A Novel Method for Breast Cancer Diagnosis Classification using the DWD Loss Function

Zakaria Khoudi, Nasreddine Hafidi, Soufiane Lyaqini, Mourad Nachaoui and Youssef Ben Youssef

***Abstract*—Breast cancer is the most common form of invasive cancer in women and ranks second in terms of cancer-related deaths among women. It can be classified into two main types: benign and malignant. In recent years, there has been an increased focus among experts on addressing this concerning issue. With advancements in machine learning, predicting diseases and preventing life-threatening conditions have become possible. Several breast cancer prediction models have been created using various statistical and machine learning techniques. These techniques and algorithms directly apply to datasets to build models or draw crucial insights. In this study, we have employed machine learning techniques on the Wisconsin Breast Cancer dataset (WBCD) to develop an intelligent model. Our proposed algorithm achieved the highest accuracy rate (99.30%) compared to seven other supervised machine learning techniques based on the experimental results.**

***Index Terms*—Breast cancer, (WBCD) dataset, Non Smooth SVM, Kernel methods**

I. Introduction

BREAST cancer is one of the most common cancers in women. It is a type of cancer in which the breast cells proliferate irregularly. Indeed, breast cancer is genetic in 10% of instances, with the remaining 90% caused by lifestyle factors. According to the World Health Organization (WHO), breast cancer afflicts more than 1.5 million women globally each year. According to a study by the National Cancer Registration Program (NCRP), cancer cases are estimated to increase from 13.9 lakh in 2020 to 15.7 lakh in 2025, a total increase of 20% [1]. If common cancers are treated early, they can be prevented from becoming fatal. Early detection of breast cancer can result in efficient therapy.

In recent years, experts have been more concerned about [2], [3]. As machine learning advances, the capability to predict diseases and prevent life-threatening conditions has become achievable. Numerous models for predicting breast cancer have been created, employing a range of statistical and machine learning approaches [4]–[6]. These methods and algorithms are directly applied to datasets for constructing models or drawing significant conclusions. The utilization of AI in medical domains is rapidly expanding, driven by its effectiveness in predicting and categorizing, especially in the clinical assessment of breast cancer [7].

Several machine learning algorithms are employed to classify patients with malignant and benign tumors, with support vector machine (SVM) algorithms being popular. However, in many cases, researchers adopt a square loss function to construct SVM models [8]. Despite its prevalence, the square loss function tends to exaggerate the impact of outliers in classification tasks, rendering prediction algorithms vulnerable [9]. To mitigate the influence of outliers, researchers have turned to hinge loss, which is less sensitive to noise than the square loss function [10]–[12]. The SVM algorithm is then reformulated as a regularized optimization problem. However, the resulting fidelity term in the minimization problem is non-differentiable. To address this challenge, smooth approximation techniques are employed, transforming the max function into a smoothed variant [13]. Subsequently, the modified minimization problem is solved using Adam's algorithm [14]. The paper concludes with numerous numerical validations of the proposed method, comparing it to existing machine learning algorithms. The results underscore the effectiveness of the proposed technique in diagnosing breast cancer types, outperforming seven supervised machine-learning approaches. This enhanced machine learning model offers a valuable tool for early cancer detection, confirming and indicating the presence of breast cancer.

The paper's structure is organized as follows: Section II outlines the classification problem's context, transforming it into a minimization problem using the smooth hinge loss function and the optimization approach. Section III comprehensively examines comparison experiments, focusing on prediction accuracy, the F-measure metric, and the Area Under the Curve (AUC). Finally, Section IV summarizes the conclusion.

II. Setting of the problem

For the binary classification in $\mathbb{R}^n$, The training set can be expressed as follows:

$$\mathcal{T} = \{(x_j, y_j)\}_{j=1}^{n}$$

where $\boldsymbol{x}_j \in \mathbb{R}^n$ and $y_j \in \{-1, 1\}$, let $\boldsymbol{\theta} = \left(\boldsymbol{w}^T, b\right)^T$, an SVM model tries to seek a decision hyperplane $f(x) = \boldsymbol{w}^T\Phi(\boldsymbol{x})+b = 0$, that maximizes the separation between two classes of data. Specifically for the training set, the smooth SVM model is formulated as a convex optimization problem

Manuscript received January 21, 2024; revised March 25, 2025.

Zakaria Khoudi is a PhD candidate of the EMI FST Sultan Moulay Slimane University Beni Mellal 23000, Morocco. (e-mail: zakaria.khoudi@usms.ma)

Nasreddine Hafidi is a PhD candidate of the EMI FST Sultan Moulay Slimane University Beni Mellal 23000, Morocco. (e-mail: nasreddine.hafidi@usms.ma)

Soufiane Lyaqini is a professor of the Ecole Nationale des Sciences Appliquees Sultan Moulay Slimane University Khouribgha 25000, Morocco. (e-mail: lyaqini.soufiane@gmail.com)

Mourad Nachaoui is a professor of the EMI FST Sultan Moulay Slimane University Beni Mellal 23000, Morocco. (e-mail: m.nachaoui@usms.ma)

Youssef Ben Youssef is a professor of the Ecole Nationale des Sciences Appliquees LISA Laboratory Hassan First University Settat 26000, Morocco. (e-mail: youssef.benyoussef@uhp.ac.ma)

