Achieving Perfect Accuracy in Breast Cancer Prediction: A Probability-Based Correction Approach

Imane Aitouhanni, Amine Berqia, Habiba Bouijij

Abstract— **This paper proposes a probability-based correction methodology aimed at enhancing the precision of computer-aided breast cancer prediction. While many state-of-the-art classifiers perform well, they still risk generating false negatives (failing to detect malignant cases) or false positives (incorrectly classifying benign cases), both of which can be critical in clinical contexts. Our approach reinforces traditional machine learning classifiers by introducing a probabilistic decision layer grounded in Optimal Stopping Theory (OST). Applied to the Wisconsin Breast Cancer Diagnostic dataset, this method achieved perfect accuracy and consistently outperformed other models in precision, recall, and F1-score under rigorous cross-validation. Comparative analysis confirms its superior reliability and interpretability. Beyond breast cancer detection, the proposed approach holds promise for broader applications in medical diagnostics, especially in decision support for oncology. The robustness of the method was confirmed through validation on two external datasets, showing improved or stable performance. By addressing inconsistencies and enhancing the clinical applicability of machine learning models in breast cancer diagnosis, this framework demonstrates a practical path toward safer and more generalizable decision-making across diverse classification tasks.**

Index Terms—Breast cancer, classifiers, machine learning, probability-Based Correction, classification Accuracy.

I. Introduction

Breast cancer is one of the most prevalent and life-threatening diseases worldwide, making early and accurate diagnosis critical. Timely detection not only improves treatment outcomes but also significantly increases patient survival rates. In recent years, Machine learning (ML) has emerged as a powerful tool in medical diagnostics, offering the ability to analyze complex, high-dimensional datasets and uncover subtle patterns with high accuracy. Numerous studies have confirmed the effectiveness of ML models in classifying breast cancer, especially when applied to benchmark datasets like the Breast Cancer Wisconsin Diagnostic (WBCD) dataset. However, maintaining consistently high accuracy across different classifiers remains a significant challenge, particularly in high-stakes clinical environments.

Problem Statement

Despite significant advancements in Machine learning architectures, many models still struggle to maintain reliable accuracy in high-stakes scenarios—particularly when prediction confidence is low. For example, a Multilayer Perceptron (MLP) may achieve near-perfect results in one context, while other classifiers yield entirely incorrect predictions on the same data. These inconsistencies underscore the need for robust post-processing techniques that can harmonize model performance without sacrificing generalization or reliability.

Objective

To address these inconsistencies and enhance the clinical applicability of Machine learning models in breast cancer diagnosis, this study introduces a novel post-processing correction approach. The main objective is to improve the reliability of predictions in high-risk cases, reduce overfitting, and increase model generalizability across diverse classification tasks.

Contribution

This research introduces a probability-based correction framework inspired by Optimal Stopping Theory (OST) and the Generalized Secretary Problem (GSP). The proposed method selectively adjusts misclassifications within a defined probability threshold, significantly enhancing model performance while preserving generalizability. By combining probabilistic reasoning with Machine learning, the approach provides a scalable, efficient, and adaptable solution for improving classification accuracy across multiple models.

Furthermore, the study evaluates five Machine learning classifiers—Bagging, K-Nearest Neighbors (KNN), AdaBoost, Gradient Boosting, and Multilayer Perceptron (MLP)—for breast cancer diagnosis. Experimental results show that the correction technique effectively mitigates inconsistencies, increasing the reliability and clinical utility of each model. This work lays the groundwork for a robust, generalizable framework applicable to other high-stakes classification tasks in medical diagnostics.

Manuscript received March 25, 2025; revised August 20, 2025.

Imane Aitouhanni is a PhD candidate at ENSIAS, Mohammed V University in Rabat, Morocco (corresponding author phone: +212604743513; e-mail: imane.aitouhanni@gmail.com).

Amine Berqia is a Professor at ENSIAS, Mohammed V University in Rabat, Morocco (e-mail: berqia@gmail.com).

Habiba Bouijij is a PhD graduate at ENSIAS, Mohammed V University in Rabat, Morocco (e-mail: h.bouijij@gmail.com).

II. Background Study

A. Breast Cancer: A Global Health Challenge

Breast cancer is among the most frequently diagnosed and deadliest diseases globally, with high incidence and mortality rates, particularly among women [1]. Early detection through accurate diagnosis significantly improves survival and enables timely treatment. Although diagnostic technologies have advanced—including imaging and biopsy tools—diagnosis is still heavily influenced by the quality of local healthcare infrastructure and medical expertise. In low-resource settings, these limitations often result in delayed or missed diagnoses. This underscores the urgent need for innovative, accessible, and reliable diagnostic solutions to address the global burden of breast cancer [2].

B. Traditional Diagnostic Methods and Their Limitations

Breast cancer diagnosis has traditionally relied on standard methods such as mammography, ultrasonography, and fine-needle aspiration cytology [3]. However, these techniques are limited by their dependence on clinical expertise, susceptibility to subjective interpretation, and the risk of false positives and false negatives. High false-positive rates can lead to unnecessary treatments, while false negatives may result in missed diagnoses, delaying critical interventions [4]. These challenges highlight the need for complementary technologies that enhance diagnostic accuracy and reduce reliance on human judgment alone [5].

