# The Impact of Scaling Techniques on Breast Cancer Prediction Algorithms

**[1]Oluwaseyi Ezekiel Olorunshola, [2*]Okeh Dominic Ebuka, [3]Adeniran Kolade Ademuwagun, [4]Fatimah Adamu-Fika and [5]Muhammad Abdullahi Kabir**

[1, 2, 3]Computer Science Department, Air Force Institute of Technology, Kaduna, Nigeria
[4, 5]Cyber Security Department, Air Force Institute of Technology, Kaduna, Nigeria

email: [1]seyisola25@yahoo.com, [2*]okehoroko2019@gmail.com, [3]kademuwagun@gmail.com, [4]f.adamu-fika@afit.edu.ng, [5]abdullahimk82@gmail.com

**Corresponding author*

*Received: 10 December 2024 | Accepted: 24 April 2025 | Early access: 04 May 2025*

**Abstract -** *Breast cancer develops when the genetic material of breast cells undergoes mutations, causing the cells to grow uncontrollably and form tumors. Efforts however have been made to combat it by developing machine learning models to help clinicians with early detection. This study investigates the impact of scaling techniques on the performance of algorithms used for breast cancer prediction. Two scaling approaches were compared with models utilizing the raw, unscaled data. The result revealed that the different scaling techniques had minimal effect on the prediction performance after Hyperparameter tuning. This suggests that for the specific dataset and algorithms used, potential sources of bias were analyzed and the classifiers adapted their internal parameters to compensate for the difference in feature scaling. The model's performance was evaluated using four metrics which are Accuracy, Recall, Precision, and F1-score through the 5-fold cross-validation. The results of this study showed that the Random Forest an ensemble model outperformed all other individual classifier after hyperparameter tuning was performed, it had an Accuracy value of 0.9578, a Recall value of 0.9297, a Precision of 0.9571, and an F1-score of 0.9425.*

**Keywords:** Breast cancer, benign, malignant, hyperparameter tuning, ensemble.

## 1 Introduction

Breast cancer is a type of tumor that occurs in the tissues of the breast. It is the most common type of cancer found in women around the world (Fatima et al., 2020). It is the most recognized global malignancy and the leading cause of cancer deaths. Despite this, undergraduate and postgraduate exposure to breast cancer is limited, impacting the ability of clinicians to accurately recognize, assess, and refer appropriate patients (Katsura et al., 2022). It is also a fact that most breast cancer cases are discovered late (Mahesh et al., 2022). The early symptoms of breast cancer may not be apparent, but it commonly presents itself as a lump in the breast and is usually painless (Huang et al., 2024). It has been discovered that 90% of breast masses are benign, such as fibroadenomas, cysts, and fibrocystic change (WHO, 2021), and although extremely common, breast pain in isolation without other signs is rarely a presentation of breast cancer (Fonseca et al., 2019). It is highly curable when they are diagnosed early before they metastasize (Harbeck et al., 2019). The diagnosis of breast cancer is time consuming due to the limited availability of diagnostic systems such as dynamic MRI, X-rays, etc. (Das et al., 2024), but recent research indicates that early detection of the disease can lead to a positive prognosis and a higher survival rate. The integration of machine learning (ML) algorithms in breast cancer prediction holds promising potential for improving accuracy. The effectiveness of these algorithms can however be significantly impacted by data preprocessing techniques, particularly scaling. This important preprocessing step ensures that all features are on a similar scale, which can potentially enhance model convergence and interpretability.

This research study focuses on investigating the impact of scaling techniques on the performance of breast cancer prediction algorithms. The study involves the analysis of the performance of five different classifiers using two distinct scaling techniques. A comprehensive comparison is made between the scaled and unscaled models to discern the impact of scaling. Additionally, the study further explores the effect of hyperparameterization on the

scaled and unscaled models in order to better understand the relationship between scaling techniques and model performance. Following the initial analysis, a comprehensive examination of the scaled and unscaled models is carried out, integrating hyperparameter tuning to refine the models. As a result of this thorough analysis, a stacking ensemble ML model tailored for breast cancer diagnosis is developed. The ensemble model is constructed using the two best performing classifiers after hyperparameter tuning with the most effective scaling technique identified during the study. The results of this research demonstrated that employing specific feature engineering techniques can significantly enhance the overall performance of the ensemble model while also helping to mitigate potential biases. It is evident that the choice of scaling technique plays a critical role in influencing the model's capacity to capture relevant information from the data and make unbiased predictions across various subgroups. The objective of this research is two-fold: firstly, to analyze ML algorithms that produced the best result and secondly to build a superior ML model by combining the best two analyzed methods using the stacking ensemble method that can predict breast cancer using the Wisconsin Breast Cancer Diagnostic (WBCD) dataset.

The study includes implementing and evaluating different ML algorithms and ultimately creating an ensemble model to enhance prediction accuracy. The dataset used in the study is the WBCD from the University of California Irvine (UCI) repository and was preprocessed, analyzed and prepared for training and testing. The remaining part of this paper is arranged as follows; Section 2 contains a literature review while Section 3 contains the methodology. Results are analyzed and discussed in Section 4 while Section 5 concludes the paper.

