Excellent: International Journal of Computational Intelligence and Sustainable Innovation
Vol. 1, No. 1, June 2026
ISSN: XXXX-XXXX, DOI: 10.

COMPARATIVE ANALYSIS OF RANDOM FOREST, SUPPORT VECTOR MACHINE, AND LOGISTIC REGRESSION ALGORITHMS IN BREAST CANCER CLASSIFICATION USING THE WISCONSIN BREAST CANCER DATASET

Devina Zuhra Utami[1], Dolli Akbar Nasution [2], Amelia Putri [3], M. Alfarezi LBN Gaol [4]
[1234]State University of Medan, Medan, Indonesia

Article Info

Article history:

Received June 12st, 2026
Revised June 18th, 2026
Accepted June 22th, 2026

Keywords:

Breast Cancer Classification;
Random Forest;
Support Vector Machine (SVM);
Logistic Regression;
Wisconsin Breast Cancer Dataset (WBCD)

ABSTRACT

Breast cancer is one of the leading causes of cancer death in women worldwide, so early detection is an important factor in improving patient survival. This study aims to compare the performance of Random Forest, Support Vector Machine (SVM), and Logistic Regression algorithms in breast cancer classification using the Wisconsin Breast Cancer Dataset (WBCD). The dataset used consisted of 569 patient medical records with 30 numerical features obtained from fine needle aspiration (FNA) examinations. The research methodology includes data collection, data cleaning, label encoding, preprocessing using StandardScaler, and dividing the dataset into 80% training data and 20% test data. Model performance evaluation was carried out using accuracy, precision, recall, F1-score, confusion matrix, and Area Under the Curve (AUC). The results showed that Random Forest and SVM obtained the highest accuracy of 97.37%, while Logistic Regression achieved an accuracy of 96.49%. Random Forest and SVM produce a 100% precision score for the ferocious class, which means there are no false positive predictions. All three algorithms achieved the same recall value of 92.86% for malignant cases, reflecting its good ability to detect breast cancer. In addition, all models obtained an AUC value above 0.99, indicating excellent classification performance. Overall, Random Forest and SVM show higher performance consistency than Logistic Regression, so both can be considered as effective approaches in supporting early detection of breast cancer systems.

***Corresponding Author:**

Devina Zuhra Utami
Email: devinazuhra@gmail.com

1. INTRODUCTION

Breast cancer is one of the most common types of cancer in women. According to WHO data, in 2020 there were around 2.3 million new cases of breast cancer worldwide and it continues to increase every year, so early detection is an important factor in increasing patients' chances of recovery and reducing mortality rates [1]. Early detection of breast cancer is essential to lower mortality rates and increase patients' chances of recovery [2]. Tumors are classified as malignant or benign. To detect malignant cancers, doctors need to use an active determination approach. However, even for specialists, identifying malignancy is very difficult [3]. The development of information technology and the availability of large amounts of medical data encourage the use of machine learning as a diagnostic tool that is able to identify disease patterns quickly and accurately. Therefore, the application of classification algorithms in breast cancer detection is one of the topics that is widely researched in the fields of health and data science [4][5].

One of the datasets that is widely used in breast cancer classification research is the Wisconsin Breast Cancer Dataset (WBCD). This dataset contains various characteristics of the cell nuclei from Fine Needle Aspiration

(FNA) examination which are used to identify benign and malignant tumors. The popularity of the dataset as a research benchmark allows the evaluation of various machine learning algorithms to be carried out objectively and consistently. In addition, the characteristics of the data that are structured and have gone through the validation process make WBCD one of the standard datasets in the development of breast cancer classification models [6].

Research conducted by Abrori and Subhiyakto (2025) showed that Logistic Regression obtained an accuracy of 98%, while Random Forest obtained an accuracy of 96% in the classification of breast cancer. Another study by Desiani et al. (2025) found that Support Vector Machine was able to provide better performance than Logistic Regression in several breast cancer classification scenarios. In addition, parameter optimization in Random Forest and SVM can significantly improve model performance. The results of the study showed that each algorithm has different characteristics and advantages in handling breast cancer data [7].

Although various studies have been conducted before, most have focused only on comparing two algorithms, such as Random Forest with Logistic Regression or Support Vector Machine with Logistic Regression. The study, which compared the three algorithms simultaneously using the Wisconsin Breast Cancer Dataset with uniform evaluation metrics, is still relatively limited. This condition causes no definite conclusion about the algorithm with the most optimal performance in the classification of breast cancer in the dataset.

