Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 1, January 2026, pp: 257-269 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Impact of Different Kernels on Breast Cancer Severity Prediction Using Support Vector Machine

Kunti R. Mahmudah[1], Sugiyarto Surono[2], Rusmining[3], and Fatma Indriani[4]
[1] Department of Master Program in Mathematics Education, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2] Mathematics Study Program, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[3] Mathematics Education Study Program, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[4] Department of Computer Science, Faculty of Mathematics and Natural Science, Lambung Mangkurat University, Banjarbaru, Indonesia

Corresponding author: Kunti R. Mahmudah (e-mail: kunti.mahmudah@mpmat.uad.ac.id), **Author(s) Email**: Sugiyarto Surono (e-mail: sugiyarto@math.uad.ac.id), Rusmining (e-mail: rusmining@pmat.uad.ac.id), Fatma Indriyani (e-mail: f.indriani@ulm.ac.id)

Abstract **Breast cancer poses a critical global health challenge and continues to be one of the most prevalent causes of cancer-related deaths among women worldwide. Accurate and early classification of cancer severity is essential for improving treatment outcomes and guiding clinical decision-making, since timely intervention can significantly reduce mortality rates and enhance patient survival. This study evaluates the performance of Support Vector Machine (SVM) models using different kernel functions of Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid for breast cancer severity prediction. The impact of feature selection was also examined, using the Random Forest algorithm to select the top features based on Mean Decrease Accuracy (MDA), which serves to reduce redundancy, improve interpretability, and enhance model efficiency. Experimental results show that the RBF kernel consistently outperformed other kernels, especially in terms of sensitivity, a critical metric in medical diagnostics that emphasizes the ability of the model to identify positive cases correctly. Without feature selection, the RBF kernel achieved an accuracy of 0.9744, a sensitivity of 0.9772, a precision of 0.9722, and an AUC of 0.9968, indicating strong performance across all evaluation metrics. After applying feature selection, the RBF kernel further improved the accuracy to 0.9754, the sensitivity to 0.9770, the precision to 0.9742, and the AUC to 0.9975, which demonstrated enhanced generalization and reduced overfitting, highlighting the benefits of targeted feature reduction. While the Polynomial kernel yielded the highest precision (up to 0.9799), its lower sensitivity (as low as 0.9237) indicates a greater risk of false negatives, which is particularly concerning in cancer detection. These findings underscore the importance of optimizing both kernel function and feature selection. The RBF kernel, when combined with targeted feature selection, offers the most balanced and sensitive model, making it highly suitable for breast cancer classification tasks where diagnostic accuracy is vital.**

Keywords **Breast cancer; Support Vector Machine; kernel comparison; MDA-based features selection; severity prediction.**

I. Introduction

Breast cancer poses a critical global health challenge and continues to be one of the most prevalent causes of cancer-related deaths among women worldwide. According to the World Health Organization (WHO), in 2020, there were approximately 10 million cancer-related deaths. Among all cancer types, breast cancer remains the most commonly diagnosed, with an estimated 2.26 million new cases and 685,000 fatalities reported worldwide [1]. Accurate diagnosis and prognosis of breast cancer are crucial to ensuring effective treatment and improving patient outcomes. Early detection significantly increases survival rates; patients diagnosed at stage I or II have a five-year survival rate of up to 90%, whereas for stage IV, it drops to approximately 31% [1], [2]. However, early-stage breast cancer diagnosis remains a challenge due to the absence of noticeable symptoms, making it imperative to develop more efficient and cost-effective diagnostic methods.

Recent advancements in data science have demonstrated promising results in breast cancer detection and classification [3]. Support Vector Machine (SVM) is among the most extensively applied

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

machine learning techniques for classifying breast cancer types [4], [5], [6]. Its ability to manage high-dimensional datasets and model intricate nonlinear patterns has made SVM a popular choice in numerous bioinformatics applications, especially within cancer genomics research [4].

Several studies have confirmed the effectiveness of SVM in breast cancer diagnosis, with classification accuracy ranging from 52.63% to 98.24% [5]. One of the key factors influencing its accuracy is the selection of the kernel function [4]. In previous research, various linear and nonlinear kernel functions have been evaluated to optimize the performance of SVM in predicting breast cancer outcome [4]. In addition, several studies have examined the use of Least Squares Support Vector Machines (LS-SVM) in breast cancer diagnosis, achieving classification accuracy as high as 98.53% on image data [6]. These results underscore the effectiveness of SVM-based methods in generating accurate diagnoses and prognoses for breast cancer.

The accuracy of the SVM model is affected by various factors, and the choice of kernel is one of the most significant [7], [8], [9], [10]. In general, SVM kernels can be categorized into linear and nonlinear kernels [11]. The linear kernel offers a straightforward structure and is known for its computational efficiency, which is suitable for capturing linear relationships in data [12]. However, its performance may be limited when addressing intricate nonlinear patterns, which are often present in a high-dimensional dataset [13]. Nonlinear kernels enhance flexibility by mapping data into a higher-dimensional space, which enables SVM to capture intricate relationships that may be overlooked by linear kernels [11]. Similarly, the polynomial kernel allows modeling feature interactions at different levels, which can be beneficial in handling complex data structures [13]. However, in breast cancer studies, further investigation is needed to examine the influence of different kernel functions on the performance of SVM-based classification models and the interplay between feature selection.

This study aims to investigate the effects of various kernel functions used in SVM on the accuracy of breast cancer severity prediction. By comparing multiple kernel implementations in SVM models, we seek to identify the most effective approach for improving classification performance. Additionally, we will evaluate the selection and significance of features to ensure that only the most relevant factors are used in the classification process. The outcomes of this study are anticipated to provide valuable insights for SVM-based classification techniques in medical applications, particularly in facilitating the early diagnosis and prognosis of breast cancer. This study makes three key contributions. First, it provides a comprehensive evaluation of various SVM kernel functions, both linear and nonlinear, to determine their effectiveness in classifying breast cancer severity. Second, it integrates feature selection using the Random Forest algorithm to examine how optimized feature subsets interact with kernel choice, thereby improving classification accuracy, sensitivity, and generalization. Third, it contributes to the development of reliable and clinically relevant diagnostic models by emphasizing early detection and minimizing false negatives, offering practical insights for enhancing breast cancer diagnosis and prognosis.

This study is structured as follows: Section II presents related works, Section III presents the dataset, the proposed methods, the feature selection, and kernel types. Section IV reports the SVM classification results on different schemes and the interpretation of the findings. Section V presents a comparison with related studies, a discussion of limitations, and concludes with the main findings.

II. Related Works

Machine learning has been widely adopted in medical research to enhance disease diagnosis, prognosis, and treatment planning [14], [15]. Breast cancer becomes the most prevalent and deadly cancer type worldwide and has received growing recognition in the artificial intelligence (AI) and machine learning communities [16], [17], [18], [19], [20]. Among various classification techniques, Support Vector Machine (SVM) has demonstrated promising results in predicting breast cancer severity due to its robustness in handling high-dimensional and complex datasets [21].

