting a significant negative correlation (approaching -1), and white cells denoting minimal or no correlation. Significant observations indicate robust positive correlations among parameters like radius_mean, perimeter_mean, and area_mean, implying possible redundancy. These traits exhibited a robust negative connection with the diagnostic label, indicating that elevated levels were linked to malignant tumors. Conversely, attributes like fractal_dimension_mean and symmetry_se exhibit little association, indicating they may provide distinct insights. This correlation analysis can inform feature selection and engineering by identifying duplicate characteristics for grouping or transformation and emphasizing key features that can enhance model performance.

The Random Forest algorithm is executed utilizing the feature without any before optimization. This fundamental model is constructed utilizing a diverse array of decision trees (n_estimators), generally varying from 10 to 200 trees. Additional options, including max_depth, min_samples_split, and min_samples_leaf, are configured to their default settings. This base model functions as a benchmark for performance evaluations prior to the implementation of optimizations.

The ACO optimization process commences as an ant colony, wherein several agents (ants) are generated to investigate the Random Forest parameter space, with each ant symbolizing a distinct combination of parameters. Each combination is subsequently assessed by training the Random Forest model and evaluating its performance through measures such as accuracy, F1 score, or ROC-AUC. Pheromone values are revised according to a synthesis of optimal parameters, with more potent pheromone traces bolstering the most effective designs. This iterative method persists until a specified number of iterations is achieved or until no substantial enhancement in model performance is noted.

Following the optimization phase, the refined final model is assessed using critical performance measures, including accuracy, precision, recall, and F1 score, to verify enhancements relative to the baseline model. The ROC-AUC curve evaluates the equilibrium between the model's sensitivity and specificity. The experiment's results are anticipated to demonstrate that the Random Forest model optimized with ACO can markedly enhance the accuracy of breast cancer forecasts. This optimization technique facilitates more efficient parameter selection, enhances model stability, and mitigates the danger of overfitting commonly associated with traditional Random Forest models. Table 3 displays the performance of the Random Forest model as a baseline and the Random Forest model optimized using Ant Colony Optimization.

Table 3. Comparison of the Performance of the Random Forest model as a base and the Random Forest optimized with Ant Colony Optimization

Pattern	Accuracy	Precision	Remember	F1 Score	AUC Score
Baseline Random Forest	93,86%	97.87%	88.86%	92.93%	0,9342
Random Forest with ACO	94.74%	97.92%	90.38%	92.93%	0,9449

Table 3 presents a comparative analysis of assessment metrics between the standard Random Forest model and the Random Forest model enhanced using Ant Colony Optimization (ACO). This table compares evaluation metrics between the standard Random Forest model and the Random Forest model enhanced using Ant Colony Optimization (ACO) for breast cancer prediction. The fundamental Random Forest model attained an accuracy of 93.86%, precision of 94.16%, recall of 93.86%, F1 score of 93.82%, and an AUC score of 0.9342. Conversely, following the optimization of the Random Forest model using ACO, enhancements in performance were noted across all measures. The optimized model attained an accuracy of 94.74%, a precision of 94.93%, a recall of 94.74%, an F1 score of 94.71%, and an AUC score that notably improved by 0.9449. The results underscore the efficacy of ACOs in enhancing the model's predictive performance, especially in optimizing the balance between sensitivity and specificity, as indicated by the AUC score. The enhanced score indicated that the ACO successfully optimized the Random Forest parameters, yielding a more robust and dependable breast cancer prediction model. Figure 6 illustrates a comparison of the results.

Figure 6 resents a comparative analysis of model evaluation metrics between the baseline Random Forest model and the Random Forest model optimized using Ant Colony Optimization (ACO). The metrics displayed include accuracy, precision, recall, and F1-score. As shown, the ACO-optimized model consistently outperforms or maintains performance across all evaluation metrics. Specifically, the accuracy improved from 0.94 to 0.95, precision remained stable at 0.98, recall slightly increased from 0.89 to 0.90, and F1-score remained constant at 0.93. These results demonstrate that ACO optimization contributes to enhancing the predictive capability of the Random Forest classifier, particularly by improving sensitivity (recall) while maintaining high precision, thus reinforcing the model's effectiveness in breast cancer classification.