Based on these problems, this study aims to conduct a comparative analysis of the Random Forest, Support Vector Machine, and Logistic Regression algorithms in breast cancer classification using the Wisconsin Breast Cancer Dataset. The evaluation was carried out using accuracy, precision, recall, and F1-score metrics to determine the performance of each algorithm. The results of the study are expected to provide recommendations on the most effective classification model and become a reference in the development of a decision support system for early detection of breast cancer.

2. METHOD

The research method used in this study uses a comparative analysis approach to evaluate the performance of the algorithm which is widely studied in this medical analysis including Random Forest, Support Vector Machine (SVM), then Logistic Regression [8]. This approach aims to find patterns of health risks based on the extraction of patient clinical data so that it can make a real contribution to the development of decision support systems for the early diagnosis of breast cancer [9]. This research began by searching for datasets through Kaggle, which then goes through the data cleaning stage to ensure the quality and accuracy of the analysis. Furthermore, data labeling and initial exploration were carried out to understand the patterns and distribution in the dataset. Pre-processing Data is also applied, such as normalization or handling of lost values, to improve the quality of the model's inputs [10].

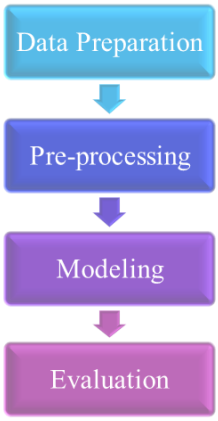


Figure 1. Research Flow

The research stage includes the separation of the dataset into training data and test data with a ratio of 80:20. A portion of 80% is used to train the model, while the remaining 20% is used to test the model's performance on data that has never been seen during the training process. The model performance assessment was carried out through accuracy calculations, confusion matrix, and classification reports to obtain a comprehensive comparison of the three algorithms used. Logistic Regression performs well in handling linear relationships between variables and providing clear interpretation of results, while Random Forest excels at capturing non-linear patterns and working effectively with complex data [11]. A study on breast cancer datasets showed that a combination of confusion matrix-based evaluations and other metrics, such as accuracy and classification reports, was helpful in determining the most accurate model [12].

ISSN: XXXX-XXXX

2.1 Load Dataset

Data collection is the process of collecting raw data from various sources for use in applications machine learning [13]. The systematic stages of the research began with the process of collecting data on medical records of breast cancer patients, both from regional health facilities [14]. The first stage is data collection and loading. Most standard research relies on international-scale public datasets sourced from the platform Kaggle or UCI Machine Learning Repository, namely Wisconsin Breast Cancer Dataset (WBCD/WDBC) [15]. This dataset contains dozens of essential continuous numerical parameters, ranging from cell radius dimensions, area area, circumference (perimeter), texture variation, fineness of curves (smoothness), to the fractal dimension point, along with a special column that records the target status of the patient's diagnosis identification target [16].

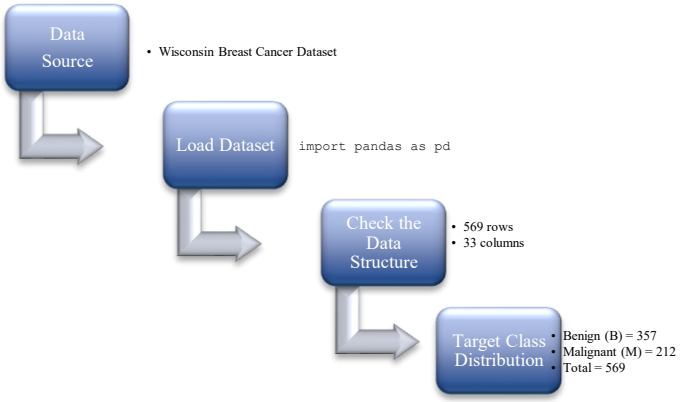


Figure 2. Data Collection Process

This dataset contains 569 patient data with 33 columns, consisting of 30 numerical features of tumor cell measurement results, 1 diagnostic target column, and 2 irrelevant columns (id and Unnamed:32). The dataset is usually stored in document format and contains hundreds of clinical record samples that are the result of extraction from breast cancer cell visuals [6]. The kanker.csv data file is loaded into the program using the `pd.read_csv()` from the Pandas library. The result is a DataFrame containing 569 rows and 33 columns that are ready to be analyzed. The dataset has 33 columns divided into 30 numerical features of cell measurements (with suffixes _mean, _se, and _worst), 1 diagnosis target column (M = Malignant, B = Benign), 1 irrelevant id column, and 1 completely blank Unnamed:32 column.