A. Support Vector Machine (SVM) for Breast Cancer Classification

As a supervised learning method, SVM is frequently employed for binary classification problems such as distinguishing between benign and malignant tumors [22]. Several studies have validated the effectiveness of SVM for breast cancer classification. Research [23] explored the application of SVM in cancer genomics, highlighting its robustness in analyzing high-dimensional data and achieving promising classification accuracy.

A study by [24] introduced a novel correlation-based kernel designed explicitly for cancer diagnosis, demonstrating superior performance over classical kernels across five real-world gene expression datasets. Subsequently, the researchers introduced the Hadamard kernel as a parsimonious alternative for predicting breast cancer outcomes, underscoring the increasing importance of tailored kernel functions in improving the performance of SVM-based cancer classification models [11]. Additionally, [25] employed the Least Squares Support Vector Machine for classifying breast cancer and obtained an impressive accuracy of 98.53% when applied to microscopic

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

images. More recently, [26] combined deep learning approaches with SVM to improve breast cancer detection, demonstrating the adaptability of SVM in hybrid machine learning frameworks.

B. Kernel Function Selection in SVM

Despite these SVM advances, one of the most critical challenges in applying SVM to medical classification problems is the selection of an appropriate kernel function, as different kernels yield varying levels of classification performance [5].

Kernel functions significantly influence SVM's ability to generalize and classify data accurately. The choice of kernel determines how input features are transformed, impacting the capability of a model to identify complex patterns [27]. Kernel functions in SVM can be categorized into two general types: linear and nonlinear. Linear kernels are computationally efficient and suitable for datasets with well-separated classes, but they fail to detect intricate non-linear patterns commonly present in biomedical data [28]. In contrast, non-linear kernels enable SVM to map data into higher-dimensional feature spaces, which results in improved classification performance [29]. Study of [30] provided an extensive analysis of kernel functions, concluding that RBF kernels tend to be superior in handling high-dimensional biomedical data due to their flexibility in capturing complex patterns. Moreover, polynomial kernels have shown effectiveness in modeling feature interactions, making them suitable for datasets with quadratic or cubic relationships [31]. Hadamard kernels, which have been applied in genomic studies, have also demonstrated promising results in cancer classification tasks [11]. However, [32] emphasized the importance of tuning kernel parameters correctly, as improper configuration can lead to poor generalization and reduced classification accuracy.

C. Feature Selection for SVM-Based Breast Cancer Prediction

In addition to selecting the kernel function, feature selection plays a pivotal role in improving SVM performance in medical classification. High-dimensional datasets, such as those used in cancer research, often contain redundant or irrelevant features that can negatively impact model accuracy [33]. Various feature selection techniques have been employed to optimize SVM models. Random Forest has been widely used for feature selection before applying SVM, demonstrating improved classification performance by eliminating irrelevant features [34]. Study of [27] integrated Particle Swarm Optimization (PSO) with SVM to optimize feature selection, achieving significant improvements in breast cancer prediction. Similarly, [35] developed a hybrid Cat-and-Mouse optimization algorithm to refine feature selection in SVM-based models, further enhancing classification accuracy in biomedical applications. These findings suggest that effective feature selection, in conjunction with an appropriate kernel function, is crucial for optimizing the performance of SVM in breast cancer classification.

D. Limitations in Existing Research and Research Gaps

Despite the widespread use of SVM in breast cancer classification, several research gaps remain. Many studies apply SVM with pre-selected kernels without systematically evaluating their impact on classification performance. A comprehensive analysis comparing different kernel functions in breast cancer severity prediction is still lacking. Furthermore, while feature selection techniques are extensively studied, the interaction between feature selection methods and kernel selection in SVM models has not been sufficiently explored. The combined effect of optimal feature selection and kernel selection remains an open question. Additionally, hybrid machine learning approaches, such as integrating SVM with evolutionary optimization algorithms, have shown promise but remain underutilized in the field of medical classification. To address these gaps, the present study conducts a comprehensive comparative analysis of several commonly used SVM kernel functions to evaluate their effectiveness in predicting the severity of breast cancer. Additionally, it will adapt feature selection techniques to optimize SVM performance while reducing computational complexity. The study will also investigate the combined effect of kernel selection and feature selection on classification accuracy. By addressing these research gaps, this work will contribute to the advancement of SVM-based breast cancer classification models, ultimately aiding in the development of more accurate and interpretable machine learning solutions for medical applications.

III. Methods

A. Data Preparation and Collection

This study utilizes a publicly available dataset from the UCI Machine Learning Repository (https://archive.ics.uci.edu/dataset/16/breast+cancer+wisconsin+prognostic), which has been widely used in previous breast cancer classification research. The dataset consists of 198 samples, each representing a breast cancer case diagnosed. It contains 33 features related to tumor characteristics, including morphological attributes and histopathological measurements. The dataset includes both benign and malignant cases, making it suitable for severity classification tasks. Prior to model development, data preprocessing is conducted to handle missing values, normalize feature distributions, and ensure that the dataset is appropriately structured for analysis.

B. Feature Selection

To enhance model accuracy while minimizing computational complexity, a feature selection process

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

is conducted before applying the model. Therefore, the Random Forest (RF) algorithm can be used to identify the most relevant and informative features [36]. In RF, features are selected based on their impact on classification performance, enabling the removal of redundant or less significant variables [37]. By retaining only the most influential features, the subsequent Support Vector Machine (SVM) classifier achieves improved generalization capability and computational efficiency.

RF commonly assesses feature importance using two key metrics: Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDG). MDA quantifies a feature's impact by randomly permuting its value and evaluating the decrease in the model's accuracy. If the permutation of a feature significantly decreases accuracy, the feature is considered important. Formally, the importance of feature x_j using MDA is calculated using Eq. (1) as follows [38]:

$$MDA(x_j) = \frac{1}{T} \cdot \sum_{t=1}^{T} \frac{\sum_{i \epsilon OOB} I(y_i = f(x_i)) - \sum_{i \epsilon OOB} I\left(y_i = f\left(x_i^j\right)\right)}{|OOB|} \quad (1)$$

where T represents the number of trees within the forest, OOB is the set of out-of-bag samples, y_i the true class label of instance i, $f(x_i)$ is the predicted class label of instance iii when using the original feature set, $f\left(x_i^j\right)$ is the feature vector of instance iii where the values of feature x_j have been randomly permuted.

MDG, in contrast, evaluates a feature's importance based on the cumulative reduction in Gini impurity attributed to that feature across all trees in the forest. Gini impurity indicates the likelihood of misclassification when an instance is randomly assigned a label based on the dataset's label distribution. The importance of a feature f_j using MDG is calculated using Eq. (2) as follows [38]:

$$MDG(x_j) = \frac{1}{T}[1 - \sum_{t=1}^{T} Gini(j)^t] \quad (2)$$

where $Gini(j)^t$ represents the decrease in the Gini impurity attributed to the feature x_j in tree t. The Gini decrease is accumulated over all nodes where x_j is used for splitting.