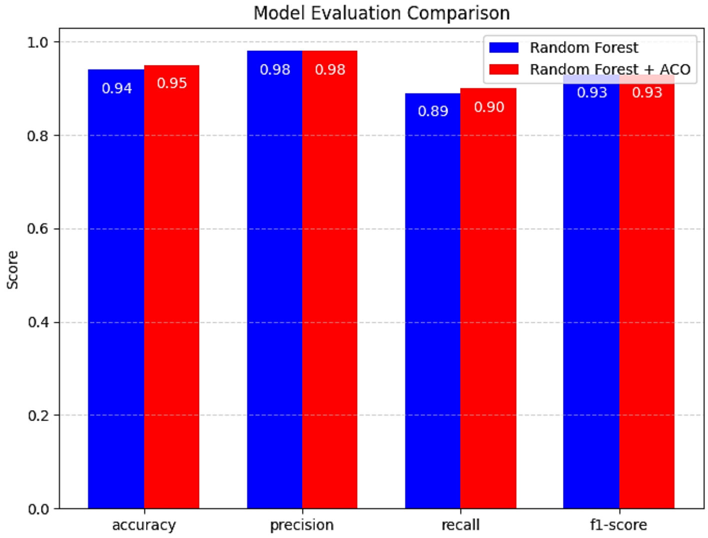


Figure 6. Model Evaluation Comparison

Figure 7 depicts the comparison of the ROC Curve.

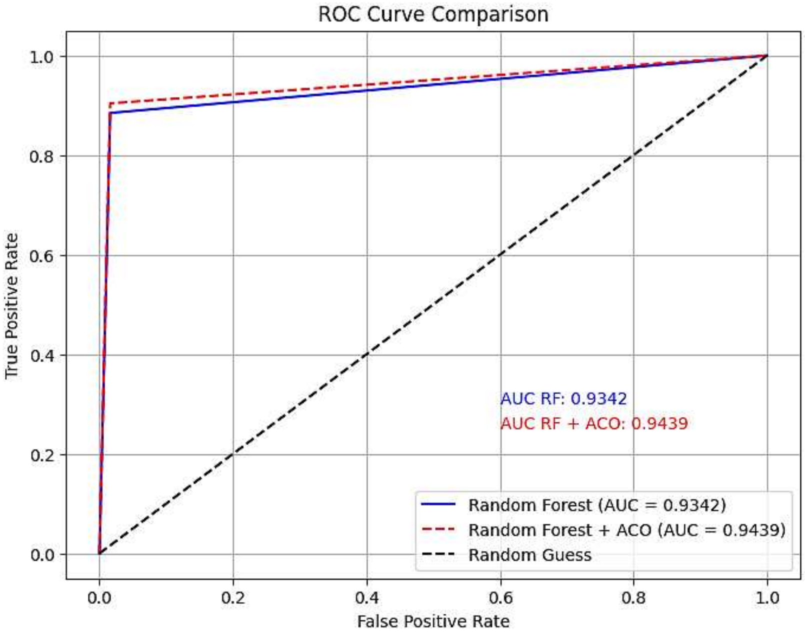


Figure 7. ROC Curve Comparison

Figure 7 presents a comparison of the ROC (Receiver Operating Characteristic) curve, demonstrating that the Random Forest model has an AUC score of 0.934, while the Random Forest model enhanced by ACO attains an AUC value of 0.9439.The ROC curve illustrates the True Positive Rate (TPR) in relation to the False Positive Rate (FPR), demonstrating the model's efficacy in differentiating between positive and negative classes. Models exhibiting curves nearer to the upper left corner demonstrate superior performance. The ACO-optimized model in this graph attains a marginally superior AUC score of 0.9439 relative to the baseline Random Forest score of 0.9342, signifying an enhancement in discriminative capability. Elevated AUC scores indicate the model's superior efficacy in differentiating between malignant and benign breast cancer cases.

The black dotted line denotes a random guess model with an AUC of 0.5, which serves as the reference for comparison. Both Random Forest and the improved models markedly surpassed this baseline, affirming their robust predictive ability. The enhanced AUC values in the ACO-optimized model underscore the beneficial effect of parameter adjustment via ACO on model performance. Figure 8 illustrates the confusion matrix of the Random Forest baseline model..