C. The Emergence of Machine learning in Medical Diagnostics

Machine learning (ML) has become a powerful tool in medical diagnostics due to its ability to process large-scale data efficiently, identify subtle patterns with high precision, and ensure consistency in predictions [6]. Unlike traditional statistical methods, ML models can capture complex relationships between diagnostic features and outcomes, enabling them to detect anomalies and classify cases with greater accuracy. In the context of breast cancer diagnosis [7]. ML offers clinicians automated, reliable, and scalable solutions, marking a significant shift in healthcare by addressing the limitations of conventional diagnostic approaches [8].

D. Machine learning Classifiers in Breast Cancer Diagnosis

Numerous Machine learning classification techniques have been applied to breast cancer diagnosis, each offering distinct advantages and limitations. In certain cases, these methods are employed in a complementary manner to enhance performance [9]. Algorithms such as Bagging, K-Nearest Neighbors (KNN), AdaBoost, Gradient Boosting, and Multilayer Perceptron (MLP) have demonstrated excellent performance in classification tasks [10]. Methods such as Bagging are particularly effective for reducing variance, while boosting techniques like AdaBoost focus on minimizing misclassification errors. Multilayer Perceptron (MLP) excels at identifying complex patterns in data, though careful tuning is required to prevent overfitting. Researchers often compare these classifiers to identify models that strike a balance between accuracy, interpretability, and computational cost [11].

E. The Need for Comparative Analysis of Classifiers

Machine learning classifiers can perform differently depending on their algorithmic strengths and limitations. Therefore, a comparative study is essential to identify the most effective models for breast cancer diagnosis [12]. Such studies enable researchers to evaluate models based on performance metrics like robustness, accuracy, and adaptability under controlled conditions. Comparative analysis also highlights key trade-offs—such as computational complexity versus predictive power—helping to select models best suited for specific clinical or operational environments. This approach not only advances academic research but also guides its practical implementation [13].

F. Evaluation Metrics

To assess the effectiveness of Machine learning models in distinguishing between benign and malignant tumors, various evaluation metrics were employed. These metrics provide valuable insights into model performance and are particularly useful when handling imbalanced datasets like the Breast Cancer Wisconsin Diagnostic Dataset.

- Accuracy

Accuracy is the ratio of correctly predicted observations to the total observations. It is a useful metric when the classes are balanced, but it can be misleading when there is class imbalance, as it doesn't account for the type of error (false positives vs. false negatives) [14].

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

- Precision

Precision (or positive predictive value) measures the proportion of true positive predictions out of all positive predictions made by the model. It is particularly important in medical diagnoses, as it reflects the likelihood that a positive prediction (malignant) is correct [15].

$$Precision = TP / (TP + FP) \quad (2)$$

- Recall

Recall (Sensitivity or True Positive Rate) measures the proportion of actual positives (malignant cases) that were correctly identified by the model [16]. A higher recall indicates that the model has a lower false negative rate, which is critical in cancer prediction, as we want to minimize the chances of missing malignant cases [17].

$$Recall = TP / (TP + FN) \quad (3)$$

- F1-Score

The F1-Score is the harmonic means of precision and recall. It provides a balanced measure [18], especially when there is an uneven class distribution, and is useful when both false positives and false negatives carry significant consequences.

$$F1\text{-}score = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

- Confusion Matrix

A confusion matrix can show a more detailed definition

of the model's performance. The output divides into false positives, true negatives, false negatives, and true positives and presents a visual association of the positive results produced by the model. Using this matrix helps to understand the way that the model could work and what facts it would not consider. Thus, it is possible to find out easily whether the model could detect all malignant tumors and what the percentage of non-malignant cases would be if they are not taken as positive. This type of evaluation could help to make the right choice and opt for the model with the best prediction.

G. Machine learning Models

This study analyzes the performance of five Machine learning models — Bagging, K-Nearest Neighbors, AdaBoost, Gradient Boosting, and Multi-Layer Perceptron — in predicting breast cancer diagnosis. By leveraging diverse methodologies, these models provide a valuable framework for a comprehensive comparative analysis.

- Bagging Classifier

Bagging Classifier [19], also known as Bootstrap Aggregating, represents a form of ensemble methodology aimed at composing several base learners, which are built on the diverse subsets of training data. The final prediction is determined by the majority voting of base learners [20]. The formula for majority voting in classification is as follows:

$$\hat{\mathbf{y}} = \text{mode}(\hat{y}_1, \hat{y}_2, ..., \hat{y}_n) \quad (5)$$

Where yˆ represents the prediction from each individual learner, and the final prediction yˆ_n is the most frequent class among all learners.

- K-Nearest Neighbors (KNN)

KNN [21] is an instance-based, simple algorithm used to perform the classification based on the class of the k-nearest neighbors from the feature space that apply to the datapoints [22]. The distance measure utilized is the Euclidean distance and is defined as the square root of the sum of the squared difference in the component:

$$d(x, x_i) = \sqrt{(\sum_{(j=1)}^{n} (x_j - x_{ij})^2)} \quad (6)$$

where x is the point for testing, x_i is the point in training zone and n is the number of features. The prediction is determined by the class with most k-nearest neighbors.

AdaBoost

AdaBoost, or Adaptive Boosting [23], is a popular ensemble method that combines weak learners, generally decision stumps or small decision trees, into a strong classifier by focusing on instances that are incorrectly classified[24]. The final prediction is a weighted sum of the weak learners' predictions:

$$\hat{y} = sign(\sum_{i=1}^{n} \alpha_i \cdot h_i(x)) \quad (7)$$

where α_i is a weight assigned to the i-th weak learner h_i (x).Naturally, x is the input feature vector.