In this study, feature importance was quantified using the Mean Decrease Accuracy (MDA) metric with a permutation-based approach. MDA was chosen as the basis for feature selection, as the focus of the classification task is to improve prediction accuracy. In this method, the trained RF model's predictive accuracy is first computed on the original dataset. Then, for each feature individually, its values are randomly permuted across all samples while keeping the rest of the data unchanged. This process disrupts the relationship between that feature and the target variable. The model's accuracy is recalculated, and the difference between the original accuracy and the accuracy after permutation represents the importance score for that feature. A larger accuracy drop indicates greater importance in classification.

For this study, the RF model was implemented with 500 trees (n_estimators = 500), a maximum depth selected via cross-validation, and a fixed random seed (random_state = 42) to ensure reproducibility. The permutation importance computation was repeated 30 times per feature, and the results were averaged to obtain stable MDA scores. Once MDA values were calculated for all 30 features, they were ranked in descending order. The top 20 features with the highest MDA scores were retained for subsequent SVM classification experiments. This threshold was chosen to balance model complexity and predictive performance, as preliminary trials indicated that including more than 20 features provided negligible accuracy gains while increasing training time. The RF feature importance scores were computed exclusively on the training portion of the data within each cross-validation fold. Specifically, for each training fold, an RF model was fit, permutation-based MDA scores were calculated, and the top 20 features were selected. These selected features were then used to train the SVM model on that fold. The corresponding validation fold remained completely unseen during both feature ranking and SVM training. This fold-by-fold approach ensured that no information from the validation or test data influenced the feature selection process, thereby avoiding overly optimistic performance estimates.

C. Kernel of SVM

Kernel functions are a fundamental component of the SVM, as they define how the input data is transformed and separated into a higher-dimensional space. The choice of kernel significantly influences the model's ability to detect linear or nonlinear patterns in data, especially in high-dimensional biomedical datasets such as those used in breast cancer classification. This study evaluates four widely used kernel functions: linear, polynomial, radial basis function (RBF), and sigmoid.

1. Linear Kernel

The linear kernel is the most basic form of kernel function, defined as the inner product between two vectors and formulated in Eq. (3) as follows [39]:

$$K(x_i, x_j) = x_i^{\intercal} x_j \quad (3)$$

where x_i and x_j represent two input feature vectors in the dataset, each corresponding to a sample. It is suitable for linearly separable data and is computationally efficient. In medical datasets with well-separated classes, the linear kernel often performs well. However, it may not be sufficient for capturing the complex, nonlinear relationships typical of cancer data.

2. Polynomial Kernel

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

The polynomial kernel allows SVM to fit curved boundaries and model interactions between features. It is calculated using Eq. (4) as follows [39]:

$$K(x_i, x_j) = (\gamma x_i^{\intercal} x_j + r)^{\boldsymbol{d}} \quad (4)$$

where γ is a scale parameter, r is a coefficient (often called bias), and d is the polynomial degree. Higher-degree polynomials enable the model to learn more complex relationships but may risk overfitting if not carefully tuned.

3. Radial Basis Function (RBF) Kernel

The RBF, also referred to as the Gaussian kernel, maps input data into an infinite-dimensional space and is calculated using Eq. (5) as [39]:

$$K(x_i, x_j) = \exp\left(-\gamma(x_i - x_j)^{\intercal}(x_i - x_j)\right) \quad (5)$$

where γ controls the width of the kernel. The RBF kernel is highly effective for capturing nonlinear patterns and is particularly suitable when the decision boundary is not clearly linear. It is widely used in bioinformatics and medical diagnostics.

4. Sigmoid Kernel

Inspired by neural networks, this kernel is calculated using Eq. (6) as follows [40]:

$$K(x_i, x_j) = \tanh(\gamma x_i^{\intercal} x_j + r) \quad (6)$$

where $\tanh$ denotes the hyperbolic tangent function. This kernel resembles the structure of a two-layer neural network and is capable of modeling nonlinear patterns.

D. Hyperparameter Tuning

To ensure fair comparison and optimal performance, the Support Vector Machine classifiers were tuned using kernel-specific grids and repeated cross-validation. For the linear kernel, the penalty parameter C was explored over $\{2^{-5}, 2^{-3}, \ldots, 2^{15}\}$. For the radial basis function (RBF) kernel, C was varied over the same range, while γ was varied over $\{2^{-15}, 2^{-3}, \ldots, 2^{3}\}$. For the polynomial kernel, C and γ used the same search ranges, with polynomial degree $d \in \{2,3,4\}$ and $coef0 \in \{0,1\}$. The sigmoid kernel was tuned over the same C and coef0 as the polynomial kernel. All parameter combinations were evaluated using 10-fold cross-validation repeated three times, with the mean area under the ROC curve (AUC) as the selection criterion. Preprocessing (min-max scaling), class balancing (SMOTE), and feature selection were applied within each training fold only to prevent information leakage. The final models were retrained on the whole training set using the best hyperparameters and then evaluated on the held-out test set.

E. Model Evaluation and Performance Metrics

To evaluate the SVM model's performance with different kernels, four standard classification metrics are used. Accuracy evaluated the overall model performance, which represents the proportion of correct predictions. Sensitivity measures the model's ability to accurately detect malignant cases, simultaneously minimizing both false positives and false negatives. Additionally, the Area Under the Curve (AUC) is calculated to assess the model's ability to classify cases based on the severity levels. As formulated in Eq. (7), accuracy indicates the ratio of correct predictions to the total number of prediction outcomes. It serves as a key metric for evaluating the model's capability in differentiating samples from various sources. Achieving high accuracy is crucial for the detection of breast cancer and determining its level of severity.

Sensitivity, as formulated in Eq. (8), indicates the model's capacity to accurately classify true positives, which indicate the existence of breast cancer, but are misclassified as absent. Precision, as formulated in Eq.9, reflects the model's ability to accurately classify the positive instances, i.e., cases where breast cancer is misclassified as existing despite its absence. This metric is particularly important in clinical settings where unnecessary alarms can lead to additional anxiety, costly follow-up procedures, or overtreatment. High sensitivity is important to reduce the risk of inappropriate treatment of breast cancer. The three metrics discussed above are summarized in Table 1 and calculated using the following formulas.