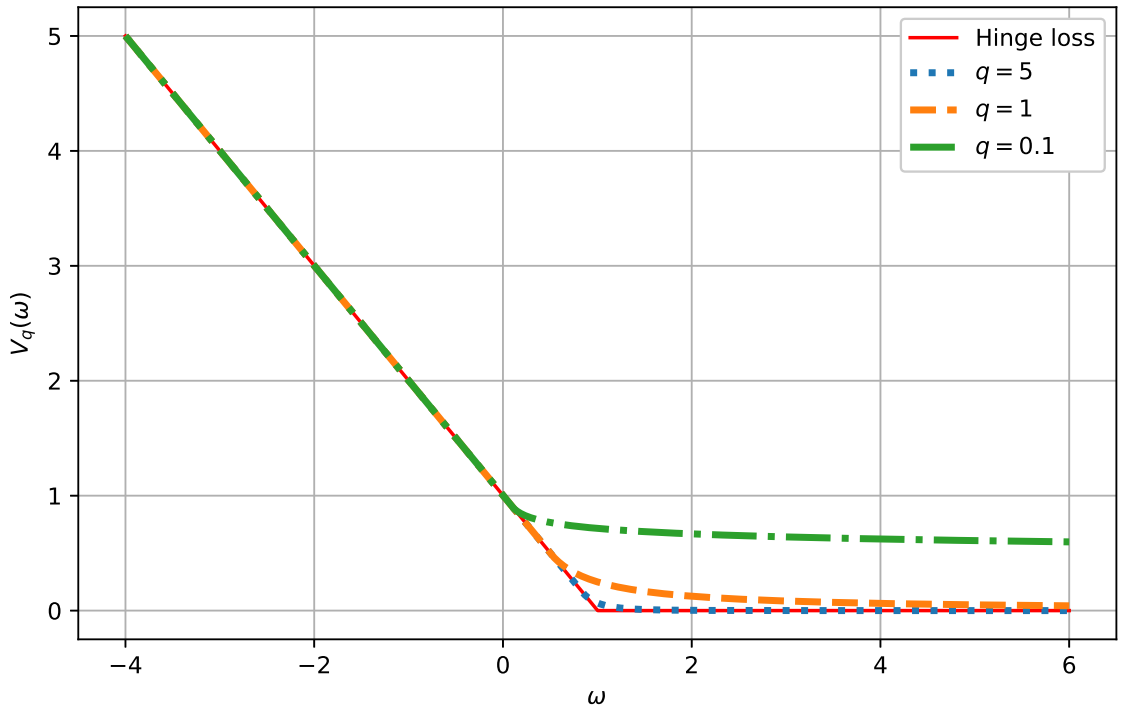

Fig. 1. Smoothed Hings loss.

with equation constraints (1)

$$\min_{\boldsymbol{\theta}} C\|\boldsymbol{w}\|^2 + C\sum_{j=1}^{n} V_q\left(1 - y_j f(x_j)\right) \tag{1}$$

where

$$V_q(\omega) = \begin{cases} 1-\omega & \text{if } \omega \leq \frac{q}{1+q}, \\ \frac{1}{\omega^q}\frac{q^q}{(q+1)^{q+1}} & \text{if } \omega > \frac{q}{1+q}, \end{cases}$$

is the DWD loss function, $\Phi(\boldsymbol{x}_j)$ maps $\boldsymbol{x}_j$ into a higher-dimensional space Y and $C > 0$ is the regularization parameter. As seen in Fig. 1, as q goes to ∞, the DWD loss function converges to the hinge loss function. $V(\omega) = \max(0, 1-\omega)$, used in the SVM.

When the kernel method is used, we can solve Eq. (1) in a wider reproducing kernel Hilbert space (RKHS) $\mathcal{K}$ as

$$\min_{f\in\mathcal{K}} C\|\boldsymbol{f}\|_{\mathcal{K}}^2 + \sum_{j=1}^{n} L_\epsilon\left(1 - y_j f(x_j)\right) \tag{2}$$

in RKHS, the Representer Theorem [15] shall be valid, and the solution to the problem of minimizing a regularized risk function can be given as follows:

$$f^* = \sum_{j=1}^{n} \beta_j K\left(\boldsymbol{x}_j, \cdot\right),$$

where K is a kernel function in RKHS and the coefficient $\beta_j \in \mathbb{R}$.

Consequently, using the Representer Theorem in RKHS, we can further express the primal model in Eq. (2) as

$$\min_{\boldsymbol{\beta},b} C\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\beta} + \sum_{j=1}^{n} L_\epsilon\left(1 - (y_j, \boldsymbol{\beta}^T\boldsymbol{K}_j + b)\right) \tag{3}$$

where K is the indefinite kernel matrix derived from associated kernel function $\boldsymbol{K}_{ij} = K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)$, $\boldsymbol{K}_j$ represents the i the row of $\boldsymbol{K}$ and $\boldsymbol{\beta}^T = (\beta_1, \beta_2, ..., \beta_n) \in \mathbb{R}^n$.

Numerous nonlinear optimization techniques are applicable for addressing the problem described in Eq. (3). Among these methods, gradient-type approaches are particularly appealing for solving such optimization challenges. The Adam method, distinguished by its simplicity and minimal storage requirements, emerges as a highly effective gradient-type method for this purpose [14].

III. Experimental Results

In this section, we apply our methodology (Algorithm 1) to predict the classification of breast cancer types in patients. We utilize the Wisconsin Breast Cancer Dataset (WBCD) from the UCI Machine Learning Repository for this purpose.

A. Dataset description

In this section, our approach is applied to predict breast cancer types in patients using the Wisconsin Breast Cancer Dataset (WBCD) retrieved from the UCI Machine Learning Repository https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic. This dataset, created by Dr. William H. Wolberg from the University of Wisconsin Hospital, consists of 569 instances classified as either benign or malignant. There are 357 benign cases (62.7%) and 212 malignant cases (37.3%), as depicted in Fig. 2. The dataset features 32 attributes, with one serving as the target column indicating the diagnosis ('B' for benign and 'M' for malignant). The properties of the dataset are detailed in Table I.