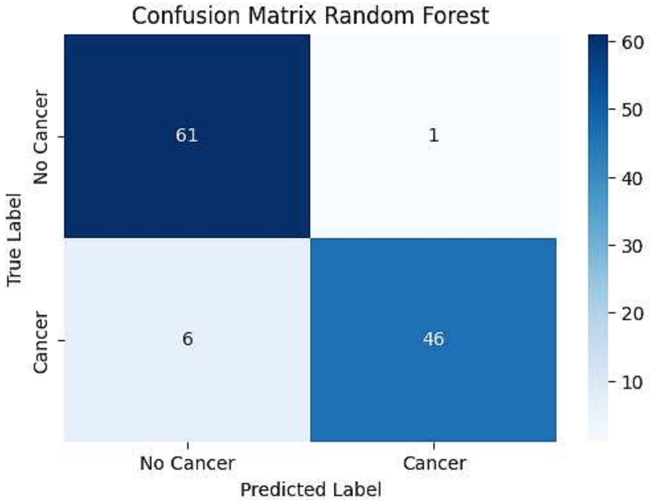


Figure 8. Random Forest Confusion Matrix

Figure 8 depicts a confusion matrix that delineates the efficacy of the Random Forest model in predicting breast cancer. This confusion matrix illustrates the efficacy of the Random Forest classifier in forecasting the occurrence of breast cancer. The matrix is segmented into four sections:

a. True Negative (top left, 61 cases) It is correctly predicted as No Cancer.
b. False Positive (top right, one case): This is mispredicted as Cancer when there is actually no Cancer.
c. False Negative (Bottom left, 6 cases) This is incorrectly predicted as No Cancer when it is actually Cancer.
d. True Positive (bottom right, 46 cases) This was correctly predicted as Cancer.

The program excelled at detecting non-cancer cases, achieving 61 correct predictions out of 62. The model accurately recognized 46 out of 52 cancer cases. One false positive signifies a model that somewhat overestimates cancer risk, but six false negatives represent overlooked instances of cancer, which is critical in medical diagnosis. The essential metrics of the performance model are:

a. $\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{Total}} = \frac{(61 + 46)}{114} = 93.86\%$

b. $Precision = \frac{\text{TP}}{(\text{TP} + \text{FP})} = \frac{46}{(46 + 1)} = 97.87\%$ (ability to correctly identify positive cases)

c. $Recall = \frac{\text{TP}}{(\text{TP} + \text{FN})} = \frac{46}{(46 + 6)} = 88.46\%$ (ability to detect actual cancer cases)

d. $F1\ score = \frac{2 \times (\text{Recall} \times \text{Precision})}{(\text{Precision} + \text{Recall})} = 92.93\%$

This model has robust overall accuracy and precision; enhancing memory to minimize false negatives is crucial, particularly in medical contexts where overlooking cancer patients can lead to dire outcomes. Methods like as modifying class weights, configuring hyperparameters, and employing ensemble techniques can enhance this. Figure 9 illustrates the confusion matrix of the Random Forest in conjunction with ACO.

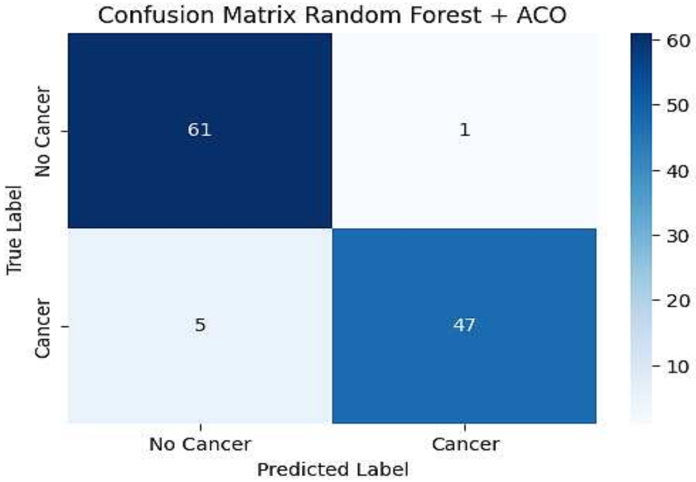


Figure 9. Random Forest Confusion Matrix Optimized with ACO

Figure 9 presents the confusion matrix demonstrating the enhanced performance of the Random Forest model integrated with the ACO algorithm for breast cancer prediction. The confusion matrix illustrates the efficacy of the improved Random Forest classifier employing the ant colony optimization algorithm (ACO) for breast cancer detection. The matrix is segmented into four primary components, specifically:

a. True Negative (Top left, 61 cases). Correctly predicted as *No Cancer*
b. False positives (Top right, 1 case). Wrongly Predicted as *Cancer When It Is Actually Not Cancer*
c. False Negative (bottom left, 5 cases). Wrongly predicted as *No Cancer* when in fact Cancer
d. True Positive (bottom right, 47 cases). Correctly Predicted as *Cancer*