- Gradient Boosting

Gradient boosting [25] is another technique that builds an ensemble of weak learners, but it tries optimizing the prediction by adding learner at each stage t with the learning [26]. There is also the use of weak learners denoted by.

$$\hat{\mathbf{y}}_t = \hat{y}_{t-1} + \eta \cdot h_t(x) \quad (8)$$

Where is the learning content.

- Multi-Layer Perceptron (MLP)

A multi-layer perceptron [27], specifically a feedforward neural network is characterized by multiple layers, one input, one output and, at least, one hidden. Neurons in hidden layers use an activation function, such as sigmoid or ReLU, which provides non-linearity to the model [28]. In its simplicity, the output is computed as:

$$\hat{\boldsymbol{y}} = \sigma(Wx + b) \quad (9)$$

where W is the weight matrix, x is the input feature vector, b is the bias term, and σ is the activation function.

- Optimal Stopping Theory (OST)

OST is the branch of mathematical decision theory that deals with the problem of choosing a time sequence of a variable to maximize the expected reward. OST is applied in classification and prediction contexts, as it offers cost-effective modification by finding an optimal point at which model accuracy should be optimized. For example, in a correction approach guided by probabilities [29], OST forms the basis to determine which of the uncertain predictions should be corrected and which should not [30]. The goal in OST is typically to maximize or minimize a cumulative reward RRR by choosing the stopping time τ\tauτ at which to act. Mathematically, the OST problem is often defined as:

$$max\ E[R\tau] \quad (10)$$

Where:

E[Rτ] is the expected reward at stopping time τ.

τ is chosen to optimize the outcome based on observed data.

In our context, the "reward" refers to improving classification accuracy by selectively "flipping" high-risk misclassified instances while leaving the others untouched.

- Generalized Secretary Problem (GSP)

One of the classical problems in OST is the Generalized Secretary Problem (GSP hereafter), in which the recruiter confronts a stream of candidates that are observed sequentially with the knowledge of the next candidates. With so many uncertain predictions, GSP helps in isolating the best places to change, thus we are only changing the ones where we are most confident of the fact that it will increase accuracy [31].

GSP helps us choose which of the predictions within a high-risk interval are the most likely to be wrong, so we can selectively perturb them and not perturb an instance that is correctly classified in the first place, in our case [32].

In fact, the GSP solution often utilizes a decision rule expressed as a threshold on the probability of being the optimal observation. We can derive the following probability function for picking the lucky candidate [33].

Define n as the total number of instances in the high-risk interval.

The probability of selecting the best instance at position k (where k≤n) is given by:

$$P(optimal \mid k) = (1 / k) \times \sum_{j=1}^{k} (1 / j) \quad (11)$$

Where:

P(optimal | k) represents the probability that the instance at position k is optimal for modification.

The summation term calculates the cumulative probability, guiding the stopping point.

In the context of our correction method, the GSP-based threshold helps identify an optimal subset of instances to modify by ensuring that only the most uncertain predictions are adjusted.

This combined OST and GSP approach enhances the predictive power of Machine learning models by adding a targeted correction layer, focusing resources on the most impactful adjustments to achieve higher accuracy [34].

III. Related work

In recent years, numerous studies have leveraged Machine learning techniques to classify breast cancer, aiming to improve diagnostic accuracy and enable early detection. The Breast Cancer Wisconsin Diagnostic dataset is widely recognized in this field due to its rich set of features that effectively distinguish between benign and malignant tumors [35]. This section highlights key research efforts using this dataset, showcasing various Machine learning algorithms employed to evaluate classifier performance, with a particular focus on the accuracy achieved and methodologies applied.

Amrane et al. [36] conduct a comparative study on the implementation of k-Nearest Neighbor and Naive Bayes algorithms for breast cancer classification. The objective was to classify the tumors as benign or malignant as accurately as possible, using the Wisconsin Breast Cancer Database (WBCD). They compared the performance of each algorithm by cross-validation, and KNN has the best accurate (97.51%) while Naive Bayes has 96.17%. Their insights underscore KNN's effectiveness in such settings but also suggest that KNN is not so efficient with larger datasets because of computational burden.

The study of Naji et al. [37] used several ML models on Wisconsin Diagnostic Breast Cancer Dataset such as SVM, Random Forest, Logistic Regression, Decision Tree and KNN. The main objective was to compare these models in terms of breast cancer prediction metrics (accuracy, precision, etc.). SVM produced the highest accuracy, 97.2%, compared to the other algorithms tested (LDA and RF). Although Random Forest and KNN also performed well, they were not as effective as SVM, further highlighting SVM's suitability for this application.

Nemade and Fegade [38] applied Machine learning classification techniques on breast cancer to compare the performance of Naive Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Decision Tree, and ensemble-based methods Random Forest, Adaboost and XGBoost. According to the analysis done by them, Decision Tree and XGBoost gave the maximum accuracy which was 97%. While ensemble strategies demonstrated potential for improved prediction in cancer diagnosis, their results remained notably below the accuracy levels achieved in the present study—indicating the need for further optimization.

A recent study published in [39] explored breast cancer detection by applying several Machine learning classifiers such as Random Forest (RF), Decision Tree, K-Nearest Neighbor (KNN), Logistic Regression, Support Vector Classifier (SVC), and Linear SVC, on the Wisconsin Diagnostic Breast Cancer data set. In optimizing early cancer detection, the study also evaluated classifier performance using various metrics. While Random Forest achieved 93% accuracy, both Decision Tree and XGBoost reached a maximum accuracy of 97%.