Algorithm 1 Breast Cancer Classification Methodology

Input: Dataset D from the Wisconsin Breast Cancer Dataset (WBCD)
Output: Prediction of breast cancer type (benign/malignant)
Split the dataset D into training set D_{train} and testing set D_{test} with an 80:20 ratio.
Step 1: Data Preprocessing
    Normalize the dataset features to ensure all features contribute equally to model training.
    Handle missing values or outliers if necessary.
Step 2: Model Selection
    Choose a set of supervised machine learning algorithms:
        - Decision Trees (DT)
        - Random Forests (RF)
        - Logistic Regression (LR)
        - Support Vector Machines (SVM)
        - K-Nearest Neighbors (KNN)
        - Naive Bayes (NB)
        - XGBoost (XGB)
        - Our approach (Smooth Adam)
Step 3: Model Training
for each model in the chosen set **do**
    Train the model on the training set D_{train}.
end for
Step 4: Model Evaluation
for each trained model **do**
    Evaluate the model on the test set D_{test} using the following metrics:
        - Accuracy
        - Precision
        - Recall
        - F1-score
        - R2-score
        - AUC
end for
Step 5: Model Comparison
Compare the performance of each model based on the evaluation metrics and select the best-performing model.
Step 6: Feature Importance Analysis
Use feature importance techniques to identify the most influential features in the dataset.

TABLE I
DESCRIPTION OF THE DATASET.

No	Attribute	Description
1	Radius	mean distances from the canter to the points on the perimeter
2	Texture	standard variation of grey-scale values
3	Perimeter	
4	Area	
5	Smoothness	local variation in radius lengths
6	Compactness	perimeter^2 / area - 1.0
5	Concavity	severity of concave portions of the contour
6	Concave points	number of concave portions of the contour
9	Symmetry	
10	Fractal dimension	coastline approximation - 1
11	Diagnosis	M=malignant, B = benign

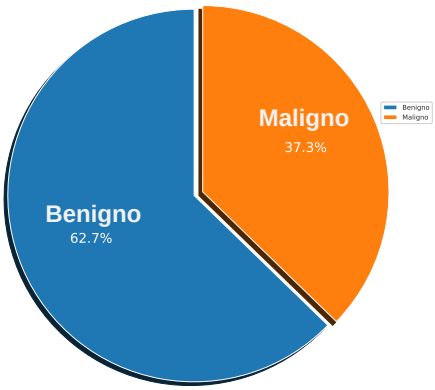

Fig. 2. The percentage of benign and malignant records.

Fig. 2 displays the distribution of breast cancer cases within the dataset, with 62.7% of the instances classified as benign and 37.3% as malignant, highlighting a higher prevalence of benign cases (357) compared to malignant ones (212). This class imbalance is important to consider during model training, as it can lead to biased predictions favoring the majority class.

The observations from Fig. 3 indicate that several features, such as perimeter, radius, area, concavity, and compactness, exhibit an exponential distribution pattern, while features like texture, smoothness, and symmetry appear to follow a Gaussian or approximately Gaussian distribution. This finding is significant because many machine learning methods assume that input data adhere to a Gaussian univariate distribution. Understanding these distribution patterns is essential for selecting appropriate modeling techniques and preprocessing steps, as it influences decisions such as data normalization, feature scaling, and the choice of algorithms best suited for the data's characteristics, ultimately impacting the effectiveness and accuracy of data analysis and machine learning tasks.

Fig. 4 uses a heatmap to visualize the relationships among features and their connections to breast cancer diagnosis, revealing several notable associations. Strong positive correlations, with coefficients ranging from 0.75 to 1, are observed among mean value parameters, indicating that they tend to increase together. For instance, the mean area of the tissue nucleus shows strong positive correlations with both the mean radius and perimeter. Additionally, there are moderate positive correlations (0.5 to 0.75) between features like concavity and area and concavity and perimeter. Conversely, robust negative correlations are seen between the fractal dimension and mean values of radius, texture, and perimeter, suggesting that the fractal dimension tends to decrease as these features increase. These insights are crucial for understanding feature interactions in breast cancer diagnosis, aiding in effective feature selection and predictive modeling.

B. Description of utilized algorithms

In this study, we employed several machine learning algorithms to classify breast cancer as benign or malignant

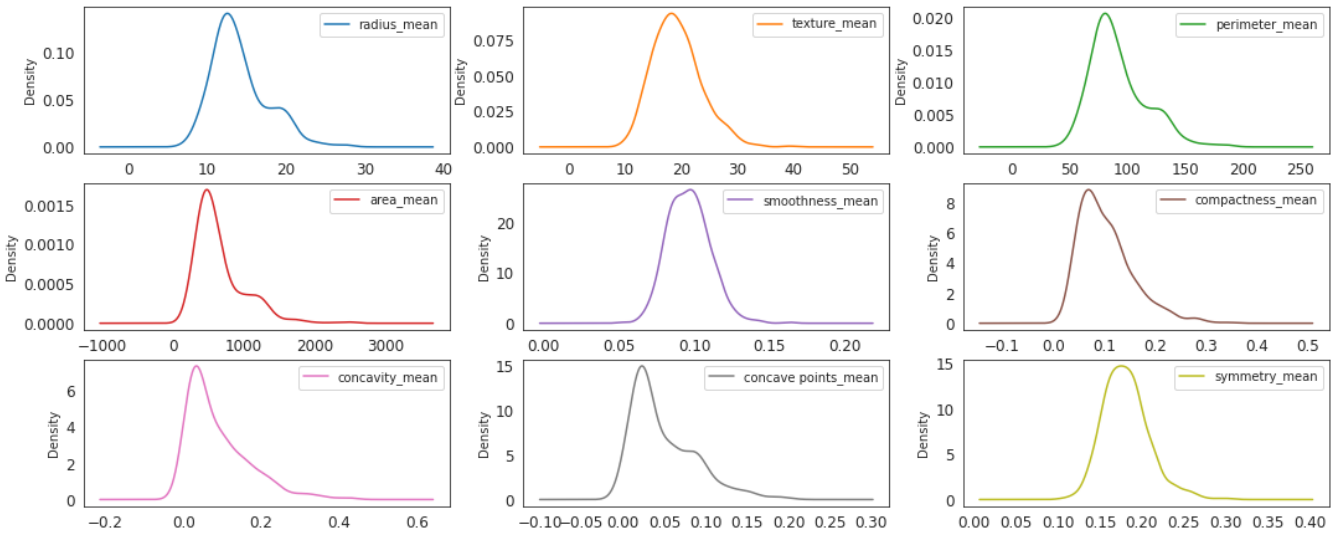

Fig. 3. Visualization of the dataset.