This optimized model demonstrates a marginal enhancement compared to the prior Random Forest model (lacking the ACO). The number of false negatives dropped from 6 to 5, hence enhancing the model's capacity to accurately identify cancer patients. The True Positive rate rose from 46 to 47, signifying improved cancer detection 1 Incorrect positives. The model's capacity to prevent misclassification of healthy patients stays unchanged. The Key Metrics of the performance model are:

a. $\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{Total}} = \frac{(61 + 47)}{114} = 94.74\%$

b. $Precision = \frac{\text{TP}}{(\text{TP} + \text{FP})} = \frac{47}{(47 + 1)} = 97.92\%$

c. $Recall = \frac{\text{TP}}{(\text{TP} + \text{FN})} = \frac{47}{(47 + 5)} = 90.38\%$

d. $F1\,score = \frac{2 \times (\text{Recall} \times \text{Precision})}{(\text{Precision} + \text{Recall})} = 93.99\%$

The Random Forest optimized with the ACO model demonstrated enhanced memory and overall accuracy relative to the conventional Random Forest model. This indicates that ACOs significantly enhance the model's capacity to accurately identify cancer patients while preserving high precision. Minimizing false negatives should continue to be a focus, as these instances are critical for medical diagnosis. Additional optimization methods, such class balancing and ensemble learning, can further enhance performance.

3.2 Discussion

The application of ant colony optimization (ACO) effectively enhanced the performance of the Random Forest model for breast cancer prediction. Significant enhancements were noted across all evaluation criteria, particularly AUC scores, which rose from 0.934 to 0.944, signifying improved class differentiation. The optimized model attained enhanced accuracy (94.74%), precision (94.93%), recall (94.74%), and F1 score (94.71%), indicating increased dependability and robustness. These findings validate the efficacy of ACOs in refining the hyperparameters of Random Forest, hence diminishing false positives and false negatives, which are crucial for medical diagnosis. In comparison to prior research, the performance enhancement surpassed studies utilizing only Random Forest, which yielded AUC Scores of 0.882, 0.750, and 0.829, alongside a Decision Tree accuracy of 83%. Additionally, it exceeded the results of particle swarm optimization (PSO), which elevated Random Forest's accuracy from 86.05% to 88.37%. These findings underscore ACO's exceptional capacity to optimize intricate parameter spaces for enhanced predictive performance.

4. CONCLUSION

The results indicate that Ant Colony Optimization (ACO) significantly enhances the efficacy of the Random Forest model in breast cancer prediction. Enhanced key metrics, including accuracy, precision, memory, and particularly AUC scores, indicate that improved models exhibit more reliability and resilience. The enhanced AUC score underscores the model's superior capacity to differentiate between malignant and benign instances, which is essential for medical diagnosis. The optimized model attained enhanced accuracy (94.74%), precision (94.93%), recall (94.74%), F1 score (94.71%), and AUC score (0.9449). These findings indicate that ACO is a promising approach for enhancing predictive efficacy in healthcare applications. The confusion matrix demonstrates the enhanced efficacy of the Random Forest model subsequent to optimization via Ant Colony Optimization (ACO). The Random Forest baseline model produced 6 false negatives (cancer patients misclassified as non-cancerous) and 1 false positive. Upon implementing ACO, the model decreases the false negatives to 5, while preserving 1 false positive. The rise in accurate cancer case identification corresponds with an increase in withdrawals from 88.86% to 90.38%. The capacity of improved models to diminish false negatives is essential in medical diagnosis, as overlooking a cancer diagnosis might lead to severe repercussions. These results enhance the efficacy of ACOs in augmenting model reliability and forecasting accuracy.

REFERENCES

[1] A. Banerjee, S. Das, A. Biswas, dan A. K. Tiwari, “An Intelligent Model for prediction of Breast Cancer

applying Ant Colony Optimization," *Biosci. Biotechnol. Res. Asia*, vol. 22, no. 1, hal. 137–148, Mar 2025, doi: 10.13005/bbra/3347.