# 2 Literature Review

Ahsan et al. (2021) carried out an experiment using data preprocessing steps like feature reduction, data conversion, and data scaling to create a standard dataset. This was important in reducing inaccuracy in final prediction. In the research, eleven ML algorithms which were Logistic Regression (LR), Linear Discriminant Analysis, K-Nearest Neighbors (KNN), Classification and Regression Trees , Naive Bayes (NB), Support Vector Machine (SVM), XGBoost, Random Forest (RF), Gradient Boost, AdaBoost, Extra Tree Classifier were analyzed under six different data scaling methods which were Normalization (NR), Standard scale (SS), MinMax (MM), MaxAbs (MA), Robust Scaler (RS), and Quantile Transformer (QT). The result showed that Classification and Regression Tree, along with RS or QT, outperforms all other ML algorithms with 100% accuracy, 100% precision, 99% recall, and 100% F1 score.

Ambarwari et al. (2020) demonstrated that data scaling techniques like MinMax normalization and standardization significantly impact data analysis. The research utilized machine learning algorithms such as KNN, Naïve Bayes, ANN, and SVM with an RBF kernel. The findings showed that Naïve Bayes maintained the most consistent performance without data scaling, while KNN was more stable than both SVM and ANN. Nevertheless, their computational results indicated that combining MinMax scaling with SVM yielded the best overall performance.

Shahriyari et al. (2019) demonstrated that normalization significantly influences the performance of various machine learning classifiers. Their study involved twelve different ML algorithms, including several commonly used in heart disease prediction, and applied multiple normalization techniques. The results highlighted a strong relationship between the choice of normalization method and the effectiveness of the algorithms. Among the eleven supervised models, SVM achieved the highest accuracy at 78%. However, Naïve Bayes stood out by offering the best overall performance in terms of both accuracy and the shortest training time.

Balabaeva et al. (2020) explored the impact of various scaling techniques on heart failure patient datasets. Their research employed advanced machine learning algorithms including XGBoost, Logistic Regression, Decision Trees, and Random Forest, alongside scaling methods such as Standard Scaler, MinMaxScaler, MaxAbsScaler, RobustScaler, and QuantileTransformer. The results indicated that Random Forest performed better when combined with StandardScaler and RobustScaler. In contrast, the performance of the Decision Tree algorithm remained unaffected by the choice of scaling method.

Singh and Singh (2020) carried out an analysis to investigate the impact of fourteen data normalization methods on classification performance while also considering the full feature set, feature selection, and feature weighting. Also, a modified Ant Lion optimization that searches feature subsets and the best feature weights along with the parameter of Nearest Neighbour Classifier was presented in the research. The Experiments were performed on 21 publicly available real and synthetic datasets, and results were analyzed based on accuracy, the percentage of feature reduced, and runtime. From the results, it was observed that no single method outperformed the others. Therefore, a set of the best and the worst methods combining the normalization procedure and empirical analysis of results was suggested. After this, it was observed that the better performers were the $z$-Score and Pareto Scaling for the full feature set and feature selection, and tanh and its variant for feature weighting. The Mean Centered,

Variable Stability Scaling and Median and Median Absolute Deviation methods along with un-normalized data were the worst performers.

Yang et al. (2021) proposed a prediction model for breast cancer recurrence based on clinical nominal and numeric features. In the study, the data used consisted of 1,061 patients from the Breast Cancer Registry from Shin Kong Wu HoSu Memorial Hospital between 2011 and 2016, in which 37 records were denoted as breast cancer recurrence. The approach used consisted of three stages. First, data pre-processing and feature selection techniques to consolidate the dataset was carried out. Among all features, six features were identified for further processing in the following stages. Next, resampling techniques were applied to resolve the issue of class imbalance. Finally, the construction of the two classifiers, AdaBoost and cost-sensitive learning, to predict the risk of recurrence and carrying out the performance evaluation in three-fold cross-validation. By applying the AdaBoost method, an accuracy of 0.973 and sensitivity of 0.675 was achieved. By combining the AdaBoost and cost-sensitive method of the proposed model, a reasonable accuracy of 0.468 and substantially high sensitivity of 0.947 which guarantee almost no false dismissal was achieved.

Elsadig et al. (2023) selected eight classification algorithms that had been used to predict breast cancer to be under investigation. These classifiers include single and ensemble classifiers. A trusted dataset has been enhanced by applying five different feature selection methods to pick up only weighted features and neglect others. Accordingly, a dataset of only 17 features has been developed, SVM was ranked at the top by obtaining an accuracy of 97.7% with classification errors of 0.029 False Negative (FN), and 0.019 False Positive (FP). Therefore, it was noteworthy that SVM was the best classifier and it outperformed even the stack classier.

Using the WBCD dataset, Strelcenia and Prakoonwit (2023) presented an effective feature engineering method to extract and modify features from data and the effects it has on different classifiers. The feature was used to compare six popular ML models for classification. The models compared were LR, RF, DT, K-NN, MLP, and XGBoost. The results showed that the DT model, when applied to the proposed feature engineering, was the best performing, achieving an average accuracy of 98.64%.