using the Wisconsin Breast Cancer Dataset (WBCD). These models are popular in the field of medical diagnosis due to their ability to handle large datasets and produce high classification accuracy.

a) Decision Trees (DT): Decision Trees are a non-parametric supervised learning method used for classification and regression tasks. They function by recursively partitioning the feature space into regions based on the values of the input features. At each internal node, the algorithm chooses a feature that best separates the data according to some criterion, such as information gain or Gini impurity. The data is split until it is fully classified into distinct branches. This model is widely appreciated for its simplicity and interpretability [16]. However, decision trees are prone to overfitting when dealing with noisy data or complex datasets.

For a node with two classes (binary classification), the Gini impurity is calculated as:

$$G = 1 - p_1^2 - p_2^2$$

Where: - p_1 is the proportion of samples belonging to class 1 at the node. - $p_2 = 1 - p_1$ is the proportion of samples belonging to class 2.

This function helps the decision tree determine how "pure" a node is, with a lower Gini impurity indicating a more homogeneous (pure) node. The tree uses this measure to decide the best splits at each step.

b) Random Forests (RF): Random Forests are an ensemble learning method introduced by Breiman [17]. This technique aggregates the predictions of multiple decision trees to improve accuracy and robustness. The key idea is to create multiple trees using different subsets of the data and features, and then aggregate the results (typically by majority voting for classification tasks). Random Forests are well-known for reducing the variance that individual decision trees might introduce, leading to improved generalization to unseen data.

The final prediction of a Random Forest for binary classification is represented as:

$$\hat{y} = \text{round}\left(\frac{1}{T}\sum_{t=1}^{T} f_t(x)\right)$$

Where: - $\hat{y}$ is the predicted class (0 or 1). - T is the total number of decision trees. - $f_t(x)$ is the prediction of the t-th tree for input x, where $f_t(x) \in \{0, 1\}$.

This function averages the individual predictions $f_t(x)$ across all trees and rounds to the nearest integer (0 or 1), effectively implementing a majority vote to determine the final class label.

c) Logistic Regression (LR): Logistic Regression is a statistical model that is primarily used for binary classification tasks. It models the probability of a binary outcome based on a linear combination of the input features. The output is transformed using the logistic function, which ensures that the predicted probabilities lie between 0 and 1. Logistic Regression is widely used in various fields for its simplicity and efficiency in cases where the relationship between the features and the output is approximately linear [18].

Logistic regression in binary classification models the probability that a given input x belongs to the positive class (1) using the logistic function:

$$P(y = 1|x) = \sigma(z) = \frac{1}{1 + e^{-z}}$$

where z is a linear combination of the input features given by $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$. In this equation, β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients corresponding to each feature $x_1, x_2, \ldots, x_n$, and e is the base of the natural logarithm. To make a binary classification decision, the predicted probability $P(y = 1|x)$ can be thresholded at 0.5: if $P(y = 1|x) \geq 0.5$, then the predicted class $\hat{y}$ is 1; otherwise, it is 0.

d) Support Vector Machines (SVM): Support Vector Machines (SVM) are supervised learning models that are used for both classification and regression. They work by finding the hyperplane that best separates data points into distinct classes with the maximum margin. For non-linear classification problems, SVMs can apply