Table 1. Confusion matrix

	Predicted as Positive	Predicted as Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (8)$$

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

AUC, as formulated in Eq. (9), measures how well the model differentiates between classes, which offers a comprehensive view of performance across various threshold performances. A higher AUC value reflects stronger discriminatory power in identifying the existence versus absence of breast cancer

F. Experimental Workflow

The study follows a structured workflow to ensure a systematic analysis of breast cancer severity classification using SVM with kernel comparison, as shown in Fig. 1, as follows. The first step involves data preprocessing, where missing values are handled and features are standardized. Following this, feature

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

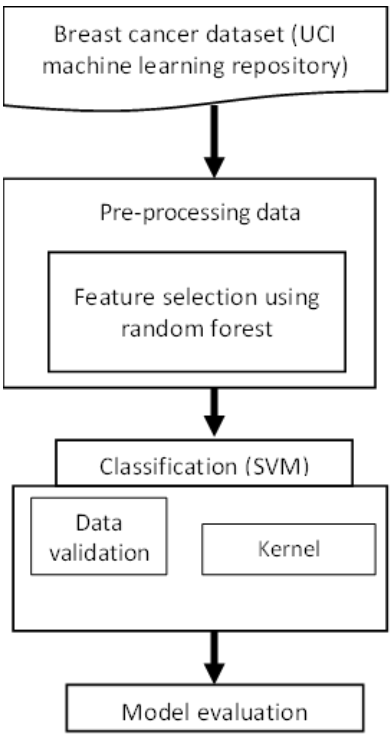

Fig. 1. Flowchart of data analysis

selection is conducted using the Random Forest algorithm. The importance of features is measured based on the Mean Decreased Accuracy (MDA) score. Once the most significant features are identified, the SVM model is trained and tested using different kernel functions. A comparative analysis of kernel functions is performed to determine which approach provides the best classification performance. Finally, model evaluation metrics are analyzed to validate the effectiveness of the proposed approach. This methodology provides a comprehensive approach to optimizing SVM-based breast cancer severity classification by integrating feature selection and kernel function evaluation.

IV. Result and Analysis

A. Preprocessing data

The dataset contained no missing values, as confirmed during the initial inspection. To ensure uniform feature scaling and prevent variables with larger numeric ranges from dominating the SVM optimization process, all 30 numerical features were normalized using min-max scaling to the range of [0,1]. The normalization was applied independently to each feature according to Eq. 10 as follows.

$$x' = \frac{x - \min A}{\max A - \min A} \quad (10)$$

where A is a variable (feature/column) in the dataset, $x \in A$ is the original feature value, and x' is the normalized value in the range [0,1].

To address class imbalance between malignant and benign cases, the Synthetic Minority Over-sampling Technique (SMOTE) was applied with a $K = 5$ nearest neighbors and a duplication size of 100%. This generated synthetic samples for the minority class, which were combined with the original dataset and shuffled to randomize observation order. All preprocessing steps were performed before feature selection and model training, ensuring that the training and testing splits (0.8 and 0.2) were drawn from the same standardized feature space. Random seeds were fixed (set.seed(123)) for reproducibility.

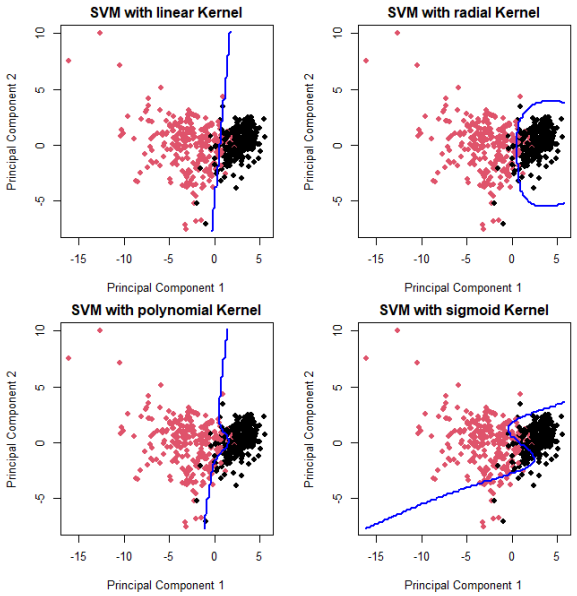

Fig 2. SVM decision boundaries based on kernel

B. Kernel Selection

Before evaluating model performance, the dataset was visualized using Support Vector Machine (SVM) decision boundaries generated by four kernel functions: Linear, Polynomial, RBF, and Sigmoid. These visualizations, shown in Fig. 2, provide an initial understanding of how each kernel transforms the space of features and segregates the two classes.

The Linear kernel produces a straight boundary, indicating its suitability for linearly separable data. While it creates a clear margin between some regions, its rigid structure may struggle in areas with overlapping or complex distributions. The Polynomial kernel introduces curved decision boundaries that can capture more intricate relationships between features. However, depending on the polynomial degree, this flexibility may lead to either over-simplification or overfitting. The RBF kernel demonstrates highly

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

adaptive boundaries that curve around data clusters, suggesting a strong capability to model non-linear patterns. Visually, it appears most effective in separating densely packed or irregularly shaped regions. The Sigmoid kernel generates boundaries with a soft, S-shaped curve. Though it introduces nonlinearity, the separation is less distinct compared to other kernels, potentially limiting its ability to divide the classes clearly. These plots provide valuable insights into how each kernel function shapes the classification space, helping to motivate the need for a more detailed performance comparison in the following sections.

C. Features Selection Using Random Forest

The MDA metric from RF is employed as a criterion for selecting important features to enhance classification performance while minimizing computational complexity. This method assesses the effect of each feature on the accuracy of the model by randomly permuting its samples and measuring the subsequent drop in prediction performance [41]. A larger drop in accuracy indicates a more important feature. The ranking shown in Fig. 3 reveals that among the 30 original features, several exhibit significantly higher contributions to classification accuracy. Notably, features F22, F21, F24, F23, and F28 exhibited the highest MDA values, indicating their strong influence on the model's predictive performance, whose score is presented in Table 2 below.

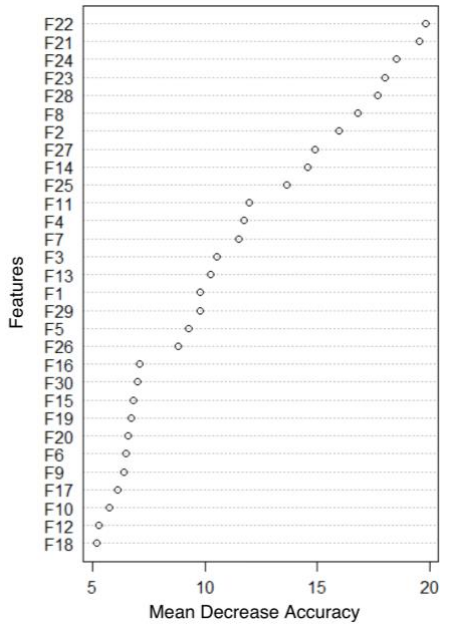

Fig 3. MDA and MDG ranking of random forest

To optimize the balance between accuracy and computational efficiency, the top 20 features were selected based on their MDA rankings. These features demonstrated substantial predictive importance, with MDA values ranging from approximately 0.075 to 0.004. By excluding the lower-ranked features, which contributed minimal or negligible improvement, the dataset's dimensionality was reduced without compromising classification integrity.