2.2 Data Cleaning

Once the data is collected, the next step is data processing and cleaning. Raw data often contains a lot of noise, inconsistencies, and imperfections that must be addressed before it can be used to train the model [17]. The data cleansing stage aims to convert the raw data into a clean format that is ready for use by the algorithm machine learning. From the results of the analysis of the data parameter structure, it is often found that the existence of attributes that do not have informative value and a logical correlation to medical classification, such as the sequence number column or patient id, to the finding of fields with absolute empty values (e.g. the Unnamed attribute: 32) [15]. The id attribute is not used in model training because it has no predictive significance to the diagnosis class [18].

ISSN: XXXX-XXXX Doi: 10.XXX

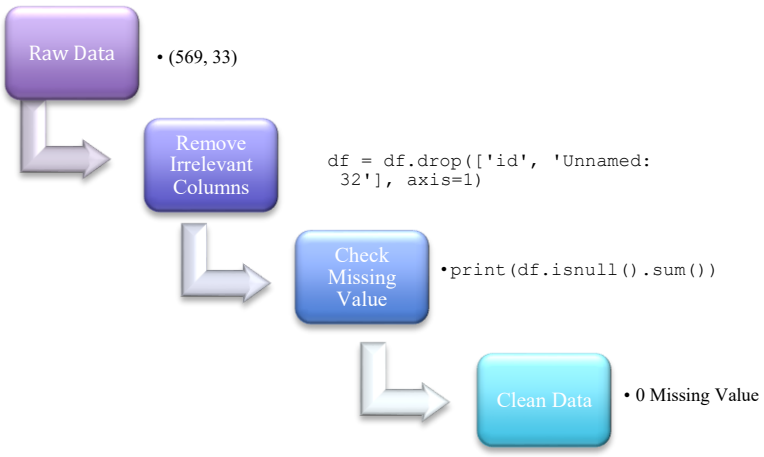


Figure 3. Data Cleansing Process

After deletion, the number of columns is reduced from 33 to 31. After the removal of the column, a blank value is checked using `print(df.isnull().sum())`. The results show that the zero value in all columns of this dataset is relatively clean and does not require an imputation process. This fundamental process is carried out by checking for possible line duplication between patient samples and using programming functions to detect data points that are left blank or missing values [10].

2.3 Data Labeling

Analysis to deepen understanding of feature patterns and the distribution of data structures is described using descriptive statistical methods in the Exploratory Data Analysis (EDA) [9]. Along with the exploration, data labeling is engineered [19]. Selection of the right features is necessary to optimize the algorithm in classifying breast cancer diseases [20]. Patient diagnosis columns that originally contained the letter component string alphabet (such as category 'B' which represents Benign or 'M' for Malignant) is converted and encoded into binary number type formats 0 and 1 [14]. Transformation steps or encoding These target variables are implemented to facilitate the algorithm process of recognizing matrices, accelerate the efficiency of computer classification programs, and prepare ready-to-use target models [8]. This conversion is done using `df['diagnosis'] = df['diagnosis'].map({'M': 1, 'B': 0}) print(df.head()).`

2.4 Pre-Processing Data

Data pre-processing is a useful step to eliminate noise, inconsistencies, and redundancies to achieve high-quality data that improves performance [21]. This process aims to reduce image noise and improve overall image clarity [22]. This set of ready-to-process data is then broken down using a random computational method to build learning stability and the formulation of fair testing functions [19]. The population of the algorithm is divided into the percentage of the training data (training set) as an educational arena for cancer pattern models, and the rest of the allocation is used independently as test data (testing set) for future medical diagnostic validity testing [6]. The data is divided into 80% training data and 20% test data using train_test_split with parameters `stratify=y` to ensure that the proportion of classes remains balanced in both subsets. Finally, normalization is carried out using `StandardScaler`. The end result is `X_train_scaled` and `X_test_scaled` that are ready to be used by the model.

2.5 Algorithm Implementation

Algorithms are at the heart of machine learning. An algorithm is a set of mathematical instructions or procedures used to find patterns in data and build models that can make predictions or decisions. The implementation of the algorithm involves the use of Random Forest, Support Vector Machine, and Logistic Regression to predict breast cancer. The data is divided into 80% training data and 20% test data

ISSN: XXXX-XXXX

using train_test_split. Finally, normalization is done using `StandardScaler`. The goal is to compare the accuracy of the two algorithms in predicting the results of the classification.

2.5.1 Random Forest

Random Forest is one of the popular algorithms in the field of Machine Learning that belongs to the group of ensemble learning, which is an approach that combines several models to improve prediction performance. The main goal of this approach is to make the final prediction results more stable, accurate, and less easily affected by data variations [23]. In evaluation, metrics such as accuracy, precision, recall, and F1-score are used. Random Forest shows superior performance, especially when it comes to handling complex and large datasets [10].