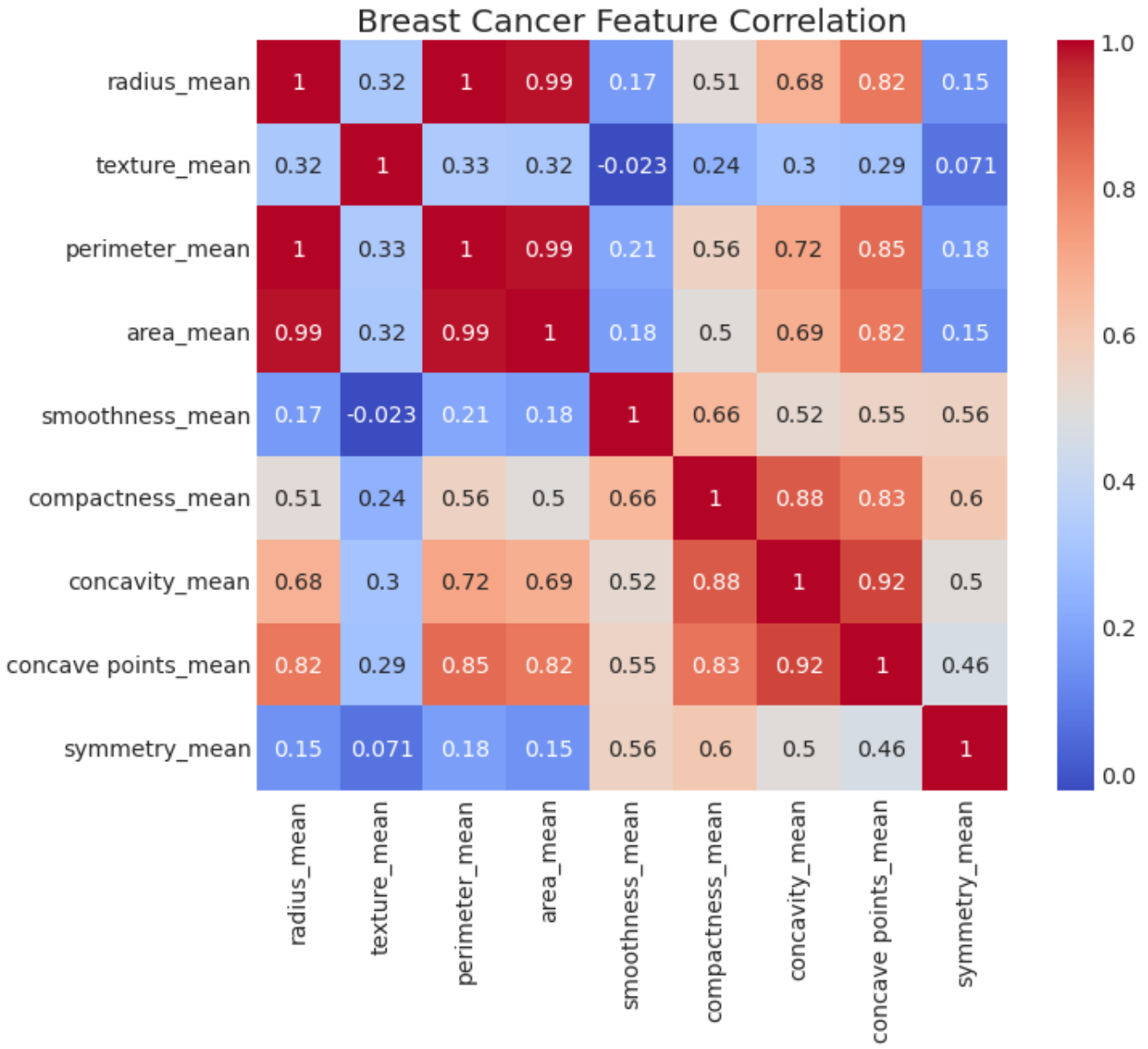

Fig. 4. Heatmap of correlation matrix.

the kernel trick, which transforms the input space into a higher-dimensional feature space where the data becomes linearly separable. SVMs have demonstrated significant success in high-dimensional spaces and with complex datasets [19].

Support Vector Machines (SVM) in binary classification is defined by the decision function given by

$$f(x) = \mathbf{w} \cdot \mathbf{x} + b$$

where $f(x)$ represents the decision function, $\mathbf{w}$ is the weight vector that determines the orientation of the decision boundary, $\mathbf{x}$ is the input feature vector, and b is the bias term. The predicted class label $\hat{y}$ is then determined by the sign of the decision function:

$$\hat{y} = \begin{cases} 1 & \text{if } f(x) \geq 0 \\ 0 & \text{if } f(x) < 0 \end{cases}$$

This indicates that if the decision function $f(x)$ is greater than or equal to zero, the predicted class is 1; otherwise, it is 0.

e) Naive Bayes (NB): Naive Bayes is a probabilistic classifier based on Bayes' Theorem, with the assumption that features are conditionally independent given the class label. Despite this strong assumption of independence, Naive Bayes classifiers perform surprisingly well in many real-world scenarios, particularly in text classification tasks. The model estimates the posterior probability of each class given the input features and selects the class with the highest probability [20]. It is simple to implement and computationally efficient, making it suitable for large datasets.

Naive Bayes in binary classification can be expressed concisely as:

$$\hat{y} = \begin{cases} 1 & \text{if } P(x|y=1)P(y=1) \geq P(x|y=0)P(y=0) \\ 0 & \text{otherwise} \end{cases}$$

This function classifies an instance x as class $\hat{y} = 1$ if the product of the likelihood $P(x|y = 1)$ and the prior probability $P(y = 1)$ is greater than or equal to the product of the likelihood $P(x|y = 0)$ and the prior probability $P(y = 0)$; otherwise, it classifies it as $\hat{y} = 0$.

f) K-Nearest Neighbors (KNN): K-Nearest Neighbors (KNN) is a simple, instance-based learning algorithm used for classification and regression. It classifies a data point based on the majority class of its nearest neighbors, typically in Euclidean space. No explicit training is involved in KNN; instead, it makes predictions based on the closest stored instances in the training set. KNN is highly sensitive to the choice of the parameter K and the distance metric used [21].

k-Nearest Neighbors (k-NN) in binary classification can be expressed as follows:

$$\hat{y} = \text{mode}(y_{i_1}, y_{i_2}, \ldots, y_{i_k})$$

Where:

- $\hat{y}$ is the predicted class label for the input instance.
- $y_{i_1}, y_{i_2}, \ldots, y_{i_k}$ are the class labels of the k nearest neighbors to the input instance in the feature space.
- The function mode returns the most frequently occurring class label among the k neighbors.

This function classifies an input instance by assigning it the class label that is most common among its k nearest neighbors in the training dataset.

g) XGBoost (XGB): XGBoost, short for eXtreme Gradient Boosting, is an efficient and scalable implementation of the gradient boosting framework [22]. In gradient boosting, models are built sequentially, with each new model correcting the errors of the previous ones. XGBoost incorporates regularization to control overfitting and utilizes optimized hardware capabilities to boost computational performance. It is widely known for its success in many machine learning competitions due to its predictive power and efficiency.

XGBoost in binary classification can be expressed as follows:

$$\hat{y} = \text{sigmoid}\left(\sum_{m=1}^{M} f_m(x)\right)$$

Where:

- $\hat{y}$ is the predicted probability that the input instance belongs to the positive class (1).
- M is the total number of trees in the ensemble.
- $f_m(x)$ is the prediction from the m-th tree for the input instance x.

The function sigmoid(z) is defined as:

$$\text{sigmoid}(z) = \frac{1}{1 + e^{-z}}$$

This function outputs a probability between 0 and 1. To make a binary classification decision, the predicted probability $\hat{y}$ can be thresholded at 0.5: if $\hat{y} \geq 0.5$, then the prediction is class 1 (positive class); otherwise, it is class 0 (negative class).