Jaiswal et al. (2023) proposed an improved version of the XGBoost ensemble algorithm called I-XGBoost. The study focused on enhancing identification accuracy through three crucial phases: data pre-treatment, feature extraction, and target role. The performance evaluations used the WBCD and compared the results with various classification techniques, including precision, recall, f1-score, and accuracy, as well as ML algorithms such as SVM, LR, K-NN, NB, DT, RF, AdaBoost, and XGBoost. The results indicated that I-XGBoost achieved an impressive accuracy score of 98.24%, while the LR classifier reached an accuracy score of 97%.

Laghmati et al. (2023) presented a supervised ML Computer Aided Design system for breast cancer classification based on feature selection, PCA, grid search for hyperparameter tuning, and cross-validation. The proposed system draws on seven ML classifiers ANN, K-NN, SVM, DT, RF, XGboost, and Adaboost. Two ensemble models were developed by concatenating the prediction of each ML model using majority voting and stacking with LR S-LR for the final prediction. The performance of the system was evaluated by computing various evaluation metrics, mainly accuracy, specificity, precision, recall, Matthews Correlation Coefficient, Jaccard, and F1-score. Wisconsin and Mass mammography datasets were used. The results indicated that the XGboost model achieved the highest recall of over 96% for the Mammographic Mass dataset, while for the WBCD, both the AdaBoost and the S-LR models outperformed the others with a Recall of 95.35%. The stacking with LR ensemble model obtained the highest accuracies of 93.37% for the Mammographic Mass dataset and 97.37% for the WBCD.

Omondiagbe et al. (2019) investigated SVM (using radial basis kernel), ANNs, and NB using WBCD Dataset to integrate these ML techniques with feature selection/feature extraction methods and compared their performances to identify the most suitable approach. The paper proposed a hybrid approach for breast cancer diagnosis by reducing the high dimensionality of features using linear discriminant analysis (LDA) and then applying the new reduced feature dataset to SVM. The proposed approach obtained an accuracy of 98.82%, a sensitivity of 98.41%, a specificity of 99.07%, and an area under the receiver operating characteristic curve of 0.9994.

Chaurasia and Pal (2021c) developed a stack-based ensemble techniques and feature selection methods for the comprehensive performance of the algorithm and comparative analysis of breast cancer datasets with reduced attributes. In the article, the SVM, K-NN, NB and perceptron were the four ML algorithms combined to make the new model, called blending (stacking). Finally, LR was used to predict the stacked model. It was significant that the sub-models produced different results that were not correlated predictions. The stacking technique was best when all the sub-models were skilfully combined. The article used the five-feature selection technique because it affected the model's overall performance.

## 2.1 Research Gap

While existing studies have extensively explored the effects of different scaling techniques on ML algorithms, there has not been any notable gap in understanding the interaction between hyperparameter tuning and scaling techniques**.** Therefore, in this research, after comparing the different scaling techniques, the various algorithms were hyperparameterized to observe if hyperparameter tuning can mitigate or eliminate the differences in performance caused by different scaling methods, such as MaxAbsScaler and StandardScaler.

# 3 Methodology

The research design, environment, and dataset are described in this Section. And also, the algorithm and performance metrics are also examined. The Python Jupyter Notebook was used to analyze the datasets in order to determine the best performing classifier of the 5 classifiers against the 4 performance metrics for breast cancer prediction. Jupyter Notebook was used for analysis because Python programming language possesses power and flexibility for building and deploying advanced ML models for breast cancer analysis. It also offers advanced techniques, scalability, integration, deployment and sharing.

## 3.1 Research Design

This research utilizes quantitative research methodology to conduct a comparative analysis of various ML algorithms, assessing their performance using specific metrics to predict breast cancer mortality.

## 3.2 Dataset Description

The WBCD dataset used in this study was sourced from the UCI Repository. It consists of clinical and demographic features of breast cancer patients. The dataset comprises features extracted from digitized Fine Needle Aspirate biopsies images. This multivariate dataset consists of 569 instances and 33 features which includes ID number, Diagnosis (M/B), Radius (mean), Texture (mean), Perimeter (mean), Area (mean), Smoothness (mean), Compactness (mean), Concavity (mean), Concave points (mean), Symmetry (mean), Fractal dimension (mean), Radius (standard error), Texture (standard error), Perimeter (standard error), Area (standard error), Smoothness (standard error), Compactness (standard error), Concavity (standard error), Concave points (standard error), Symmetry (standard error), Fractal dimension (standard error), Radius (worst), Texture (worst), Perimeter (worst), Area (worst), Smoothness (worst), Compactness (worst), Concavity (worst), ab. Concave points (worst), ac. Symmetry (worst), ad. Fractal dimension (worst), and it has 0 mismatches and 0 missing values.

Table 1 shows the feature of the dataset.