2.5.2 Support Vector Machine

Support Vector Machines (SVM) is a supervised learning method that can be used in regression and classification cases. SVM is an algorithm that works using nonlinear mapping to convert original training data to higher dimensions[24]. Based on the training data, it can be labeled from one or several categories, this algorithm produces a model that can make a prediction, whether the test data belongs to which category [25].

2.5.3 Logistic Regression

Logistic regression is a type of regression that links one or more independent variables (independent variables) with dependent variables in the form of categories, usually 0 and 1 [16]. Model evaluation was conducted using metrics such as accuracy, precision, recall, and F1-score, which provide an important picture of the model's effectiveness in classifying breast cancer data [10].

2.6 Evaluation

The effectiveness of the implemented model (Random Forest, SVM, dan Logistic Regression) is measured thoroughly using a combination of accuracy metrics, confusion matrix, and classification report which includes precision, recall, and F1-score. Through these metrics, performance Logistic Regression is assessed based on its ability to handle linear relationships and clarity of interpretation of results. Meanwhile, Random Forest evaluated for its excellence in capturing non-linear patterns as well as its effectiveness in addressing the complexity of breast cancer clinical data. All of these measurement results are used as the main reference to determine the most accurate and stable classification model to support medical decisions. After the validation process is carried out, proceed to the evaluation stage using confusion matrix, ROC curve, and matrix evaluation [26].

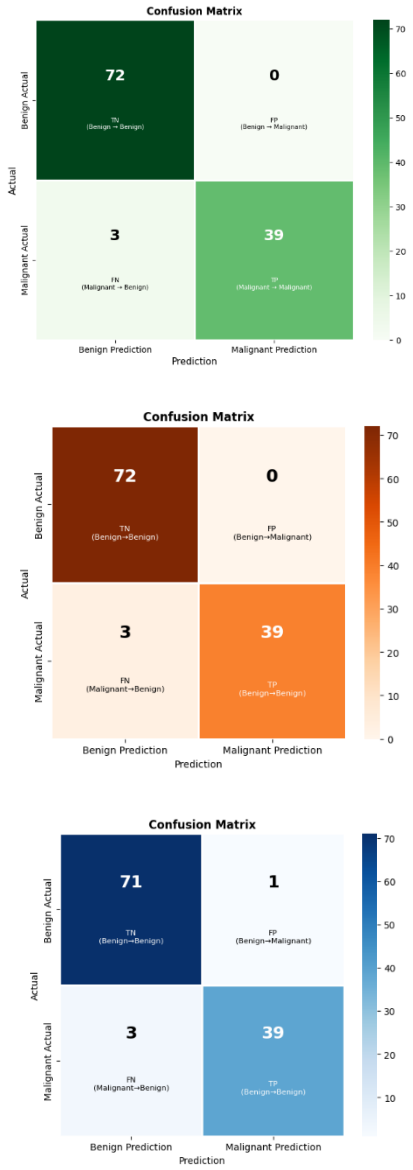
3. RESULTS AND DISCUSSION

This section describes the performance of Random Forest, Support Vector Machine (SVM), and Logistic Regression to predict breast cancer based on accuracy, precision, recall, F1-score, and visualizations such as confusion matrix, matrix evaluation, and ROC Curve. The analysis was carried out to compare the advantages and disadvantages of each algorithm, especially in detecting positive (malignant cancer) and negative (benign cancer) classes. In addition, the discussion included factors that affect the performance of the model, such as the suitability of Logistic Regression with linear data, the ability of SVM to maximize the margin of separation, and the ability of Random Forest to handle non-linear relationships. The practical relevance of the model, the limitations of the research, and the implications of prediction errors are also described.

3.1 Performa Model

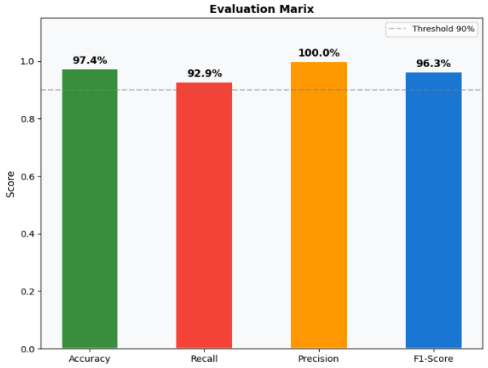
In this sub-chapter, an analysis was carried out on the performance of three classification algorithms based on their ability to identify two main categories, namely B (Benign) and M (Malignant). The evaluation was carried out using data that had been divided into 80% training data as many as 455 data and 20% test data as many as 114 data. Train data is used to build models while test data is used to measure the model's ability to make predictions on data that has never been seen before.