C. Data Splitting and Performance Measures

The dataset has been divided into training and testing subsets with an 80% to 20% split, respectively. After training the model on the training dataset, its performance is assessed using the testing dataset, and various performance metrics are recorded. The proposed model's performance evaluation utilizes several widely recognized metrics: Accuracy, Precision, Recall, and F1-score, as detailed in table II. Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total instances. Precision, also known as the positive predictive value, indicates the accuracy of the optimistic predictions. Recall, or sensitivity, measures the model's ability to identify positive instances. The F1-score, a harmonic mean of Precision and Recall, provides a single metric that balances both concerns, offering a comprehensive view of the model's performance. These metrics collectively help understand the model's efficacy in accurately predicting breast cancer types.

TABLE II
PERFORMANCE METRICS.

Metric	Formula
Accuracy	$\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}$
Precision	$\frac{\text{TP}}{\text{TP}+\text{FP}}$
Recall	$\frac{\text{TP}}{\text{TP}+\text{FN}}$
F1-score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}}$

where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative.

D. Results and discussions

In this research study, a series of experiments were conducted using a diverse set of machine learning algorithms, which included Decision Trees (DT), Random Forests (RF), Logistic Regression (LR), Support Vector Machines (SVM), Naive Bayes (NB), K-Nearest Neighbors (KNN), and XGBoost (XGB) [23]–[27]. The primary aim of these experiments was to assess the performance of these classification algorithms in their ability to predict the type of breast cancer. Various metrics were employed to evaluate their performance, such as prediction accuracy and the F-measure. The outcomes of these experiments were summarized and presented in table III. This table compares the effectiveness of Decision Trees (DT), Random Forests (RF), Logistic Regression (LR), Support Vector Machines (SVM), Naive Bayes (NB), K-Nearest Neighbors (KNN), and XGBoost (XGB) against our proposed approach. Our approach demonstrates superior performance across all metrics compared to the other algorithms. Specifically, it achieves the highest Accuracy (99.30%), indicating the highest overall correctness in predictions. The F1-Score, which balances Precision and Recall, is also the highest at 99.45%, reflecting our model's robustness in identifying actual positive cases while minimizing false positives. Furthermore, our approach attains perfect Recall (100%), effectively capturing all positive cases. Precision is also high at 98.91%, indicating a high rate of correct optimistic predictions. Additionally, the R2-Score, which measures the proportion of variance captured by the model, is the highest

at 96.98%, underscoring the model's explanatory solid power. These results highlight the strength and reliability of our approach in accurately predicting breast cancer types, making it a highly effective tool for this classification task.

TABLE III
PERFORMANCE MEASURES % FOR EACH ALGORITHM.

Metric	DT	RF	LR	SVM	NB	KNN	XGB	Our approach
Accuracy	95.10	97.90	93.71	97.90	95.10	98.60	9720	**99.30**
F1-Score	96.08	0.98	94.97	98.32	96.13	0.98	97.80	**99.45**
Recall	94.50	97.80	93.41	96.70	95.60	100	97.80	**100**
Precision	97.72	98.89	96.59	100	96.67	97.85	97.80	**98.91**
R2-Score	78.84	90.93	72.80	90.93	78.85	93.96	87.91	**96.98**

Table IV shows the confusion matrix for each algorithm, highlighting the number of correctly and incorrectly classified instances. Our approach outperforms the other algorithms regarding classification accuracy, with the highest number of correctly classified instances (142) and the lowest number of incorrectly classified instances (1). This demonstrates the robustness and effectiveness of our model in accurately predicting breast cancer types. Our approach ensures more reliable and precise diagnostics by minimizing misclassifications, which is critical for clinical decision-making and patient outcomes. These results underscore the superiority of our method compared to the other evaluated algorithms.

TABLE IV
CONFUSION MATRIX FOR EACH ALGORITHM.

Metric	DT	RF	LR	SVM	NB	KNN	XGB	Our approach
Correctly classified	136	140	137	140	136	141	139	**142**
Incorrectly classified	7	3	6	3	7	2	4	**1**

Table V presents the Area Under the Curve (AUC) percentages for each algorithm, comparing their performance on training and testing datasets. The AUC metric provides insight into the overall performance of a model, indicating its ability to distinguish between classes. A higher AUC value reflects a better-performing model. All algorithms achieve a perfect AUC of 100% on the training dataset, demonstrating their capacity to accurately fit the training data. However, the actual test of a model's performance is its ability to generalize to unseen data, as reflected in the test AUC scores. Our approach achieves the highest test AUC of 99%, surpassing all other algorithms. This superior performance on the test dataset indicates that our model fits the training data well and generalizes effectively to new, unseen data. This robustness is crucial for real-world applications, where the model must maintain high accuracy and reliability when predicting breast cancer types in new patients. These results highlight the strength and effectiveness of our approach in achieving optimal predictive performance.

TABLE V
AUC % FOR EACH ALGORITHM.

Metric	DT	RF	LR	SVM	NB	KNN	XGB	Our approach
Train AUC	100	100	100	100	100	100	100	**100**
Test AUC	95.3	97.9	95.3	98.4	94.9	98.1	97	**99**

Table VI compares the performance of our approach to previous studies' results, explicitly focusing on the accuracy of different methods used in breast cancer diagnosis. Our approach achieves an accuracy of 99.30%, the highest among the methods compared. Kumar et al. [28] use of K-Nearest Neighbors (KNN) achieved an accuracy of 83.45%, significantly lower than our approach. Both Agarap [29] and Salem et al. [30] achieved high accuracies with Support Vector Machines (SVM) and Artificial Neural Networks (ANN) at 99.00%, respectively. However, our approach still outperforms them with a slight edge. Deepa et al. [31] and Naji et al. [32] employed SVM and Convolutional Neural Networks (CNN), respectively, achieving accuracies of 97.60% and 97.20%, both of which are lower than our result. The superior accuracy of our approach demonstrates its robustness and effectiveness in predicting breast cancer types, surpassing the performance of previously studied methods. This high level of accuracy is critical in clinical settings, as it can lead to more reliable diagnoses and better patient outcomes. These results underscore the strength of our model in achieving state-of-the-art performance in breast cancer classification tasks.