Table 2. Features importance based on MDA scores

	MDA		MDA		MDA
F21	0.0753	F3	0.0206	F5	0.0043
F23	0.0703	F22	0.0139	F18	0.0024
F24	0.0644	F11	0.0117	F16	0.0022
F8	0.0528	F13	0.0107	F30	0.0020
F28	0.0527	F2	0.0101	F20	0.0016
F27	0.0393	F25	0.0076	F9	0.0015
F14	0.0302	F26	0.0067	F15	0.0015
F4	0.0290	F17	0.0047	F10	0.0015
F7	0.0256	F29	0.0046	F19	0.0011
F1	0.0214	F6	0.0045	F12	0.0010

This selection approach not only enhances the generalization capability of the subsequent SVM model but also minimizes the risk of overfitting and reduces training time. The retained features were then used as input for the kernel comparison in the SVM classification stage

D. Performance by Utilizing Kernel Optimization without Performing Feature Selection

To establish a baseline and evaluate the impact of feature selection, the Support Vector Machine (SVM) model was initially tested using all available features without prior dimensionality reduction. The kernel functions that form the focus of this research are Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid. The classification performance was assessed using four standard metrics: Accuracy, Sensitivity, Precision, and AUC (Area Under the ROC Curve), as summarized in Table 3.

Table 3. SVM Performances based on Feature Selection and Kernel Optimization

Kernel	Acc.	Sensitivity	Precision	AUC
Linear	0.9746	0.9772	0.9725	0.9967
Polynom	0.9516	0.9237	0.9786	0.9945
Radial	0.9744	0.9772	0.9722	0.9968
Sigmoid	0.9524	0.9758	0.9327	0.9922

The Linear kernel yielded the highest accuracy (97.46%) and sensitivity (97.72%), demonstrating excellent ability to identify malignant cases correctly. The RBF kernel performed comparably, with slightly

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

lower accuracy (97.44%) but the highest AUC (99.68%), indicating superior performance in distinguishing between classes across all thresholds. The Polynomial kernel achieved the highest precision (97.86%), reflecting its strength in minimizing false positives, but showed a noticeable drop in sensitivity (0.9237), raising concerns about its reliability in detecting all malignant cases. The Sigmoid kernel exhibited moderate and balanced results, with strong sensitivity (97.58%), although its overall accuracy and AUC were slightly lower than those of the linear and RBF alternatives. These initial results suggest that, without feature selection, Linear and RBF kernels offer the most consistent and reliable performance for breast cancer severity classification. In the next phase, the influence of feature selection on these kernel performances is examined to evaluate potential improvements in both accuracy and model efficiency.

E. Performance by Utilizing Feature Selection and Kernel Optimization

Table 4 presents the performance of four kernel functions: linear, Radial Basis Function (RBF), and polynomial, with a combination of feature selection for breast cancer classification.

Table 4. Performance analysis of SVM with feature selection and kernel selection

Kernel	Acc.	Sensitivity	Precision	AUC
Linear	0.9734	0.9749	0.9722	0.9962
Polynom	0.9525	0.9242	0.9799	0.9952
Radial	0.9754	0.9770	0.9742	0.9975
Sigmoid	0.9489	0.9646	0.9357	0.9885

The linear kernel achieves an accuracy of 97.34%, indicating strong overall performance in classifying the majority of breast cancer cases correctly. Its sensitivity of 97.49% represents a slight decrease from the previous result of 97.72%, suggesting that feature selection maintained high classification capability while potentially enhancing model efficiency and interpretability. With an AUC of 99.62%, the linear kernel demonstrates excellent discriminative power in distinguishing between malignant and benign cases, minimizing both false positives and false negatives. The RBF kernel with feature selection yields the highest accuracy among all kernels at 97.54%, outperforming both the RBF kernel without feature selection (97.44%) and the linear kernel. Its sensitivity of 97.70% is also the highest, indicating slightly better performance in detecting malignant cases. The AUC of 99.75% is marginally superior to that of the linear kernel, confirming the RBF kernel's effectiveness in distinguishing between classes.

The polynomial kernel achieves the lowest sensitivity at 92.42%, suggesting reduced effectiveness in identifying malignant cases compared to other kernels. Its accuracy of 95.25% is also lower than both the linear and RBF kernels. Although its precision is the highest at 97.99%, this may reflect a trade-off with sensitivity. The AUC of 99.52%, while still high, is below that of the linear and RBF kernels. The sigmoid kernel achieves the lowest overall performance, with an accuracy of 94.89% and a precision of 93.57%. Although its sensitivity is relatively high at 96.46%, the AUC of 98.85% is the lowest among all kernels, indicating reduced discriminative capability. When applying feature selection and kernel optimization, all four kernels demonstrate strong classification performance. The RBF kernel offers the best overall balance with the highest accuracy, sensitivity, and AUC. The linear kernel also performs consistently well, particularly in terms of accuracy and AUC. While the polynomial kernel excels in precision, it is less effective in sensitivity. The specific performance priorities of the classification task should therefore guide the choice of kernel.

V. Discussion

This study evaluated the performance of Support Vector Machine (SVM) classifiers using four kernel functions: linear, Polynomial, Radial Basis Function (RBF), and Sigmoid for predicting breast cancer severity. The evaluation was conducted under two scenarios: first, using the complete set of features, and second, after applying feature selection based on Mean Decrease Accuracy (MDA) scores from a Random Forest model, where the top 20 most important features were retained to enhance model efficiency and focus.

The findings indicate that the choice of kernel function has a significant influence on model performance, particularly in medical classification tasks where sensitivity is crucial. Sensitivity measures the model's ability to correctly identify malignant cases, which is particularly important in breast cancer diagnosis, where false negatives can result in delayed or missed treatments. Among the tested kernels, the RBF consistently yielded the highest sensitivity: 97.72% without feature selection and 97.70% with feature selection. These results highlight the RBF kernel's strong ability to detect malignant cases accurately while maintaining other performance metrics. This aligns with prior studies such as [42], which reported sensitivity up to 98.7% in optimized SVM models for medical classification.

After applying feature selection, the RBF kernel achieved the highest accuracy of 97.54% and the highest AUC of 99.75%, indicating superior overall performance and excellent discrimination between malignant and benign cases. These outcomes suggest that the RBF kernel offers a balanced combination of high sensitivity, strong precision, and overall classification reliability. In contrast, the Linear kernel

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

also performed competitively, achieving 97.49% sensitivity and 99.62% AUC after feature selection. While its performance was slightly lower than that of RBF, the linear kernel remains a strong candidate due to its computational simplicity and consistent results. This is consistent with findings in similar classification contexts, such as diabetes detection [43].

The Polynomial kernel recorded the highest precision in both scenarios (97.86% without feature selection and 97.99% with), reflecting its effectiveness in minimizing false positives. However, its sensitivity was the lowest among the four kernels at 92.37% without and 92.42% with feature selection. This makes this kernel less suitable in clinical settings where missing malignant cases is highly undesirable. The Sigmoid kernel yielded moderate results across metrics and did not outperform other kernels in any category, limiting its suitability for high-stakes diagnostic applications.