ISSN: XXXX-XXXX Doi: 10.XXX

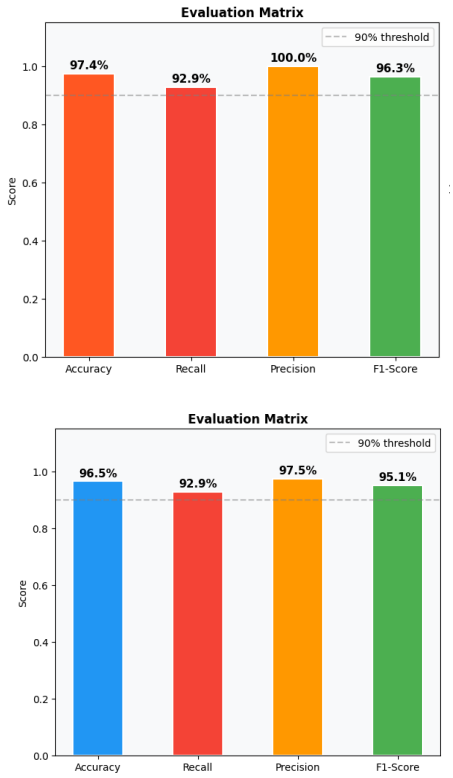


Gambar 4. Confusion matrix Random Forest, SVM, dan Logistic Regression

Based on the results of the confusion matrix, the Logistic Regression algorithm managed to produce 71 true negatives, 39 true positives, 1 false positive, and 3 false negatives. The Support Vector Machine (SVM) algorithm obtained 72 true negatives, 39 true positives, no false positives, and 3 false negatives. The same results were also shown by the Random Forest algorithm , namely, 72 true negatives, 39 true positives, 0 false positives, and 3 false negatives. Based on this comparison, SVM and Random Forest performed better than Logistic Regression in minimizing false positives, where none of the benign patients were diagnosed as malignant.

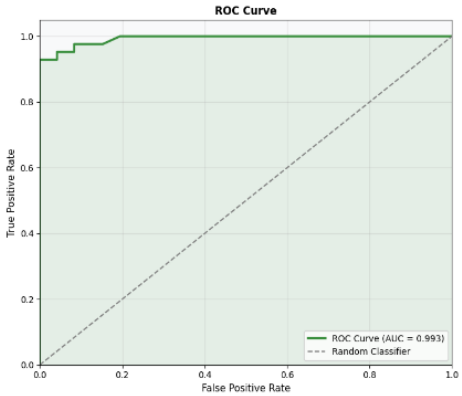


ISSN: XXXX-XXXX

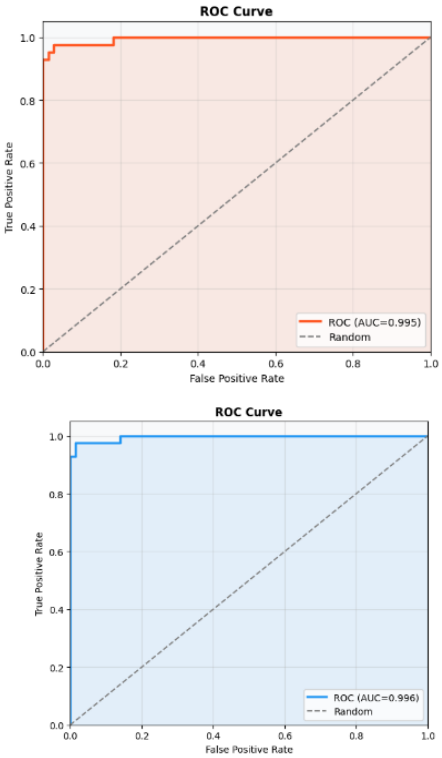


Gambar 5. Matrix evaluation Random Forest, SVM, dan Logistic Regression

Based on the results of the matrix evaluation, the SVM and Random Forest algorithms showed better performance than Logistic Regression on most measurement metrics. Both algorithms obtained a Precision value of 100% in the Malignant class , which indicates that all data predicted as malignant cancer are truly included in the malignant category, so no false positive cases were found. On the other hand, the Recall value for the Malignant class in all three algorithms reached 92.86% which indicates that each model equally missed 3 out of 42 cases of malignant cancer in the testing data. In addition, the results of the 10-fold cross validation showed that Logistic Regression obtained an average Recall of 96.21% with a standard deviation of 3.56%, SVM obtained an average Recall of 96.21% with a standard deviation of 4.66%, and Random Forest obtained an average of Recall of 93.87% with a standard deviation of 5.96%. The relatively small standard deviation values of the three algorithms indicate that the models are stable and consistent across different data divisions, so that there is no overfitting.