Table 1: Dataset snippet

| ID | Diagnosis | Radius_Mean | Texture_Mean | Perimeter_Mean | Area_Mean | Smoothness_Mean |
|---|---|---|---|---|---|---|
| 842302 | M | 17.99 | 10.38 | 122.8 | 1001 | 0.1184 |
| 842517 | M | 20.57 | 17.77 | 132.9 | 1326 | 0.08474 |
| 84300903 | M | 19.69 | 21.25 | 130 | 1203 | 0.1096 |
| 84348301 | M | 11.42 | 20.38 | 77.58 | 386.1 | 0.1425 |
| 84358402 | M | 20.29 | 14.34 | 135.1 | 1297 | 0.1003 |
| 843786 | M | 12.45 | 15.7 | 82.57 | 477.1 | 0.1278 |

## 3.3 Data Preprocessing

The data preprocessing phase involves processing the data to identify and remove any unintended and unnamed columns that might exist due to formatting issues, thereby ensuring that the dataset is clean and structured correctly. It is vital for the subsequent steps in the model-building process. After the dataset is loaded and cleaned, the next step is to convert the target variable, 'diagnosis,' from categorical to numerical values. This conversion is

necessary to make the data compatible with ML algorithms. Specifically, the diagnosis labels 'M'(malignant) and 'B' (benign) were mapped to 1 and 0, respectively. Following the conversion, the data was split into features and target variables.

## 3.4 Data Scaling Method

Data normalization is an activity in data preprocessing that changes the attribute value according to a common scale or range to improve the performance of a ML algorithm. There are different types of techniques for data normalization, but this research is only limited to two types. The StandardScaler and the MaxAbsScaler. Applying standard normalization ensures that for each feature, the mean is 0 and the variance is 1, resulting to all features being on the same scale. However, this normalization does not guarantee obtaining any specific minimum and maximum feature values. The StandardScaler was selected because it standardizes features by removing the mean and scaling to unit variance. This is particularly beneficial for SVM and MLP, which are sensitive to features scales and require normalized input for optimal performance. MaxAbsScaler was chosen because it is useful when working with sparse data or when faster, simpler transformation is needed. It scales each feature by its maximum absolute value and keeps the sign of the data, which preserves sparsity and is more computaionally efficient. NB, RF, and DT are relatively scale invariant, meaning they are not significantly afftected by feature magnitudes. However, scaling was still applied for uniformity and to observe whether indirect effects (eg on feature interaction) would influence performance. The use of both scaling techniques allowed for a comparative analysis of their influence across both scale sensitive and scale insensitive algorithms, providing deeper insight into their practical implications in breast cancer prediction.

To optimize the performance of the classification algorithms, hyperparameter tuning was performed using grid search and cross validation. This approach systematically explores a predefined set of hyperparameter values and evaluate model performance using 5-fold cross validation to ensure robustness and to avoid overfitting. For each algorithm, the best combination of hyperparameters was selected based on the higest cross validation score.

The standardScaler normalization is determined by the formula:

$$z = (x - \mu)/\sigma \tag{1}$$

Where:

i. z is the standardized value
ii. $x$ is the original value of the feature
iii. μ is the mean of the feature in the training set
iv. σ is the standard deviation of the feature in the training set.

The MaxAbsScaler scales the data by setting the maximum absolute value of each feature to 1. It analyzes the training data and finds the absolute maximum value for each feature. The MaxAbsScaler is determined by the formula:

$$X_{Scaled} = X/|X\_\max| \tag{2}$$

where:

i. $X_{Scaled}$ is the scaled value
ii. $X$ is the original value of the feature
iii. $|X\_\max|$ is the absolute value of the maximum value of the feature in the training set.

The framework in Figure 1 shows the workflow from dataset input, through normalization techniques, to output produced.

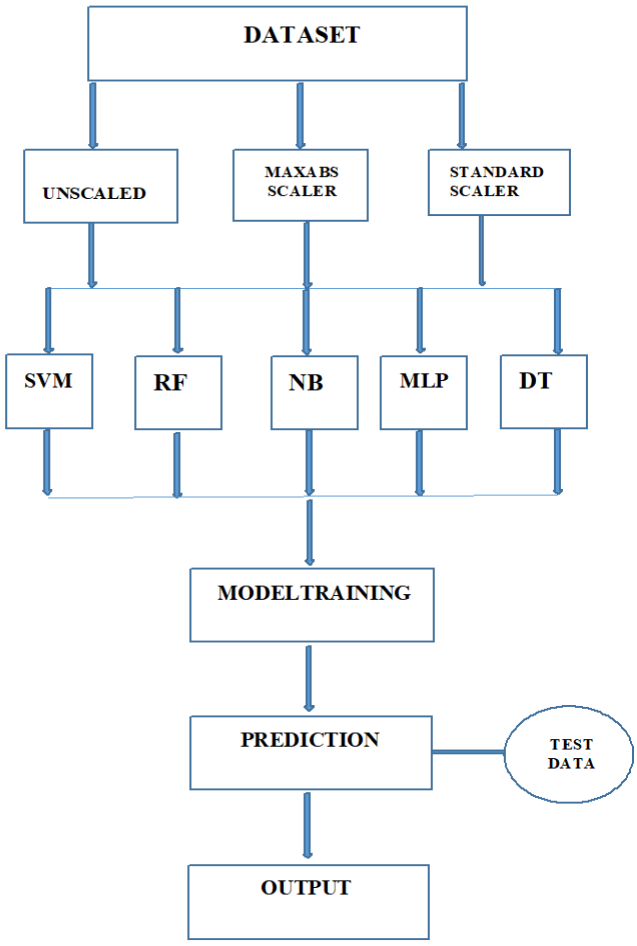


Figure 1: The framework of the proposed Breast Cancer Classification Model.