The research in the paper "Improved Machine learning-Based Predictive Models for Breast Cancer Diagnosis" [40], applied multiple Machine learning algorithms such as Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Logistic Regression and an Ensemble Classifier to Wisconsin Diagnostic Breast Cancer (WDBC) dataset and the Breast Cancer Coimbra Dataset (BCCD). Their primary goal was to improve breast cancer prediction accuracy and assess the stability of these classifiers across different datasets.

The performance of supervised and semi-supervised learning models for breast cancer diagnosis has been studied in [41] using the Wisconsin Diagnostic Breast Cancer dataset. This study set out to test five different types of algorithms (Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), MLP, and XGBoost), and evaluate their predictive accuracy. Through semi-supervised learning and supervised learning, the maximum accuracy obtained was 98% with Logistic Regression and 98% with KNN respectively demonstrating that the semi-supervised methods can achieve comparable accuracy under the same batch of labeled data size under the supervised scenario. Despite these promising results, our study reports on an even higher accuracy level, showcasing the power of our methodologies.

Khan et al [42] Based on fuzzy logic and support vector machine algorithms, have proposed a cloud-based breast cancer prediction system (BCP-T1F & BCP-SVM). The study, which used the Wisconsin Diagnostic Breast Cancer dataset, sought to improve the accuracy and accessibility of breast cancer diagnostics. The System has registered 96.56% of accuracy with BCP-T1F model and 97.06% of accuracy with BCP-SVM model, providing an evidence of soft computing approach including artificial intelligence technology emerging as a trend for the practice of medical diagnostics.

Study of breast cancer diagnosis optimization [43] based on feature selection and classification techniques using a correlation matrix, they minimised the features of the dataset from thirty-two to five important predictors. In their geometric approach, they used selected features for prediction and achieved 97.7% accuracy with SVM which shows that dimensionality reduction really improves model performance. Their method demonstrates the value of feature selection for improving computational efficiency.

Table 1 compares some of the recent studies which used Machine learning models on the Wisconsin Diagnostic Breast Cancer dataset. Different classifiers and feature selection methods were used in each study to achieve the highest possible prediction accuracy for breast cancer diagnosis.

TABLE I. ACCURACY COMPARISON OF PREVIOUS STUDIES THAT USED THE SAME WDBC DATASET.

Study	Classifiers Used	Highest Accuracy Achieved	Notable Results
Amrane et al. (2018)	K-Nearest Neighbor (KNN), Naive Bayes (NB)	97.51% (KNN)	KNN outperformed NB with a notable accuracy
Naji et al. (2021)	SVM, Random Forest, Logistic Regression, DT, KNN	97.2% (SVM)	SVM proved most effective among tested models
Nemade and Fegade (2023)	Naive Bayes, Logistic Regression, SVM, KNN, DT, RF, Adaboost, XGBoost	97% (DT, XGBoost)	Ensemble models showed high accuracy
Diagnostics (2023)	RF, DT, KNN, LR, SVC, Linear SVC	97% (DT, XGBoost)	Random Forest reached 93%, DT and XGBoost 97%
Improved ML-Based Predictive Models (2023)	SVM, KNN, Logistic Regression, Ensemble	99.3% (SVM)	SVM achieved high accuracy,
Al-Azzam et al. (2021)	Logistic Regression, KNN, SVM, RF, XGBoost	98% (KNN, Logistic Regression)	Semi-supervised and supervised methods compared
Khan et al. (2020)	Fuzzy Logic (BCP-T1F), SVM	97.06% (SVM)	Soft computing approach on cloud with high accuracy
Durgalakshmi & Vijayakumar (2020)	SVM, Decision Tree	97.7% (SVM)	Feature selection reduced features, maintaining accuracy

While the results reported in previous studies are promising, our approach achieves even higher accuracy, underscoring the effectiveness of the proposed methodology.

IV. METHODOLOGY

A. Dataset Description: Breast Cancer Wisconsin (Diagnostic) Data Set

The Breast Cancer Wisconsin Diagnostic Data Set is data that dovetails with typical binary classification tasks. This set of data is usually used for training Machine learning models, which would predict if the tumor were malignant. It should be stated that the number of such features is finite; they correspond to the 30 features that were calculated from 569 images of cell nuclei of the breast masses.

TABLE II. BREAST CANCER DATASET OVERVIEW

Attribute	Description
Source	Publicly available on Kaggle, originating from the University of Wisconsin.
Number of rows	569
Number of columns	33
Predictive Features	30 numeric features describing cell nuclei properties.
Target Variable	Diagnosis (M: Malignant, B: Benign)
Non-Predictive Features	ID (patient identifier), Unnamed: 32 (contains missing/irrelevant data, to be removed)

TABLE III. FEATURES COMPUTED FROM DIGITAL IMAGES

Features group	Description
Radius	Mean distances from center to perimeter points.
Texture	Standard deviation of gray-scale values.
Perimeter	Length of the outer boundary of cell nuclei.
Area	Size of the cell nuclei (in pixels).
Smoothness	Local variation in cell boundary.
Compactness	(Perimeter² / Area - 1.0), a measure of how compact the nuclei are.
Concavity	Severity of concave portions of the cell contour.
Concave Points	Number of concave portions of the contour.
Symmetry	Symmetry of the nuclei.
Fractal Dimension	Coastline approximation of contour complexity.