ISSN: XXXX-XXXX Doi: 10.XXX



Gambar 6. ROC Curve Random Forest, SVM, dan Logistic Regression

Analysis through the Receiver Operating Characteristic (ROC) curve shows that all three algorithms have very high classification capabilities. The Area Under Curve (AUC) values obtained consecutively were 0.996 for Logistic Regression, 0.995 for SVM, and 0.993 for Random Forest. The AUC value close to 1 indicates that the three models have an excellent ability to distinguish between benign and malignant breast cancer cases. These findings are in line with research by Al Abrori and Subhiyakto (2025), who stated that an AUC value close to 1 indicates a highly accurate classification performance on the Wisconsin Breast Cancer Dataset.

3.2 Comparison of Advantages and Disadvantages of the Model

In this sub-chapter, the advantages and disadvantages of the three algorithms are compared based on the results of performance evaluation in predicting breast cancer.

Model Accuracy Results Table

Metric	Logistic Regression	SVM	Random Forest
Accuracy	96,49%	97,37%	97,37%
Precision (Jinak)	95,95%	96,00%	96,00%
Precision (Desire)	97,50%	100,00%	100,00%
Recall	98,61%	100,00%	100,00%
Recall	92,86%	92,86%	92,86%
F1-Score (Jinak)	97,26%	97,96%	97,96%
F1-Score (Ganas)	95,12%	96,30%	96,30%
AUC	0,996	0,995	0,993

SVM and Random Forest show excellence in detecting the Benign class, as seen from the recall value of 100% and the accuracy of 96%, which means that there is not a single benign patient who is undetected. This advantage in SVM is supported by a maximum margin mechanism that optimally separates the two classes, while in Random Forest it is supported by an ensemble approach that combines the results of 100 decision trees to improve prediction stability.

ISSN: XXXX-XXXX

However, the three algorithms had the same recall value for the Malignant class of 92.86%, meaning that there were 3 cases of malignant cancer that were not detected from 42 malignant data. The notable difference lies in the precision of the Malignant class, where SVM and Random Forest achieved a perfect value of 100% none of the benign patients were wrongly predicted as malignant while the Logistic Regression had a precision of 97.50% with 1 false positive. Overall, SVM and Random Forest show more consistent performance than Logistic Regression, especially in terms of accuracy. The selection of the best algorithm depends on specific needs, such as whether the top priority is sensitivity to positive classes or overall prediction precision.

3.3 Factors Affecting Performance

In this sub-chapter, the factors that affect the performance of the three algorithms in predicting breast cancer are discussed. The characteristics of the relationship between features are the main factor that differentiates the performance of the three algorithms. Logistic Regression assumes a linear relationship between independent variables and class probabilities. In this dataset, which has several features with fairly clear linear separation, such as radius_mean and area_mean, Logistic Regression can work well. However, if there are significant non-linear relationships between features, Logistic Regression may have difficulty modeling the pattern optimally. SVMs with RBF kernels are able to map data to higher dimensions so that they can handle non-linear relationships, as evidenced by the 100% accuracy of the Malignant class. On the other hand, Random Forest uses a combination of decisions from 100 trees so that it captures complex and non-linear relationships between features. The distribution of dataset classes is also an important factor. The dataset shows an imbalance of the mild class, where the Benign class dominates by 62.7% (357 data) while the Malignant class is only 37.3% (212 data). This condition has the potential to affect the sensitivity of the model to minority classes, especially in detecting the Malignant class. To overcome this, the study used the stratify=y parameter in the train-test split process to ensure that the proportion of classes remained balanced in both training and testing data. This is in line with the findings of Al Abrori and Subhiyakto (2025) who stated that class imbalance can cause models to more easily predict the majority class, so special techniques are needed to increase sensitivity to minority classes.

3.4 Implications of Prediction Errors

Prediction errors in the context of breast cancer diagnosis have very different clinical implications between False Positive (FP) and False Negative (FN). False Positive occurs when a benign patient is predicted to be malignant, while False Negative occurs when a patient who is actually malignant is predicted to be benign. From the results of the evaluation, SVM and Random Forest managed to achieve FP = 0, meaning that none of the benign patients were misdiagnosed as malignant. In contrast, Logistic Regression yields 1 false positive. All three algorithms produced FN = 3, which means that there were 3 malignant cancer patients who were not successfully detected and predicted to be benign. This condition is much more dangerous medically because patients do not get the necessary treatment, thus potentially worsening health conditions and reducing the chances of recovery. Therefore, in the context of cancer detection, the Recall value is the most critical evaluation metric compared to other metrics. All three algorithms resulted in identical Recall of 92.86% for the Malignant class, which means that out of every 100 cases of malignant cancer, about 7 cases are potentially undetectable. This needs to be a serious concern if the model is to be implemented in clinical decision support systems.