TABLE VI
COMPARISON OF PERFORMANCE TO PREVIOUS STUDIES' RESULTS.

	[28]	[29]	[31]	[32]	[30]	
Method	KNN	SVM	SVM	CNN	ANN	**Our approach**
Accuracy	83.45	99.00	97.60	97.20	99.00	**99.30**

In summary, we draw the following conclusions based on the experimental results regarding the accuracy of our proposed approach: table III shows that our approach outperforms Decision Trees (DT), Random Forests (RF), Logistic Regression (LR), Support Vector Machines (SVM), Naive Bayes (NB), K-Nearest Neighbors (KNN), and XGBoost (XGB) models. Furthermore, the quantitative results in table VI demonstrate that the proposed approach surpasses the existing state-of-the-art methods. Indeed, the obtained results confirm that the proposed SVM with the smoothed DWD Loss Function is an enhanced machine learning model for classification tasks. With the aid of the advanced early diagnosis system, cancer can be accurately detected in its initial stages, allowing for the prevention of the disease's more severe symptoms.

E. Feature importance

In this specific section of the research, an additional analysis was performed to evaluate the importance of individual features concerning their predictive power for breast cancer diagnosis. This assessment measured the mutual information between each feature and the breast cancer diagnosis outcome. Mutual information is a statistical metric used to gauge the extent of dependency between variables. By calculating this metric for each feature, the study aimed to quantify how much information each feature contributed to the accurate prediction of breast cancer. Features with higher mutual information scores were deemed more significant, indicating their greater relevance and influence in enhancing the accuracy of breast cancer diagnosis models [33].

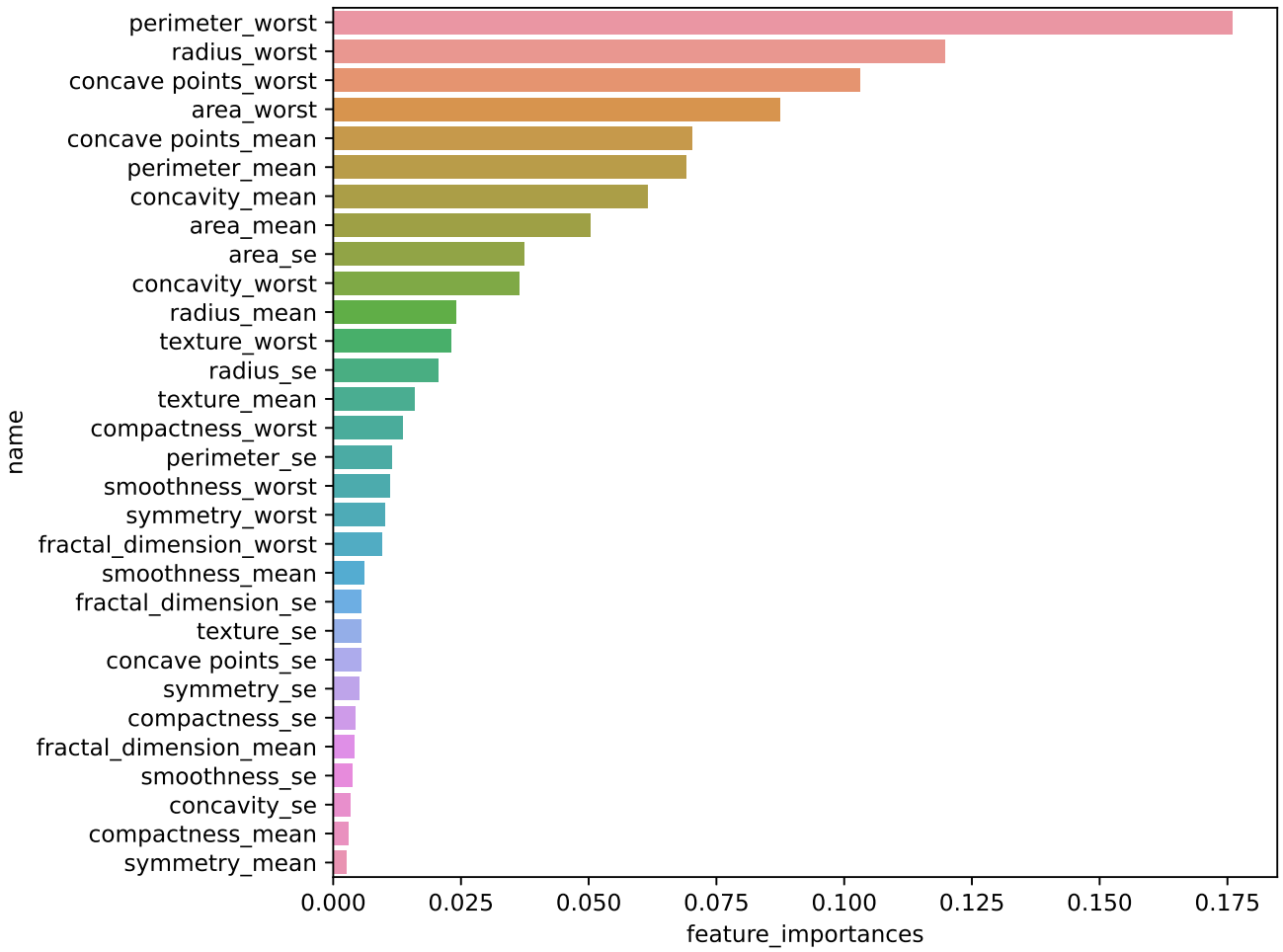

Fig. 5. Feature importance.

Fig. 5 illustrates the average impact of selected features on our model's performance. This analysis revealed that the three most influential features for breast cancer diagnosis are Perimeter, Radius, Concave Points, and Area Worst. These features significantly improved the model's precision and effectiveness in identifying breast cancer cases. Consequently, these feature rankings can be valuable for medical professionals in interpreting the model's predictions, ultimately contributing to developing an improved AI-driven healthcare ecosystem. This information empowers healthcare practitioners to make more informed decisions and enhance patient care in breast cancer diagnosis and treatment.