The performance comparison between models with and without feature selection showed only marginal improvements, particularly in AUC and sensitivity for the RBF kernel. However, the use of RF-based feature selection still provided practical benefits by reducing the dimensionality of the input space, improving computational efficiency, and enhancing the interpretability of the results. In real-world applications, especially where model transparency and speed are valued, the ability to focus on a smaller set of the most predictive features may justify the additional preprocessing step, even if absolute performance gains are modest. This trade-off highlights the role of feature selection not only in maximizing accuracy but also in promoting generalization and practical deployment.

This high AUC of the RBF kernel also indicates that it maintains superior discriminative ability across a range of classification thresholds, which means it can effectively separate malignant from benign cases under varying decision criteria. In a real-world diagnostic context, this translates to greater flexibility in tuning the classification threshold to prioritize clinical objectives. For example, in breast cancer screening, sensitivity is typically prioritized to minimize false negatives, thereby reducing the risk of missed diagnoses. Our results show that the RBF kernel not only achieves the highest AUC but also delivers the highest sensitivity among the evaluated kernels, which reinforces its suitability for early detection tasks, where failing to identify a malignant case can have serious consequences. While the polynomial kernel demonstrated the highest precision, this came at the cost of reduced sensitivity, that is potential to more missed positive cases. Thus, the trade-off between kernels involves balancing the clinical imperative for high sensitivity against the need to reduce false positives. Given the potential for patient anxiety, unnecessary follow-up tests, and associated costs arising from false positives, the choice of kernel should be informed by the specific diagnostic setting, with the RBF kernel offering the most adaptable performance profile for sensitivity-driven screening programs. Additionally, reducing the number of features improved model interpretability and likely reduced the risk of overfitting, which is important when handling complex biomedical data.

The superior performance of the RBF kernel compared to the linear, polynomial, and sigmoid alternatives can be attributed to its ability to model complex, non-linear relationships between features and class labels [44], [45]. Breast cancer morphological and textural features often exhibit non-linear interactions, which the RBF kernel effectively captures by projecting data into a high-dimensional feature space, where classes become more separable. Unlike the linear kernel, which assumes a purely linear boundary, or the polynomial kernel, which may overfit when degree is high, the RBF kernel adapts its flexibility through the γ parameter in enabling a balance between bias and variance. Prior studies in medical imaging and cancer diagnosis have also reported that the RBF kernel consistently outperforms other kernels in handling heterogeneous biomedical data [42] which is owed to its robustness to irrelevant features and capacity to handle overlapping class distributions. This property, combined with our targeted feature selection strategy, likely explains the observed balance between sensitivity, precision, and AUC, making the RBF kernel the most reliable choice for our classification task.

VI. Conclusion

This study aimed to evaluate the impact of different kernel functions (linear, Polynomial, RBF, and Sigmoid) on the performance of SVM models for predicting breast cancer severity, both with and without feature selection using Random Forest-based MDA. The results indicate that the choice of kernel function has a significant impact on classification outcomes. Among the tested kernels, the RBF kernel consistently achieved the best overall performance, with an accuracy of 97.44%, a sensitivity of 97.72%, a precision of 97.22%, and an AUC of 99.68% without feature selection. After applying feature selection, its performance further improved to an accuracy of 97.54%, sensitivity of 97.70%, precision of 97.42%, and AUC of 99.75%, which shows its strong ability to balance diagnostic sensitivity and overall reliability. The Linear kernel also performed competitively, with an accuracy of 97.34% and AUC of 99.62% after feature selection, which makes it a practical alternative due to its computational simplicity. The Polynomial kernel achieved the highest precision (97.99%) but suffered from low sensitivity (92.42%), while the Sigmoid kernel

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

produced moderate results across all metrics, therefore limiting its clinical applicability. Future research should build on these results by validating the approach on larger and more diverse datasets to strengthen generalizability across populations. Another promising direction is the development of hybrid models that integrate SVM with ensemble or deep learning methods to capture more complex nonlinear relationships while preserving interpretability. Adaptive feature selection strategies tailored to specific patient cohorts could also improve robustness and reduce overfitting risks. Beyond algorithmic improvements, future efforts should explore the hyperparameter optimization of kernel parameters, longitudinal studies that incorporate temporal patterns in patient data, and the multimodal integration of genomic, imaging, and clinical data. Importantly, translating these models into clinical decision-support systems and conducting prospective clinical trials would provide evidence of real-world applicability and clinical utility. Overall, these directions will guide future research toward the development of reliable, scalable, and clinically applicable AI-based diagnostic tools that can support early detection and personalized treatment planning in breast cancer care.

Acknowledgment

The authors would like to acknowledge the financial support from Universitas Ahmad Dahlan through research funding under grant number PD-086/SP3/LPPM-UAD/IX/2024. The academic environment, facilities, and continuous encouragement from the institution have played a significant role in enabling the successful completion of this work. This study would not have been possible without the university's strong commitment to advancing research and innovation in the fields of bioinformatics and computer science.

Funding

This research was funded by Universitas Ahmad Dahlan under grant number PD-086/SP3/LPPM-UAD/IX/2024.

Data Availability

This study utilizes a publicly available dataset from the UCI Machine Learning Repository, specifically the Breast Cancer Wisconsin Dataset. The dataset has been widely used in previous breast cancer classification research and can be accessed freely at the provided link.

Author Contribution

Kunti Robiatul Mahmudah conceptualized and designed the study, led the research process, and was responsible for data collection, analysis, and interpretation. Sugiyarto Surono contributed expertise in statistics, supported the methodological framework, and provided guidance in data analysis. Rusmining contributed expertise in mathematics, particularly in the theoretical foundation of the Support Vector Machine model and kernel functions. Fatma Indriani contributed expertise in computer science, supported the implementation of machine learning algorithms, and assisted with coding, validation, and visualization. All authors contributed to manuscript preparation, critically reviewed the content, approved the final version for submission, and agreed to be accountable for the integrity and accuracy of the work.

Declaration

Ethical Approval

This study did not involve human participants, animals, or clinical interventions. The research was conducted using a publicly available secondary dataset obtained from an established data repository. Therefore, formal ethical approval was not required.

Consent for Publication Participants

Not applicable, as the study did not involve direct participation of humans or animals.

Competing Interests

The authors declare that there is no conflict of interest associated with this research. This study was carried out with the full support of Universitas Ahmad Dahlan, whose academic environment, resources, and encouragement have greatly contributed to the successful completion of this work. The institution's commitment to advancing research and fostering innovation has been invaluable in ensuring the integrity and quality of this study.

References

[1] 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, May 2021, doi: 10.3322/caac.21660.

[2] S. Saadatmand, R. Bretveld, S. Siesling, and M. M. A. Tilanus-Linthorst, "Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173 797 patients," *BMJ*, p. h4901, Oct. 2015, doi: 10.1136/bmj.h4901.