# 4 Result and Discussion

This Section discusses and analyzes the values of the performance metrics obtained on each classifier after analyzing it on the different scaling techniques. The analysis carried out in this research investigated the influence of scaling techniques on the performance of the algorithms for breast cancer prediction. MaxAbsScaler and StandardScaler were the two different scaling approaches used. Their impact was compared to the performance of the raw, unscaled data. The results revealed that the different scaling techniques produced slight variations in the prediction performance when compared to the unscaled models. This indicates that scaling has some influence on the algorithms. Interestingly, after hyperparameter tuning, all the performance metrics yielded the same results for both scaled and unscaled predictions. This suggests that the hyperparameter tuning process potentially compensated for the lack of explicit scaling. It can also be noted that after hyperprameter tuning, the scaling techniques became minimal because the models adapt their internal parameters to compensate for the difference in feature scaling. This tuning allows models to maintain strong performance across scaled and unscaled datasets, particularly when regularization, learning rates and kernel parameters are optimally adjusted. The specific algorithms used and the distribution of the data itself also contributed to this observation. It is important to note also that, the sensitivity of different algorithms might be inherently more robust to feature scale variations than others. Also, hyperparameter tuning can sometimes adjust internal model parameters in a way that mimic the effect of scaling and this explains why the performance metrics became similar after tuning.

## 4.1 Analysis of Results before Hyperparameter Tuning

After conducting a thorough analysis of the models, it became clear that the MaxAbsScaler demonstrated the most optimal performance across all performance metrics in the SVM model. Following closely behind, the StandardScaler also exhibited strong performance in the SVM model. The RF and DT models delivered identical results across all metrics in both the scaled and unscaled techniques. However, it is worth noting that the NB model produced similar results in the scaled technique, while the unscaled technique resulted in higher values across the metrics. For the MLP model, the StandardScaler achieved the highest score across the metrics, with the

MaxAbsScaler closely trailing behind in performance. The analysis revealed that different algorithms might have varying levels of sensitivity to scaling. Some algorithms might be more robust to feature scales than others. Figure 2 shows the performance of the different scaling technique and the unscaled technique across the different classifiers, and the performance of the classifiers were evaluated based on four different metrics.

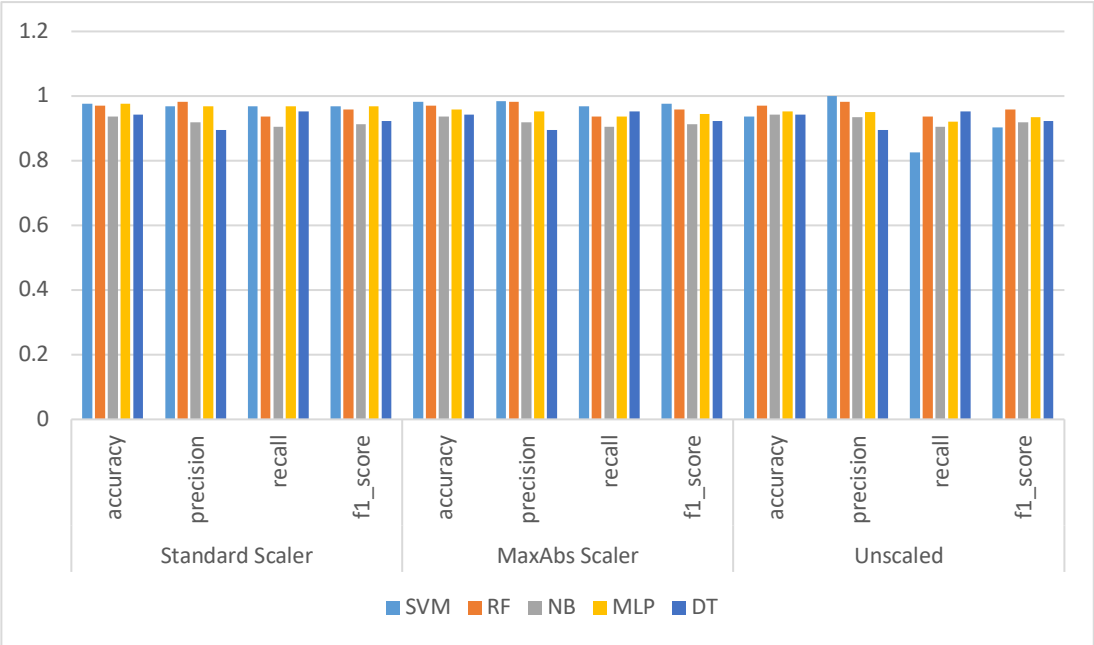


Figure 2: Analysis of the different models before Hyperparameter Tuning

The following Tables show the result of the classifiers after it has been evaluated on the four different metrics parameters. It was observed from the results that the SVM classifier had a value of 1.0 for the precision for the unscaled technique, while the StandardScaler and the MaxAbsScaler had value of 0.968 and 0.984 respectively. The two scaling techniques normalize features, but some of SVM parameters are scale sensitive. The unscaled data is not always at disadvantage when compared with scaled data especially when features are roughly on the same scale, so even though scaling is generally recommended, the result shows that it is not always superior.