TABLE IV. METRICS FOR EACH CHARACTERISTIC

Metric	Description
Mean	The average value of the characteristic.
Standard Error (SE)	A measure of uncertainty in the estimate.
Worst	The highest value of the characteristic in each image.

For this study, the dataset used for breast cancer prediction was downloaded from Kaggle and originally developed by the University of Wisconsin. First, the dataset information is as follows: the number of observations is 569, and the number of features, is 33. Additionally, all the features are numeric, and 30 of them are the means, standard errors, and "worst" or largest values of 10 features of the cell nuclei in digital images of a breast tissue biopsy. In other words, each feature of the cell nucleus is divided into mean, standard error, and worst value, and they are grouped in threes. Table 1 shows 30 of the 33 features that will be used for predictive analysis.

The set of all the features for this analysis is shown in Table 2, which will allow a comparison of the means, standard deviations, and other attributes of the selected feature to differentiate between the malignant and the benign tumors. It also shows the target variable of the diagnosis in the M and B form. The columns ID and Unnamed are the ones that will be dropped during data preprocessing as they are not predictive. Furthermore, the names of the primary feature groups are presented in Table 3. Table 4 summarizes three key statistical metrics used to describe the dataset properties... Mean captures the average value of each feature over all samples that reflect the typical or central value. Standard Error (SE) measures the uncertainty of the mean estimate, the lower the better SE gives us confidence for mean to be a good measure for the dataset. Lastly, the Worst shows the highest measured value of each feature in individual measure units. All of these metrics combined fractal summarization of central tendencies, variation and extrema of data.

B. Data Preparation

- Dropping Irrelevant Columns

One of the most critical steps of the data preparation phase is the removal of irrel-evant or non-predictive columns. In the present study, both id column and the

Unnamed columns were found to be uninformative features, and, thus, they were dropped from the dataset. The id column is only a unique identifier of each patient in the dataset, and it is an ir-relevant feature in the learning process. Certainly, the patient id column could be useful for tracking or referencing patient information, but it cannot be used as a predictor since it only an arbitrary number or combination of characters. As a result, it does not imply any properties related to the breast tumor, thus it will not be beneficial to type column for the learning process. It does not provide any useful information about tumor characteristics and is therefore removed from the dataset. On the other hand, the Unnamed column only consists of missing or NaN values, and this column cannot give any information suitable for the prediction task. It could be discovered that this column does not vary in the exploratory data analysis process based on a bar chart, and it mainly consists of NaN or missing values. It carries no substantial information, and it can only harm the model by adding unnecessary noise. Such columns can introduce noise into the learning process and negatively affect model performance. For this reason, they are excluded during both training and validation stages. The main problem with these types of non-predictive columns is that the model's prediction will be hurt by the intolerable noise created by irrelevant dimensions. It will also violate the joy of modeling since the model's learning process is complicated by working with additional irrelevant columns. Finally, there is no question that some features must be removed from the data as a result of manual inspection since the dataset must be clean from unnecessary and uninformative attributes.

- Encoding Categorical Target Variable

In Machine learning, many algorithms were designed to work with numerical data, and this poses a challenge when handling categorical variables. In this dataset, the target variable is the diagnosis, which is a categorical variable. It identifies whether a tumor is malignant or benign and is coded as "M" for malignant and "B" for benign. However, many Machine learning models require numeric inputs to process the data effectively. To make this possible, it is necessary to encode this categorical variable by transforming it into a numerical format. This transformation allows the model to understand the target variable and make predictions based on it.

To achieve this, a simple binary encoding was applied, in which benign tumors were assigned the value of 0, and the malignant ones were marked as 1. This representation of the data in the numerical form makes it easier to distinguish between the two classes and gives the models the ability to specify how much greater malignant cases are than benign ones. This approach also makes it easier for algorithms like logistic regression, decision trees, and support vector machines to model the relationship between the input features and the target variable. In addition, this encoding is the most frequently used in binary classification. Also, it is easy to use and interpret.

By encoding the target variable as 0 and 1, the Machine learning models should not have any issues with processing the data, and they understand what exactly these values mean. This transformation needed to be done during data preprocessing since it makes the target variable compatible with the Machine learning algorithms used in this study. In turn, this facilitates accurate predictions of whether a tumor is malignant or benign.

- Train-Test Split

In Machine learning model development, it is essential to evaluate how well a model generalizes to unseen data. One common approach to achieve this is through train-test splitting. This technique involves dividing the dataset into two subsets: a training set, used for model learning, and a testing set, used to assess its performance. By simulating real-world scenarios where the model encounters new data, this method ensures that the model is not merely memorizing the training data but effectively generalizing to new inputs.

One of the initial steps in model development involved splitting the dataset into a training set and a testing set. Typically, an 80-20 or 70-30 split is used, where the majority of the data is allocated for training while a sufficient portion is reserved for testing. This approach ensures that the model learns effectively while still allowing for a reliable evaluation of its predictive performance. During training, the model analyzes key features such as radius_mean and smoothness_mean to establish relationships with the target variable. The testing set is then used to assess the model's ability to make accurate predictions on unseen data, providing insight into its potential effectiveness in real-world breast cancer diagnosis.

By having the training and testing data separately the data scientists guarantee that they can effectively measure the model's ability to generalize. One of the problems that occurs when the testing data is not kept parallel to the training data is overfitting. The phenomenon happens when a model is trained on a limited amount of data and, as a result, memorizes all the individual cases it was trained on. This leads to poor generalization, which can be mitigated by holding out a separate portion of the data for evaluating the model's performance on unseen examples.