[3] K. Polat and S. Güneş, "Breast cancer diagnosis using least square support vector machine," *Digit Signal Process*, vol. 17, no. 4, pp. 694–701, Jul. 2007, doi: 10.1016/j.dsp.2006.10.008.

[4] N. Cristianini and B. Scholkopf, "Support Vector Machines and Kernel Methods: The New Generation of Learning Machines," *AI Mag*, vol. 23, no. 3, pp. 31–41, Sep. 2002.

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

[5] R. Vinge and T. McKelvey, "Understanding Support Vector Machines with Polynomial Kernels," in *2019 27th European Signal Processing Conference (EUSIPCO)*, IEEE, Sep. 2019, pp. 1–5. doi: 10.23919/EUSIPCO.2019.8903042.

[6] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of Support Vector Machine (SVM) Learning in Cancer Genomics," *Cancer Genomics Proteomics*, vol. 15, no. 1, Jan. 2018, doi: 10.21873/cgp.20063.

[7] F. S. Gomiasti, W. Warto, E. Kartikadarma, J. Gondohanindijo, and D. R. I. M. Setiadi, "Enhancing Lung Cancer Classification Effectiveness Through Hyperparameter-Tuned Support Vector Machine," *Journal of Computing Theories and Applications*, vol. 1, no. 4, pp. 396–406, Mar. 2024, doi: 10.62411/jcta.10106.

[8] C. S. Rao and K. Karunakara, "Efficient Detection and Classification of Brain Tumor using Kernel based SVM for MRI," *Multimed Tools Appl*, vol. 81, no. 5, pp. 7393–7417, Feb. 2022, doi: 10.1007/s11042-021-11821-z.

[9] A. P. Gopi, R. N. S. Jyothi, V. L. Narayana, and K. S. Sandeep, "Classification of tweets data based on polarity using improved RBF kernel of SVM," *International Journal of Information Technology*, vol. 15, no. 2, pp. 965–980, Feb. 2023, doi: 10.1007/s41870-019-00409-4.

[10] M. Ring and B. M. Eskofier, "An approximation of the Gaussian RBF kernel for efficient classification with SVMs," *Pattern Recognit Lett*, vol. 84, pp. 107–113, Dec. 2016, doi: 10.1016/j.patrec.2016.08.013.

[11] H. Jiang, W.-K. Ching, W.-S. Cheung, W. Hou, and H. Yin, "Hadamard Kernel SVM with applications for breast cancer outcome predictions," *BMC Syst Biol*, vol. 11, no. S7, p. 138, Dec. 2017, doi: 10.1186/s12918-017-0514-1.

[12] M. N. Murty and R. Raghava, *Support Vector Machines and Perceptrons*. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-41063-0.

[13] S. F. Hussain, "A novel robust kernel for classifying high-dimensional data using Support Vector Machines," *Expert Syst Appl*, vol. 131, pp. 116–131, Oct. 2019, doi: 10.1016/j.eswa.2019.04.037.

[14] B. Zhang, H. Shi, and H. Wang, "Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach," *J Multidiscip Healthc*, vol. Volume 16, pp. 1779–1791, Jun. 2023, doi: 10.2147/JMDH.S410301.

[15] R. R. Chandan *et al.*, "Reviewing the Impact of Machine Learning on Disease Diagnosis and Prognosis: A Comprehensive Analysis," *Open Pain J*, vol. 17, no. 1, May 2024, doi: 10.2174/0118763863291395240516093102.

[16] Z. Zhu, Y. Sun, and B. Honarvar Shakibaei Asli, "Early Breast Cancer Detection Using Artificial Intelligence Techniques Based on Advanced Image Processing Tools," *Electronics (Basel)*, vol. 13, no. 17, p. 3575, Sep. 2024, doi: 10.3390/electronics13173575.

[17] M. R. Darbandi, M. Darbandi, S. Darbandi, I. Bado, M. Hadizadeh, and H. R. Khorram Khorshid, "Artificial intelligence breakthroughs in pioneering early diagnosis and precision treatment of breast cancer: A multimethod study," *Eur J Cancer*, vol. 209, p. 114227, Sep. 2024, doi: 10.1016/j.ejca.2024.114227.

[18] A. Soliman, Z. Li, and A. V. Parwani, "Artificial intelligence's impact on breast cancer pathology: a literature review," *Diagn Pathol*, vol. 19, no. 1, p. 38, Feb. 2024, doi: 10.1186/s13000-024-01453-w.

[19] A. Singh, A. Singh, and S. Bhattacharya, "Research trends on AI in breast cancer diagnosis, and treatment over two decades," *Discover Oncology*, vol. 15, no. 1, p. 772, Dec. 2024, doi: 10.1007/s12672-024-01671-0.

[20] T. Fujioka *et al.*, "The Future of Breast Cancer Diagnosis in Japan with AI and Ultrasonography," *JMA J*, vol. 8, no. 1, pp. 91–101, 2024, doi: 10.31662/jmaj.2024-0183.

[21] B. Ghaddar and J. Naoum-Sawaya, "High dimensional data classification and feature selection using support vector machines," *Eur J Oper Res*, vol. 265, no. 3, pp. 993–1004, Mar. 2018, doi: 10.1016/j.ejor.2017.08.040.

[22] W. Zeng, J. Jia, Z. Zheng, C. Xie, and L. Guo, "A comparison study: Support vector machines for binary classification in machine learning," in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, IEEE, Oct. 2011, pp. 1621–1625. doi: 10.1109/BMEI.2011.6098517.

[23] B. S. Abunasser, M. R. J. AL-Hiealy, I. S. Zaqout, and S. S. Abu-Naser, "Breast Cancer Detection and Classification using Deep Learning Xception Algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, 2022, doi: 10.14569/IJACSA.2022.0130729.

[24] H. Jiang and W.-K. Ching, "Correlation Kernels for Support Vector Machines Classification with Applications in Cancer Data," *Comput Math Methods Med*, vol. 2012, pp. 1–7, 2012, doi: 10.1155/2012/205025.

[25] S. A. Korkmaz and M. Poyraz, "Least Square Support Vector Machine and Minumum Redundacy Maximum Relavance for Diagnosis of Breast Cancer from Breast Microscopic Images," *Procedia Soc Behav Sci*, vol. 174, pp. 4026–

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

4031, Feb. 2015, doi: 10.1016/j.sbspro.2015.01.1150.

[26] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," *PeerJ*, vol. 7, p. e6201, Jan. 2019, doi: 10.7717/peerj.6201.

[27] M. P. Behera, A. Sarangi, D. Mishra, and S. K. Sarangi, "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine," *Procedia Comput Sci*, vol. 218, pp. 818–827, 2023, doi: 10.1016/j.procs.2023.01.062.

[28] A. Patle and D. S. Chouhan, "SVM kernel functions for classification," in *2013 International Conference on Advances in Technology and Engineering (ICATE)*, IEEE, Jan. 2013, pp. 1–9. doi: 10.1109/ICAdTE.2013.6524743.