Table 2: Result comparison of unscaled algorithms before hyperparameter tuning

| Model | Accuracy | Precision | Recall | F1_score |
|---|---|---|---|---|
| **SVM** | 0.936 | 1 | 0.825 | 0.904 |
| **RF** | 0.971 | 0.983 | 0.937 | 0.959 |
| **NB** | 0.942 | 0.934 | 0.905 | 0.919 |
| **MLP** | 0.953 | 0.951 | 0.921 | 0.935 |
| **DT** | 0.942 | 0.896 | 0.952 | 0.923 |

Table 3: Result comparison of standard scaled algorithms before hyperparameter tuning

| Model | Accuracy | Precision | Recall | F1_score |
|---|---|---|---|---|
| **SVM** | 0.977 | 0.968 | 0.968 | 0.968 |
| **RF** | 0.971 | 0.983 | 0.937 | 0.959 |
| **NB** | 0.936 | 0.919 | 0.905 | 0.912 |
| **MLP** | 0.977 | 0.968 | 0.968 | 0.968 |
| **DT** | 0.942 | 0.896 | 0.952 | 0.923 |

Table 4: Result comparison of MaxAbs scaled algorithms before hyperparameter tuning

| Model | Accuracy | Precision | Recall | F1_score |
|---|---|---|---|---|
| **SVM** | 0.982 | 0.984 | 0.968 | 0.976 |
| **RF** | 0.971 | 0.983 | 0.937 | 0.959 |
| **NB** | 0.936 | 0.919 | 0.905 | 0.912 |
| **MLP** | 0.959 | 0.952 | 0.937 | 0.944 |
| **DT** | 0.942 | 0.896 | 0.952 | 0.923 |

## 4.2 Analysis of Results after Hyperparameter Tuning

Following hyperparameter tuning, the classification models developed using both scaling techniques (MaxAbsScaler and StandardScaler) and the model trained without any scaling produced virtually identical results. This suggests that the tuning process effectively adjusted internal parameters to account for differences in feature scale. The performance metrics for the classifiers were as follows:

SVM had an Accuracy of 0.9508, Precision value of 0.9429, Recall of 0.9249, and F1-score of 0.9333. SVM demonstrated balanced performance across all metrics, showing its effectiveness in separating classes, even without scaled input, once properly tuned. RF had Accuracy of 0.9578, Precision of 0.9571, Recall value of 0.9297, and F1-score of 0.9425. RF outperformed all other models due to its ensemble approach, robustness to scale variation, and strong generalization capability. NB had 0.9385 for Accuracy, 0.9467 Precision, 0.8870 for Recall, and F1-score of 0.9148. NB achieved high precision but lower recall, indicating a tendency to avoid false positives at the cost of some false negatives. MLP had Accuracy of 0.9314, Precision of 0.9205, Recall of 0.8961, and F1-score of 0.9065. MLP performed well overall, though slightly impacted by scale-sensitive behavior, especially in recall. DT had an Accuracy of 0.9403, Precision of 0.9311, Recall of 0.9111, and F1-score of 0.9184. DT showed reliable and interpretable performance with minimal sensitivity to feature scaling. Overall, the similarity in results confirms that hyperparameter tuning reduced the dependency on scaling techniques. RF being an ensemble of multiple DTs, consistently delivered the highest performance due to its robustness and capacity to reduce overfitting. RF being an ensemble model produced the highest results among all the techniques because of its robustness, scale invariance and ability to generalize well. The classifier further benefitted from combining the strength of the multiple DT.

Figure 3 shows a diagrammatical chart representation of the analysis of the different classifiers after the performance of hyperparameter tuning.

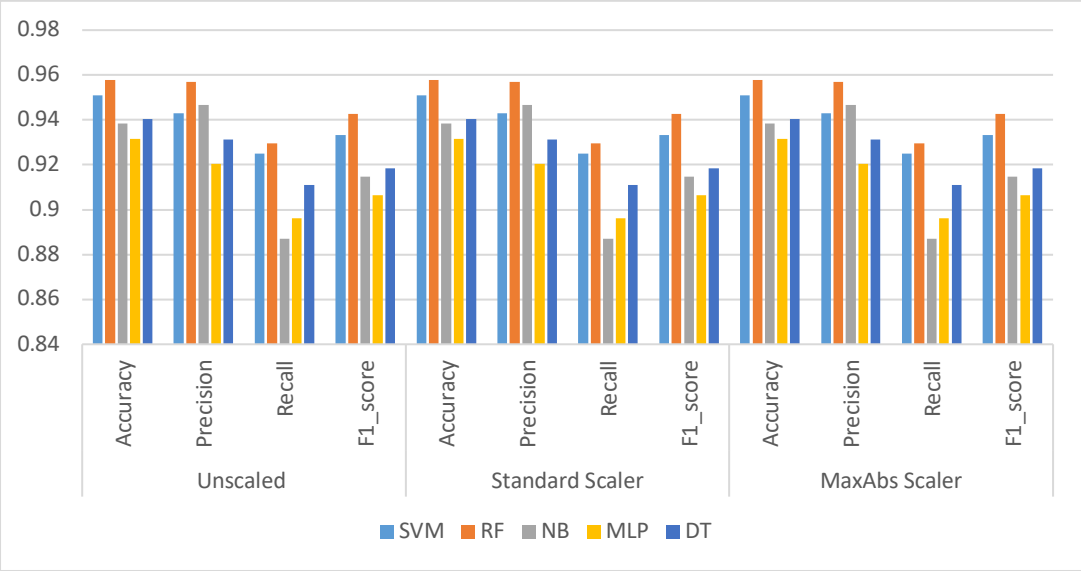


Figure 3: Analysis of the different models after Hyperparameter Tuning

The following Tables show the result of the performance of the different classifiers after hyperparameter tuning, and they are arranged as follows, Unscaled, StandardScaler, and MaxAbsScaler. The results are evaluated based on four metric parameters which are accuracy, precision, recall, and f1_score.