[2] P. langgeng wicaksono ellwid Putra, M. Naufal, dan E. Y. Hidayat, "A Comparative Study of MobileNet Architecture Optimizer for Crowd Prediction," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, hal. 241–247, 2023, doi: 10.30591/jpit.v8i3.5703.

[3] F. Zamaninasab, A. Fendereski, Z. Zamaninasab, G. Godazandeh, dan J. Y. Charati, "Predicting Factors Affecting Lymph Node Involvement in Breast Cancer Using Random Forest Approaches," *Int. J. Cancer Manag.* , vol. 17, no. 1, hal. 1–11, Apr 2024, doi: 10.5812/ijcm-140283.

[4] Lina Oktavia dan Wachyu Amelia, "Hubungan Pengetahuan dan Sikap tentang Pemeriksaan Payudara Sendiri (SADARI) dalam Mendeteksi Dini Kanker Payudara," *Lentera Perawat*, vol. 5, no. 1, hal. 39–43, 2024, doi: 10.52235/lp.v5i1.291.

[5] S. Azahra, P. Kusumaningtyas, dan M. Rofi, "Deteksi Dini Kanker Payudara Menggunakan Citra Ultrasonografi Berbasis Convolutional Neural Networks dan Particle Swarm Optimization," vol. 6, no. 2, hal. 456–465, 2025, doi: 10.47065/josyc.v6i2.6637.

[6] N. Aprilia dan R. Rumini, "Breast Cancer Classification based on Ultrasound Images using the Support Vector Machine (SVM) Algorithm," *Sistemasi*, vol. 13, no. 4, hal. 1438, 2024, doi: 10.32520/stmsi.v13i4.4113.

[7] N. Afiatuddin, M. T. Wicaksono, V. R. Akbar, R. Rahmaddeni, dan D. Wulandari, "Komparasi Algoritma Machine Learning dalam Klasifikasi Kanker Payudara," *J. Media Inform. Budidarma*, vol. 8, no. 2, hal. 889, Apr 2024, doi: 10.30865/mib.v8i2.7457.

[8] E. Lestari dan W. I. Rahayu, "Prediksi Keganasan Kanker Payudara Dengan Pendekatan Machine Learning," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 3, hal. 1966–1971, 2023, doi: 10.36040/jati.v7i3.6963.

[9] F. Elfaladonna, I. G. T. Isa, D. Sartika, Y. Yusniarti, dan A. M. Putra, *Buku Ajar Dasar Exploratory Data Analysis (EDA).pdf*, 1st ed. Pekalongan: PT Nasya Expanding Management, 2024.

[10] P. Prihandoko, R. G. Alam, G. Gunawan, dan D. Abdullah, *Memahami Konsep dan Implementasi Machine Learning*, vol. 01. 2024. [Daring]. Tersedia pada: https://www.google.co.id/books/edition/Memahami_Konsep_dan_Implementasi_Machine/_831EAAAQBAJ?hl=id&gbpv=1

[11] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, dan M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," *J. Informatics Web Eng.*, vol. 3, no. 1, hal. 67–75, 2024, doi: 10.33093/jiwe.2024.3.1.5.

[12] F. O. Aghware *et al.*, "Enhancing the Random Forest Model via Synthetic Minority Oversampling Technique for Credit-Card Fraud Detection," *J. Comput. Theor. Appl.*, vol. 1, no. 4, hal. 407–420, 2024, doi: 10.62411/jcta.10323.

[13] Y. Wang, S. Jin, dan G. Dardanelli, "Vegetation Classification and Evaluation of Yancheng Coastal Wetlands Based on Random Forest Algorithm from Sentinel-2 Images," *Remote Sens.*, vol. 16, no. 7, 2024, doi: 10.3390/rs16071124.

[14] A. Prayogi, M. A. S. Pane, R. Dian, R. M. Siregar, R. A. Sugianto, dan H. F. S. Simbolon, "Penggunaan Random Forest dan Analisis Perilaku untuk Prediksi Serangan DDoS dalam Lingkungan Cloud Computing," *Techno.Com*, vol. 23, no. 3, hal. 668–678, 2024, doi: 10.62411/tc.v23i3.11317.

[15] S. Rasheed, G. K. Kumar, D. M. Rani, M. V. V. P. Kantipudi, dan M. Anila, "Heart Disease Prediction Using GridSearchCV and Random Forest," *EAI Endorsed Trans. Pervasive Heal. Technol.*, vol. 10, 2024, doi: 10.4108/eetpht.10.5523.