[29] T. Kavzoglu and I. Colkesen, "A kernel functions analysis for support vector machines for land cover classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, no. 5, pp. 352–359, Oct. 2009, doi: 10.1016/j.jag.2009.06.002.

[30] R. Fernandes de Mello and M. Antonelli Ponti, "Statistical Learning Theory," in *Machine Learning*, Cham: Springer International Publishing, 2018, pp. 75–128. doi: 10.1007/978-3-319-94989-5_2.

[31] M. Azzeh, Y. Elsheikh, A. B. Nassif, and L. Angelis, "Examining the performance of kernel methods for software defect prediction based on support vector machine," *Sci Comput Program*, vol. 226, p. 102916, Mar. 2023, doi: 10.1016/j.scico.2022.102916.

[32] A. Goel and S. Kr. Srivastava, "Role of Kernel Parameters in Performance Evaluation of SVM," in *2016 Second International Conference on Computational Intelligence & Communication Technology (CICT)*, IEEE, Feb. 2016, pp. 166–169. doi: 10.1109/CICT.2016.40.

[33] P. El Kafrawy, H. Fathi, M. Qaraad, A. K. Kelany, and X. Chen, "An Efficient SVM-Based Feature Selection Model for Cancer Classification Using High-Dimensional Microarray Data," *IEEE Access*, vol. 9, pp. 155353–155369, 2021, doi: 10.1109/ACCESS.2021.3123090.

[34] T. Ebina, H. Toh, and Y. Kuroda, "DROP: an SVM domain linker predictor trained with optimal features selected by random forest," *Bioinformatics*, vol. 27, no. 4, pp. 487–494, Feb. 2011, doi: 10.1093/bioinformatics/btq700.

[35] B. Kalpana *et al.*, "Cat and Mouse Optimizer with Artificial Intelligence Enabled Biomedical Data Classification," *Computer Systems Science and Engineering*, vol. 44, no. 3, pp. 2243–2257, 2023, doi: 10.32604/csse.2023.027129.

[36] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognit Lett*, vol. 31, no. 14, pp. 2225–2236, Oct. 2010, doi: 10.1016/j.patrec.2010.03.014.

[37] A.-L. Boulesteix and M. Slawski, "Stability and aggregation of ranked gene lists," *Brief Bioinform*, vol. 10, no. 5, pp. 556–568, Sep. 2009, doi: 10.1093/bib/bbp034.

[38] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," *BMC Bioinformatics*, vol. 9, no. 1, p. 307, Dec. 2008, doi: 10.1186/1471-2105-9-307.

[39] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. El Ghaoui, and M. I. Jordan, "Learning the Kernel Matrix with Semidefinite Programming," *Journal of Machine Learning Research*, vol. 5, pp. 27–72, 2004.

[40] H.-T. Lin and Chih-Jen Lin., "A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods.," *Neural Comput*, vol. 3, no. 16, pp. 1–32, 2003.

[41] A. Bahl *et al.*, "Recursive feature elimination in random forest classification supports nanomaterial grouping," *NanoImpact*, vol. 15, p. 100179, Mar. 2019, doi: 10.1016/j.impact.2019.100179.

[42] R. S.V.G, R. K.Thammi, K. V. Valli, and V. Kamadi VSRP, "An SVM Based Approach to Breast Cancer Classification using RBF and Polynomial Kernel Functions with Varying Arguments," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5901–5904, 2014.

[43] G. A. Pethunachiyar, "Classification Of Diabetes Patients Using Kernel Based Support Vector Machines," in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, Jan. 2020, pp. 1–4. doi: 10.1109/ICCCI48352.2020.9104185.

[44] Zhen-Dong Zhao, Y.-Y. Lou, Jun-Hong Ni, and Jing Zhang, "RBF-SVM and its application on reliability evaluation of electric power system communication network," in *2009 International Conference on Machine Learning and Cybernetics*, IEEE, Jul. 2009, pp. 1188–1193. doi: 10.1109/ICMLC.2009.5212365.

[45] X. Ding, J. Liu, F. Yang, and J. Cao, "Random radial basis function kernel-based support vector machine," *J Franklin Inst*, vol. 358, no. 18, pp. 10121–10140, Dec. 2021, doi: 10.1016/j.jfranklin.2021.10.005.

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960

Author Biography

Kunti Robiatul Mahmudah is a committed research scholar in the Department of Mathematics Education at Universitas Ahmad Dahlan, Indonesia. She earned her master's degree in science from the Department of Mathematics, Gadjah Mada University, Indonesia, in 2017. Following her master's studies, she pursued and completed a Ph.D. in Computer Science from Kanazawa University, Japan, in 2021. Her academic pursuits and research focus underscore her dedication to advancing the fields of computational statistics and big data analytics, with a primary application of these theories in the field of bioinformatics. ORCID ID: 0000-0003-4660-5454

Sugiyarto Surono is a senior lecturer at the Mathematics Study Program, Ahmad Dahlan University. He has a deep interest in the implementation of Big Data, Machine Learning, Artificial Intelligence, and Deep Learning. As a professor, Sugiyarto actively teaches and guides Mathematics students, combining various fields of science through Data Science applications. His objective is to make students aware of the importance of this knowledge in the context of technological development. Through this approach, Sugiyarto is trying to make a significant contribution in preparing the younger generation to face the technological challenges of the future. ORCID ID: 0000-0001-6210-7258

Rusmining began her academic journey with a bachelor's degree in mathematics education from Universitas Negeri Semarang, Indonesia, where she also completed her master's degree. Her research focuses on applying statistical and computational methods to analyze mathematical literacy and develop digital learning tools using platforms such as GeoGebra and LaTeX. She actively trains teachers to integrate these technologies, improving students' critical thinking and problem-solving abilities. Rusmining's work emphasizes data-driven approaches to assessing and enhancing student performance, closely aligning with machine learning goals like pattern recognition and predictive modeling in education. ORCHID ID: 0009-0007-0459-5549

Fatma Indriani is a lecturer in the Department of Computer Science at Lambung Mangkurat University, with a strong research interest in Data Science. Before pursuing an academic career, she completed her undergraduate studies in the Informatics Department at the Bandung Institute of Technology. In 2008, she began her journey as a lecturer at Lambung Mangkurat University, contributing to the field of Computer Science through teaching and research. To further expand her expertise, she pursued a master's degree at Monash University, Australia, which she completed in 2012. Her academic journey continued with a doctorate in Bioinformatics from Kanazawa University, Japan, which she completed in 2022. With a focus on both Data Science and Bioinformatics, she actively engages in research, exploring innovative ways to leverage data-driven technologies for scientific advancement. Her dedication to academia and research enables her to make significant contributions to the advancement of knowledge in her field, while also mentoring students and collaborating on interdisciplinary projects. ORCID ID: 0009-0006-7180-6708

Manuscript received June 3, 2025; Revised September 10, 2025; Accepted October 10, 2025; date of publication January 13, 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.960