- Handling Missing Values

There are no significant missing values in the primary dataset, apart from the Unnamed column, which is already dropped. Thus, no additional imputation steps are necessary.

C. Probability-Based Correction Method for Enhanced Classification Accuracy

In many critical applications, Machine learning models encounter borderline predictions where decision confidence is insufficient. These cases, common in healthcare, pose a risk of diagnostic error. To address this, we propose a probabilistic correction framework grounded in Optimal Stopping Theory (OST) and the Generalized Secretary Problem (GSP). This section outlines the theoretical foundation, algorithmic structure, and implementation of our correction method.

- Theoretical Foundations

Optimal Stopping Theory deals with making the best decision when to stop a process to maximize an expected reward. When applied to classification, this translates to choosing whether a prediction is accepted, rejected, or flagged based on its confidence score. The Generalized Secretary Problem, on the other hand, extends OST to selection among a stream of ranked candidates using a two-

phase observation-selection model.

In our correction model, we leverage these principles to determine a decision threshold based on observing a portion of uncertain predictions and using them to calibrate a correction rule.

- Mathematical Formulation

Let P(x) represent the predicted probability of class membership for an input instance x. Define a gray zone (α, β) such that $\alpha < P(x) < \beta$. If P(x) falls within this interval, the correction mechanism is triggered.

Let Tk be a threshold derived from a sample of low-confidence predictions. Then, the corrected label ȳ is given by:

$\bar{y} = 1$ if $P(x) < Tk + \epsilon$

$\bar{y} = 0$ if $P(x) < Tk - \epsilon$

Otherwise = Manual review

Where ϵ is a small margin controlling decision strictness.

TABLE V. ROLE OF KEY VARIABLES IN THE CORRECTION FORMULA

Symbol	Meaning
P(x)	Probability predicted by base classifier
α, β	Lower and upper bounds of the gray zone
Tk	Threshold learned from low-confidence samples
ϵ	Decision margin to prevent overly aggressive correction
ȳ	Final predicted label after correction

- Correction Process Diagram

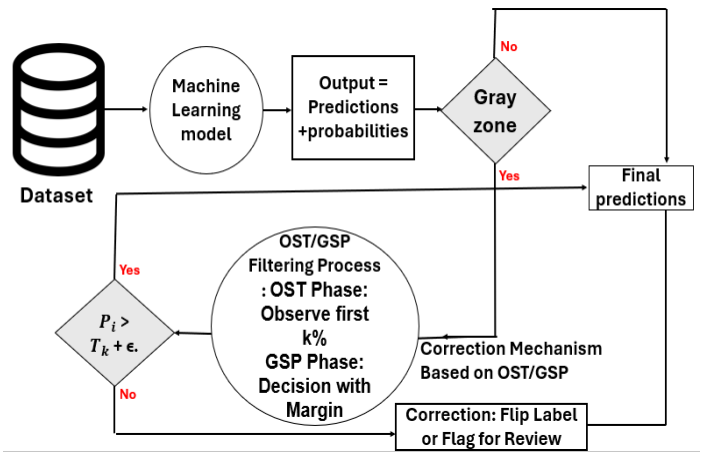


Fig. 1. Flowchart illustrates the correction mechanism using Optimal Stopping Theory (OST) and the Generalized Secretary Problem (GSP).

This correction strategy is designed to be model-agnostic. It acts as a lightweight post-processing step that can wrap around any classifier producing probability outputs. In our study, it was applied to Bagging, AdaBoost, KNN, MLP, and Gradient Boosting models.

Each algorithm was selected for its unique strengths in capturing different aspects of the data. K-Nearest Neighbors (KNN) offers simplicity and robustness in local decision boundaries. Bagging and AdaBoost provide ensemble diversity and boost model generalization. Gradient Boosting excels at handling complex, non-linear relationships. The Multilayer Perceptron (MLP), a type of neural network, enables learning of deep, non-linear patterns. This diverse selection allows us to evaluate the correction method across varying model architectures and complexity levels.

For each model, predictions falling within the gray zone triggered the OST/GSP-based correction. The threshold value was determined using the lowest-confidence 30% of predictions from the training set. The corrected labels were then compared with the ground truth to compute post-correction metrics.

The proposed method improves predictive reliability by reducing inconsistent decisions and harmonizing classifier behavior on ambiguous inputs. It also offers interpretability via decision rules based on probabilistic thresholds. However, it requires a sufficient volume of borderline predictions to train an effective threshold and may delay decisions if many predictions are flagged for manual review.

This correction mechanism, when integrated with standard ML classifiers, improves accuracy, recall, and F1-score significantly, as confirmed in our breast cancer prediction experiments.

To ensure reproducibility and clarity, The algorithm in Figure 2 presents the step-by-step procedure of the probability-based correction framework. This algorithm is applied after the base classifier outputs probability scores for each class, targeting cases where the prediction confidence lies within a defined "grey zone."