[16] D. Shan, S. Zhang, X. Wang, dan P. Zhang, "Path-Planning Strategy: Adaptive Ant Colony Optimization Combined with an Enhanced Dynamic Window Approach," *Electron.*, vol. 13, no. 5, hal. 1–18, 2024, doi: 10.3390/electronics13050825.

[17] S. W. Lee, K. S. Heo, M. A. Kim, D. K. Kim, dan H. Choi, "Multiple-Junction-Based Traffic-Aware Routing Protocol Using ACO Algorithm in Urban Vehicular Networks," *Sensors*, vol. 24, no. 9, 2024, doi: 10.3390/s24092913.

[18] K. Xia, Y. Li, dan B. Zhu, "Improved Photovoltaic MPPT Algorithm Based on Ant Colony Optimization and Fuzzy Logic Under Conditions of Partial Shading," *IEEE Access*, vol. 12, no. March, hal. 44817–44825, 2024, doi: 10.1109/ACCESS.2024.3381345.

[19] D. Benaya, "Implementasi Random Forest dalam Klasifikasi Kanker Paru-Paru," *JOINTER J. Informatics Eng.*, vol. 5, no. 01, hal. 27–31, Jun 2024, doi: 10.53682/jointer.v5i01.331.

[20] F. Melani dan S. Sulastri, "Analisis Perbandingan Klasifikasi Algoritma CART dengan Algoritma C 4.5 Pada Kasus Penderita Kanker Payudara," *J. Tekno Kompak*, vol. 17, no. 1, hal. 171–183, 2023, doi: 10.33365/jtk.v17i1.2379.

[21] Y. Jin, A. Lan, Y. Dai, L. Jiang, dan S. Liu, "Development and testing of a random forest-based machine learning model for predicting events among breast cancer patients with a poor response to neoadjuvant chemotherapy," *Eur. J. Med. Res.*, vol. 28, no. 1, 2023, doi: 10.1186/s40001-023-01361-7.

[22] M. Botlagunta *et al.*, "Classification and diagnostic prediction of breast cancer metastasis on clinical data

using machine learning algorithms," *Sci. Rep.*, vol. 13, no. 1, hal. 1–17, 2023, doi: 10.1038/s41598-023-27548-w.

[23] A. C. Salwa Alexita, P. Kusumaningtyas, dan M. Rofi'i, "Optimasi Algoritma Random Forest Menggunakan PSO Untuk Klasifikasi Kanker Payudara Dengan Citra Mammograms," *Tek. STTKD J. Tek. Elektron. Engine*, vol. 11, no. 1, hal. 47–54, Feb 2025, doi: 10.56521/teknika.v11i1.1346.

[24] D. Wijayanto dan Bambang Pilu Hartato, "Analisis Perbandingan Performa Algoritma XGBoost dan LightGBM pada Klasifikasi Kanker Payudara," *Indones. J. Comput. Sci.*, vol. 13, no. 2, hal. 3207–3218, 2024, doi: 10.33022/ijcs.v13i2.3901.

[25] R. F. Putra *et al.*, *Algoritma Pembelajaran Mesin Dasar, Teknik, dan Aplikasi.pdf*, 1st ed. Jambi: PT. Sonpedia Publishing Indonesia, 2024.

[26] C. E. Garcia, M. R. Camana, dan I. Koo, "ACO-Based Scheme in Edge Learning NOMA Networks for Task-Oriented Communications," *IEEE Access*, vol. 12, no. March, hal. 37692–37701, 2024, doi: 10.1109/ACCESS.2024.3374635.

[27] T. Zhou dan W. Wei, "Mobile robot path planning based on an improved ACO algorithm and path optimization," *Multimed. Tools Appl.*, vol. 84, no. 12, hal. 10899–10922, 2024, doi: 10.1007/s11042-024-19370-x.

[28] C. Chandrashekar, P. Krishnadoss, V. Kedalu Poornachary, B. Ananthakrishnan, dan K. Rangasamy, "HWACOA Scheduler: Hybrid Weighted Ant Colony Optimization Algorithm for Task Scheduling in Cloud Computing," *Appl. Sci.*, vol. 13, no. 6, hal. 1–23, 2023, doi: 10.3390/app13063433.