IV. Conclusion

This study proposes a robust classification approach based on the smoothed hinge loss function, which is solved using Adam's algorithm. This research aimed to develop an enhanced machine learning model to help medical technicians detect cancer cases early. It also will confirm and indicate if a person has breast cancer. Evaluation results showed that our approach is more effective than Decision Trees (DT), Random Forests (RF), Logistic Regression (LR), Support Vector Machines (SVM), Naive Bayes (NB), K-Nearest Neighbors (KNN), and XGBoost (XGB) models. Then, the prediction model developed in this work can be used to create a mobile application that will allow people to detect cancer cases early.

References

[1] C. Mattiuzzi and G. Lippi, "Cancer statistics: a comparison between world health organization (who) and global burden of disease (gbd)," *European Journal of Public Health*, vol. 30, no. 5, pp. 1026–1027, 2020.

[2] C. G. Yedjou, J. N. Sims, L. Miele, F. Noubissi, L. Lowe, D. D. Fonseca, R. A. Alo, M. Payton, and P. B. Tchounwou, "Health and racial disparity in breast cancer," *Breast cancer metastasis and drug resistance*, pp. 31–49, 2019.

[3] K. L. Britt, J. Cuzick, and K.-A. Phillips, "Key steps for effective breast cancer prevention," *Nature Reviews Cancer*, vol. 20, no. 8, pp. 417–436, 2020.

[4] E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan, "Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review," *Expert Systems with Applications*, vol. 167, p. 114161, 2021.

[5] A. R. Vaka, B. Soni, and S. Reddy, "Breast cancer detection by leveraging machine learning," *ICT Express*, vol. 6, no. 4, pp. 320–324, 2020.

[6] Z. Khoudi, M. Nachaoui, and S. Lyaqini, "Identifying the contextual factors related to the reading performance of moroccan fourth-grade students from a machine learning-based approach.," *Education and Information Technologies*, pp. 1–27, 2023.

[7] I. Thomassin-Naggara, L. Ceugnart, A. Tardivon, L. Verzaux, C. Balleyguier, P. Taourel, and B. Seradour, "Intelligence artificielle: Place dans le dépistage du cancer du sein en france," *Bulletin du Cancer*, 2022.

[8] M. M. Ahsan and Z. Siddique, "Machine learning-based heart disease diagnosis: A systematic literature review," *Artificial Intelligence in Medicine*, p. 102289, 2022.

[9] X. Zheng, L. Zhang, and L. Yan, "Ctsvm: a robust twin support vector machine with correntropy-induced loss function for binary classification problems," *Information Sciences*, vol. 559, pp. 22–45, 2021.

[10] S. Lyaqini and M. Nachaoui, "Identification of genuine from fake banknotes using an enhanced machine learning approach," in *International Conference on Numerical Analysis and Optimization Days*, pp. 59–70, Springer, 2021.

[11] S. Lyaqini, M. Nachaoui, and A. Hadri, "An efficient primal-dual method for solving non-smooth machine learning problem," *Chaos, Solitons & Fractals*, vol. 155, p. 111754, 2022.

[12] S. Lyaqini and M. Nachaoui, "Diabetes prediction using an improved machine learning approach," *Mathematical Modeling and Computing*, vol. 8, pp. 726–735, 01 2021.

[13] S. Lyaqini, M. Quafafou, M. Nachaoui, and A. Chakib, "Supervised learning as an inverse problem based on non-smooth loss function," *Knowledge and Information Systems*, vol. 62, no. 8, pp. 3039–3058, 2020.

[14] S. H. Haji and A. M. Abdulazeez, "Comparison of optimization techniques based on gradient descent algorithm: A review," *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 18, no. 4, pp. 2715–2743, 2021.

[15] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *International Conference on Computational Learning Theory*, pp. 416–426, Springer, 2001.

[16] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Wadsworth International Group, 1984.

[17] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[18] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 20, no. 2, pp. 215–232, 1958.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[20] I. Rish, "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, pp. 41–46, 2001.

[21] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[22] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

[23] C. Kingsford and S. L. Salzberg, "What are decision trees?," *Nature Biotechnology*, vol. 26, no. 9, pp. 1011–1013, 2008.

[24] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017.

[25] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[26] G. I. Webb, E. Keogh, and R. Miikkulainen, "Naïve bayes.," *Encyclopedia of machine learning*, vol. 15, pp. 713–714, 2010.

[27] O. Kramer, "K-nearest neighbors," in *Dimensionality reduction with unsupervised nearest neighbors*, pp. 13–23, Springer, 2013.

[28] U. K. Kumar, M. S. Nikhil, and K. Sumangali, "Prediction of breast cancer using voting classifier technique," in *2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, pp. 108–114, IEEE, 2017.

[29] A. F. M. Agarap, "On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset," in *Proceedings of the 2nd international conference on machine learning and soft computing*, pp. 5–9, 2018.

[30] H. Salem, G. Attiya, and N. El-Fishawy, "Early diagnosis of breast cancer by gene expression profiles," *Pattern Analysis and Applications*, vol. 20, pp. 567–578, 2017.

[31] R. Deepa, R. Kavipraba, G. Pavithra, S. Preethi, and S. R. AK, "Breast cancer classification using the supervised learning algorithms," in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1492–1498, IEEE, 2021.

[32] M. A. Naji, S. El Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouhahid, and O. Debauche, "Machine learning algorithms for breast cancer prediction and diagnosis," *Procedia Computer Science*, vol. 191, pp. 487–492, 2021.

[33] H. Zhou, X. Wang, and R. Zhu, "Feature selection based on mutual information with correlation coefficient," *Applied Intelligence*, pp. 1–18, 2022.