3.5 Research Limitations

This research has a number of limitations that need to be considered. The use of a single dataset, the Wisconsin Breast Cancer Dataset, makes it difficult to generalize the results to a wider population or to datasets with different characteristics. All three algorithms are tested using default parameters without hyperparameter tuning, so the resulting performance may not have reached the optimal point of each algorithm. However, the results of an evaluation that have reached an Accuracy above 96% and an AUC above 0.993 indicate that the default parameter is good enough for this dataset. The analysis focuses only on quantitative metrics such as accuracy, precision, recall, F1-score, and AUC, without a qualitative approach from a clinical perspective. Although there was a mild class imbalance (62.7% vs 37.3%), this study did not explicitly apply cost-sensitive learning techniques such as class_weight parameters, but only used stratify=y to maintain class proportions. Further research is recommended to use larger and more diverse datasets, explore hyperparameter tuning techniques in more depth, and explicitly apply cost-sensitive learning approaches to further enhance sensitivity to Malignant classes with higher clinical consequences.

4. CONCLUSION

This study successfully conducted a comparative analysis of the Random Forest, Support Vector Machine (SVM), and Logistic Regression algorithms in the classification of breast cancer using the Wisconsin Breast Cancer Dataset. The results of the evaluation showed that the three algorithms had excellent classification performance with an accuracy rate above 96% and an AUC value above 0.99. Random Forest and SVM performed best with an accuracy of 97.37%, while Logistic Regression obtained an accuracy of 96.49%.

Based on the results of the confusion matrix and classification report, Random Forest and SVM were able to achieve 100% accuracy in the malignant class so that it did not produce false positives. However, the three algorithms still produced the same recall value in the malignant class of 92.86%, which indicates that there are still some cases of malignant cancer that are not detected. Overall, SVM and Random Forest showed more consistent performance than Logistic Regression, especially in minimizing prediction errors in cancer diagnosis.

This study shows that machine learning algorithms can be used effectively to support the early detection process of breast cancer. For further research, it is recommended to use a larger and more diverse dataset, hyperparameter tuning on each algorithm, and apply a cost-sensitive learning approach to increase sensitivity to malignant cancer cases and produce more optimal models for clinical implementation.

REFERENCES

[1] O. Of et al., "DAN SUPPORT VECTOR MACHINE," vol. 14, no. 1, pp. 90–101, 2025.
[2] Y. Amethiya, P. Pipariya, S. Patel, and M. Shah, “Comparative analysis of breast cancer detection using machine learning and biosensors,” Intell. Med., vol. 2, no. 2, pp. 69–81, 2022, doi: 10.1016/j.imed.2021.08.004.
[3] M. Moniruzzaman Khan et al., “Machine Learning Based Comparative Analysis for Breast Cancer Prediction,” J. Healthc. Eng., vol. 2022, 2022, doi: 10.1155/2022/4365855.
[4] S. F. Khorshid, A. M. Abdulazeez, and A. B. Sallow, “A Comparative Analysis and Predicting for Breast Cancer Detection Based on Data Mining Models,” Asian J. Res. Comput. Sci., vol. 8, no. 4, pp. 45–59, 2021, doi: 10.9734/ajrcos/2021/v8i430209.
[5] H. Sung et al., “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” CA. Cancer J. Clin., Vol. 71, No. 3, pp. 209–249, 2021, doi: 10.3322/CAAC.21660.
[6] H. Chen, N. Wang, X. Du, K. Mei, Y. Zhou, and G. Cai, “Classification Prediction of Breast Cancer Based on Machine Learning,” Comput. Intell. Neurosci., vol. 2023, no. 1, 2023, doi: 10.1155/2023/6530719.
[7] A. Setiawan, A. A. Siregar, N. Setiawan, J. Nasution, and N. D. Putra, "Performance Optimization of SVM and Random Forest Models for Breast Cancer Classification Using Hyperparameter Tuning," vol. 4, no. 3, pp. 2141–2149, 2026.
[8] Nunung Nurjanah, Arphilia Nur Rani, Hanny Hikmayanti Handayani, and Anis Fitri Nur Masruriyah, "Implementation of Breast Cancer Type Classification Model Using SVM Algorithm and Web-Based Logistic Regression," Ris. and E-Journal Manaj. Inform. Computer., vol. 7, no. 4, pp. 1739–1750, 2023.
[9] Z. Enitisya, R. S. Pratama, P. Y. Faliha, K. Artaliani, R. Della, and M. Ayu, "Prediction of Breast Cancer in Indonesia using Support Vector Machine Algorithm and Logistic Regression," vol. 01, no. 02, pp. 113–121, 2025, doi: 10.30873/jodmapps.v1i2.pp103-121.
[10] Z. Z. Hulaifah Al Abrori and E. R. Subhiyakto, "Comparative Analysis of the Accuracy of Breast Cancer Prediction Using Random Forest Algorithm and Logistic Regression," J. Algorithm., vol. 22, no. 1, pp. 300–311, 2025, doi: 10.33364/algorithm/v.22-1.2164.
[11] A. Raheem, S. Waheed, M. Karim, N. U. Khan, and R. Jawed, “Prediction of major adverse cardiac events in the emergency department using an artificial neural network with a systematic grid search,” Int. J. Emerg. With., vol. 17, no. 1, pp. 1–11, 2024, doi: 10.1186/s12245-023-00573-2.
[12] T. E. Mathew and K. S. Anil Kumar, “A logistic regression based hybrid model for breast cancer classification,” Indian J. Comput. Sci. Eng., vol. 11, no. 6, pp. 899–906, 2020, doi: 10.21817/indjcse/2020/v11i6/201106201.
[13] S. Junaidi et al., Machine Learning Textbook. Jambi: PT. Sonpedia Publishing Indonesia, 2024.
[14] N. Cahyani, R. Irsyada, and A. Y. Kartini, "Implementation of Machine Learning Model as a Breast Cancer Prediction System," Digit. Transform. Technol., vol. 4, no. 2, pp. 1112–1120, 2025, doi: 10.47709/digitech.v4i2.5209.
[15] M. A. Hasan Dalfi, S. Chaabouni, and A. Fakhfakh, “Breast Cancer Detection Using Random Forest Supported by Feature Selection,” Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 2s, pp. 223–238, 2024.
[16] A. D. Achmad, "Classification of Breast Cancer Using the Logistic Regression Method," Jtriste, vol. 9, no. 1, pp. 143–148, 2022.
[17] P. Bintoro, Ratnasari, E. Wihardjo, I. P. Putri, and A. Asari, Introduction to Machine Learning. Solok, West Sumatra: PT Keras Media Literasi Indonesia, 2024.