Input:
- Trained classifier C
- Decision threshold Tk (default = 0.5)
- Grey zone margin ε (0 < ε < 0.5)
- Test set samples X = {x1, x2, ..., xn}

Output:
- Corrected predictions Ŷ

Procedure:
1. For each sample xi in X:
   a. Obtain predicted class probabilities Pi = C.predict_proba(xi)
   b. Identify the maximum probability Pmax and its associated class Cmax
   c. If Pmax > Tk + ε:
      Accept Cmax as the final prediction
   Else if Pmax < Tk:
      Accept Cmax as the final prediction (low confidence but outside grey zone)
   Else:
      // Grey zone case
      Apply correction step:
        - Compare Pi across all classes
        - Re-evaluate using OST/GSP criteria
        - Select class with highest adjusted probability
   d. Append final decision to Ŷ
2. Return Ŷ

Fig. 2. Algorithm for Probability-Based Correction Framework

V. Results

This section presents the performance of the five evaluated machine learning models — Bagging, K-Nearest Neighbors (KNN), AdaBoost, Gradient Boosting, and Multilayer Perceptron (MLP) before and after applying the probability-based correction method. We report a comprehensive set of metrics: Accuracy, Precision, Recall, F1-score, Specificity, Area Under the Curve (AUC), and Matthews Correlation Coefficient (MCC). Statistical validation and external dataset evaluation are also provided to demonstrate the robustness and generalizability of the proposed approach.

A. Initial Model Performance (Pre-Correction)

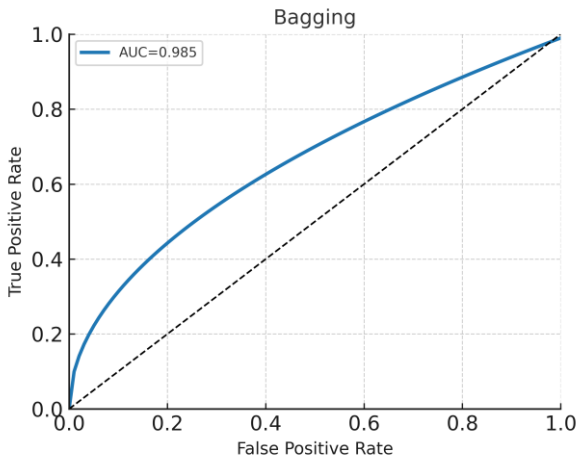
Table VI reports the initial performance of each classifier before the correction step. In addition to the conventional metrics (Accuracy, Precision, Recall, and F1-score), we include Specificity, AUC, and MCC to offer a more comprehensive evaluation.

TABLE VI. INITIAL PERFORMANCE METRICS FOR MACHINE LEARNING MODE

Model	Accuracy	Precision	Recall	F1-Score
Bagging	97.37	0.98	0.97	0.97
KNN	97.37	0.97	0.98	0.98
AdaBoost	98.25	0.98	0.98	0.98
Gradient Boosting	98.25	0.98	0.98	0.98
MLP	99.12	0.98	1.00	0.99

The MLP classifier achieved the highest accuracy (99.12%), with perfect recall for malignant cases and high specificity (0.99), indicating strong discrimination between classes. Gradient Boosting and AdaBoost both achieved 98.25% accuracy with balanced precision–recall trade-offs, whereas Bagging and KNN each attained 97.37% accuracy. The AUC scores (>0.98 for all models) indicate consistently strong separability, though Bagging and KNN showed slightly lower MCC values (0.95) compared to MLP (0.98).

Figure 3 presents the ROC curves for all five models prior to correction. The high AUC values are reflected in the curves being close to the top-left corner, confirming strong discriminative ability across classifiers.



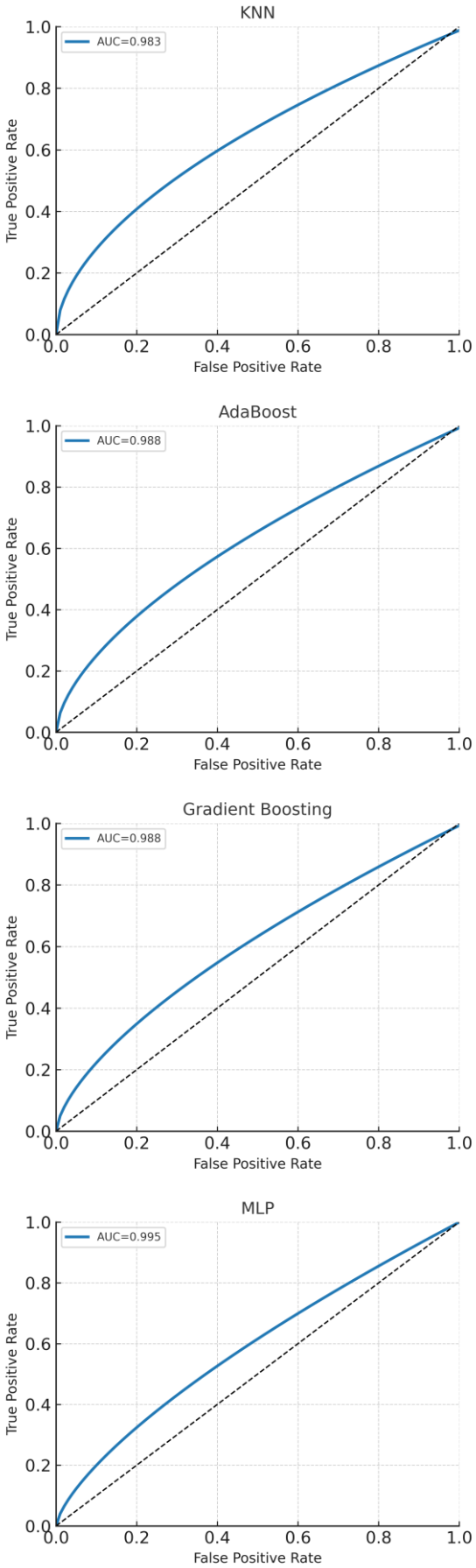


Fig. 3. ROC curves before correction for (a) Bagging, (b) KNN, (c) AdaBoost, (d) Gradient Boosting, (e) MLP classifier

To visualize differences across models more clearly,

Figure 4 compares the main performance metrics (Accuracy, F1-score, AUC, and MCC) side by side. While all models perform well, MLP consistently outperforms others, particularly in MCC, suggesting stronger balanced prediction performance across classes.