Table 5: Result comparison of unscaled algorithms after hyperparameter tuning

| Model | Accuracy | Precision | Recall | F1_score |
|---|---|---|---|---|
| **SVM** | 0.9508 | 0.9429 | 0.9249 | 0.9333 |
| **R** | 0.9578 | 0.9571 | 0.9297 | 0.9425 |
| **NB** | 0.9385 | 0.9467 | 0.8870 | 0.9148 |
| **MLP** | 0.9314 | 0.9205 | 0.8961 | 0.9065 |
| **DT** | 0.9403 | 0.9311 | 0.9111 | 0.9184 |

Table 6: Result comparison of standard scaled algorithms after hyperparameter tuning

| Model | Accuracy | Precision | Recall | F1_score |
|---|---|---|---|---|
| **SVM** | 0.9508 | 0.9429 | 0.9249 | 0.9333 |
| **RF** | 0.9578 | 0.9571 | 0.9297 | 0.9425 |
| **NB** | 0.9385 | 0.9467 | 0.8870 | 0.9148 |
| **MLP** | 0.9314 | 0.9205 | 0.8961 | 0.9065 |
| **DT** | 0.9403 | 0.9311 | 0.9111 | 0.9184 |

Table 7: Result comparison of MaxAbs Scaled Algorithms after hyperparameter tuning

| Model | Accuracy | Precision | Recall | F1_score |
|---|---|---|---|---|
| **SVM** | 0.9508 | 0.9429 | 0.9249 | 0.9333 |
| **RF** | 0.9578 | 0.9571 | 0.9297 | 0.9425 |
| **NB** | 0.9385 | 0.9467 | 0.8870 | 0.9148 |
| **MLP** | 0.9314 | 0.9205 | 0.8961 | 0.9065 |
| **DT** | 0.9403 | 0.9311 | 0.9111 | 0.9184 |

## 4.3 Comparative Analysis of Scaling and Classifiers Performance

Table 8 presents a comparative analysis of the impact of various scaling techniques on ML algorithms. While some previous studies reported higher performance metrics, the findings of this study demonstrate that scaling had minimal effect after hyperparameter tuning. This suggests that tuning effectively mitigates scale sensitivity across most models. Notably, the classifiers in this research achieved high yet realistic performance scores, indicating reliable model behavior on real-world data. In contrast, the near-perfect results reported in some studies may point to potential issues such as overfitting, data leakage, or insufficient model validation.

Table 8: Performance comparison of scaling techniques used

| References | Techniques Used | Algorithm | Accuracy |
|---|---|---|---|
| Balabaeva et al. (2020) | StandardScaler, MaxAbsScaler, Unscaled, MinMaxScaler, RobustScaler | XG boost | 100% |
| Ahsan et al. (2021) | RobustScaler, and Quantile Transformer | Classification and Regression Tree | 100% |
| Shahriyari et al. (2019) | Normalization (Technique not Specified) | SVM | 78% |
| Proposed method | MaxAbsScaler | SVM before Hyperparameter Tuning | 98% |
| | StandardScaler, MaxAbsScaler, Unscaled | RF after Hyperparameter Tuning | 96% |

# 5 Conclusions

The incidence of breast cancer has caused significant concern due to its aggressive nature and its impact on countless patients, particularly women. Fortunately, recent research and studies have focused on early detection of this global malignancy. The emergence of ML has played a significant role in detecting and identifying the disease at an early stage. This has resulted in the development of ML models that when integrated into systems will help clinicians to detect breast cancer and determine its type. This study highlights the importance of evaluating the impact of data preprocessing techniques like scaling. While scaling might not always lead to substantial performance gain, it has a valuable practice to ensure model robustness and interpretability, especially when dealing with future dataset or algorithms with potentially higher sensitivity to feature scaling. In this research, it was discovered that hyperparameter tuning might have the ability to adjust internal model parameters in a way that compensates for the lack of scaling, thereby making different scaling techniques to have little or no influence on predictions made after hyperparameter tuning. The methodical approach of data cleaning, preprocessing, model training, evaluation, and hyperparameter tuning combined, resulted in a robust and accurate predictive tool. Future work could explore the inclusion of additional features and further optimization of the model to enhance its predictive capabilities.

# References

Ahsan, M., Mahmud, M., Saha, P., Gupta, K., & Siddique, Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. Technologies, 9(3), 52. https://doi.org/10.3390/technologies9030052

Ambarwari, A.; Adrian, Q.J. & Herdiyeni, Y. (2020) Analysis of the Effect of Data Scaling on the Performance of the Machine Learning Algorithm for Plant Identification. *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)* 2020, 4, 117–122.