ISSN: XXXX-XXXX

[18] S. Y. Putri, N. Ramadani, and N. A. Ginting, "EVALUATION OF RECURSIVE FEATURE ELIMINATION FOR THE CLASSIFICATION OF BREAST CANCER USE," vol. 8, no. 01, 2026.

[19] A. Desiani, D. A. Zayanti, I. Ramayanti, F. F. Ramadhan, and G. Giovillando, "Perbandingan Algoritma Support Vector Machine (Svm) Dan Comparison of Support Vector Machine (Svm) and Logistic," J. Artificial Intelligence and Technology. Inf. Vol., vol. 4, no. 1, pp. 33–42, 2025.

[20] E. S. Septiany, H. H. Handayani, T. Al Mudzakir, and A. F. N. Masruriyah, "Optimization of Support Vector Machine Method Using Recursive Feature Elimination and Forward Selection for Breast Cancer Classification," TIN Terap. Inform. Nusant., vol. 5, no. 2, pp. 144–154, 2024, doi: 10.47065/tin.v5i2.5324.

[21] A. Sharma and P. K. Mishra, "Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis," Int. J. Inf. Technol., vol. 14, no. 4, pp. 1949–1960, 2022, doi: 10.1007/s41870-021-00671-5.

[22] S. Sasidharan Nair and M. Subaji, "Automated Identification of Breast Cancer Type Using Novel Multipath Transfer Learning and Ensemble of Classifier," IEEE Access, vol. 12, no. June, pp. 87560–87578, 2024, doi: 10.1109/ACCESS.2024.3415482.

[23] Munaldi, Random Forest and XGBoost Algorithms Using Python. Banyumas: Ganesha Kreasi Universesta, 2025.

[24] Alexsander, Ahmad Nazri, Rio Agus Panbudi, and Junadhi, "Implementation of SVM Algorithm in Predicting Stroke," J. Zetroem, vol. 6, no. 2, pp. 1–5, 2024, doi: 10.36526/ztr.v6i2.3676.

[25] A. Aditya Permana et al., Machine Learning. 2023. [Online]. Available: www.globaleksekutifteknologi.co.id

[26] A. M. Majid and I. Nawangsih, "Comparison of Ensemble Methods to Improve the Accuracy of Machine Learning Algorithms in Predicting Breast Cancer," J. SAINTIKOM (Journal of Management Science. Inform. and Computer), vol. 23, no. 1, p. 97, 2024, doi: 10.53513/jis.v23i1.9563.