Balabaeva, K. & Kovalchuk, S. (2019). Comparison of Temporal and Non-Temporal Features Effect on Machine Learning Models Quality and Interpretability for Chronic Heart Failure Patients. Procedia Computer Science. 2019, 156, 87–96.

Chaurasiya, S., & Rajak, R. (2022). Comparative analysis of machine learning algorithms in breast cancer classification. Research Square (Research Square). https://doi.org/10.21203/rs.3.rs-1772158/v1

Das, A. K., Biswas, S. K., Mandal, A., Bhattacharya, A., & Sanyal, S. (2024). Machine Learning based Intelligent System for Breast Cancer Prediction (MLISBCP). *Expert Systems with Applications*, 242, 122673. https://doi.org/10.1016/j.eswa.2023.122673

Elsadig, M. A., Altigani, A., & Elshoush, H. T. (2023). Breast cancer detection using machine learning approaches: a comparative study. *International Journal of Electrical and Computer Engineering*, 13(1), 736. https://doi.org/10.11591/ijece.v13i1.pp736-745

Fatima, N., Liu, L., Hong, S., & Ahmed, H. (2020). Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. IEEE Access, 8, 150360–150376. https://doi.org/10.1109/access.2020.3016715

Fonseca, M. M., Lamb, L. R., Verma, R., Ogunkinle, O., & Seely, J. M. (2019). Breast pain and cancer: should we continue to work-up isolated breast pain? Breast Cancer Research and Treatment, 177(3), 619–627. https://doi.org/10.1007/s10549-019-05354-1

Halim, K. N. A., Jaya, A. S. M., & Fadzil, A. F. A. (2020). Data Pre-Processing Algorithm for Neural Network Binary Classification model in Bank Tele-Marketing. *International Journal of Innovative Technology and Exploring Engineering*, 9(3), 272–277 https://doi.org/10.35940/ijitee.c8472.019320

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., Ruddy, K., Tsang, J., & Cardoso, F. (2019). Breast cancer. Nature Reviews. Disease Primers, 5(1). https://doi.org/10.1038/s41572-019-0111-2

Huang, Y., Zeng, P., & Zhong, C. (2024). Classifying breast cancer subtypes on multi-omics data via sparse canonical correlation analysis and deep learning. *BMC Bioinformatics*, 25(1). https://doi.org/10.1186/s12859-024-05749-y

Jaiswal, V., Suman, P., & Bisen, D. (2023). An improved ensembling techniques for prediction of breast cancer tissues. *Multimedia Tools and Applications*, 83(11), 31975–32000. https://doi.org/10.1007/s11042-023-16949-8

Katsura, C., Ogunmwonyi, I., Kankam, H. K., & Saha, S. (2022). Breast cancer: presentation, investigation and management. *British Journal of Hospital Medicine*, 83(2), 1–7. https://doi.org/10.12968/hmed.2021.0459

Laghmati, S., Hamida, S., Hicham, K., Cherradi, B., & Tmiri, A. (2023). An improved breast cancer disease prediction system using ML and PCA. *Multimedia Tools and Applications*, 83(11), 33785–33821. https://doi.org/10.1007/s11042-023-16874-w

Mahesh, T. R., Kumar, V. V., Vivek, V., Raghunath, K. M. K., & Madhuri, G. S. (2022). Early predictive model for breast cancer classification using blended ensemble learning. *International Journal of System Assurance Engineering and Management*, 15(1), 188–197. https://doi.org/10.1007/s13198-022-01696-0

Omondiagbe, D. A., Veeramani, S., & Sidhu, A. S. (2019a). Machine learning classification techniques for breast cancer diagnosis. IOP Conference Series: Materials Science and Engineering, 495, 012033. https://doi.org/10.1088/1757-899x/495/1/012033

Polyakova, M. V., & Krylov, V. N. (2022). Data normalization methods to improve the quality of classification in the breast cancer diagnostic system. *Applied Aspects of Information Technologies*, 5(1), 55–63. https://doi.org/10.15276/aait.05.2022.5

Shahriyari, L. (2019). Effect of normalization methods on the performance of supervised learning algorithms applied to HTSeq-FPKM-UQ datasets: 7SK RNA expression as a predictor of survival in patients with colon adenocarcinoma. *Briefings in Bioinformatics*. 2019, 20, 985–994.

Sharma, A., Goyal, D., & Mohana, R. (2024). An ensemble learning-based framework for breast cancer prediction. *Decision Analytics Journal*, 10, 100372. https://doi.org/10.1016/j.dajour.2023.100372

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, 105524. https://doi.org/10.1016/j.asoc.2019.105524

Strelcenia, E., & Prakoonwit, S. (2023). Effective feature engineering and Classification of breast cancer diagnosis: a Comparative study. *BioMedInformatics,* 3(3), 616–631. https://doi.org/10.3390/biomedinformatics3030042

World Health Organization. Breast cancer. 2021. https://www.who.int/news-room/fact- sheets/detail/breast-cancer

Yang, P., Wu, W., Wu, C., Shih, Y., Hsieh, C., & Hsu, J. (2021). Breast cancer recurrence prediction with ensemble methods and cost-sensitive learning. Open Medicine, 16(1), 754–768. https://doi.org/10.1515/med-2021-0282