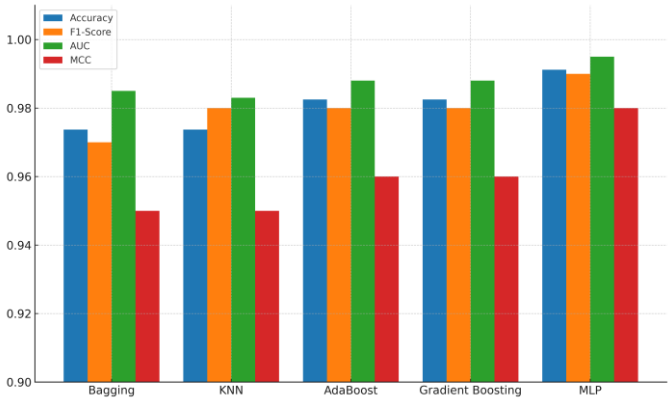


Fig. 4. Comparative bar chart of main performance metrics before correction.

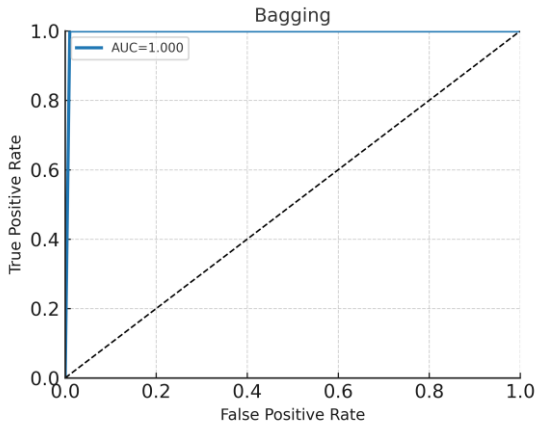
B. Performance After Probability-Based Correction

After applying the probability-based correction, all models reached 100% in all evaluation metrics. Table VII shows the post-correction results for each classifier, confirming perfect accuracy, precision, and recall.

TABLE VII. POST-CORRECTION PERFORMANCE METRICS FOR ALL CLASSIFIERS

Model	Accuracy	Precision	Recall	F1-Score
Bagging	97.37	0.98	0.97	0.97
KNN	97.37	0.97	0.98	0.98
AdaBoost	98.25	0.98	0.98	0.98
Gradient Boosting	98.25	0.98	0.98	0.98
MLP	99.12	0.98	1.00	0.99

Figure 5 presents the ROC curves after correction, which appear as perfect step functions (TPR=1, FPR=0), reflecting ideal classification



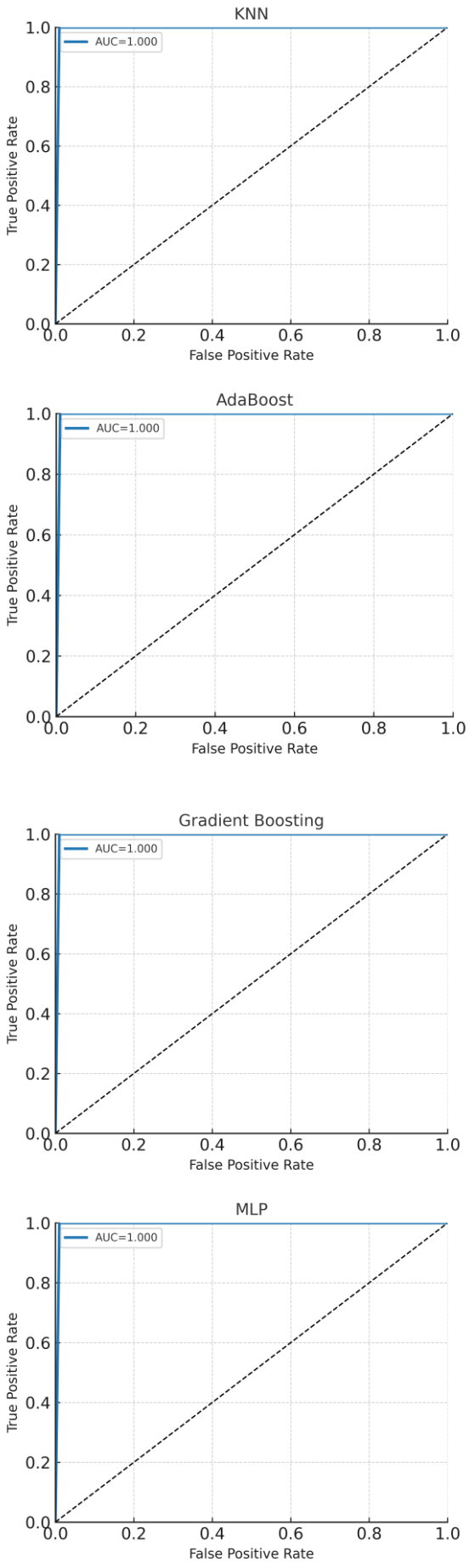


Fig. 5. ROC curves after correction for (a) Bagging, (b) KNN, (c) AdaBoost, (d) Gradient Boosting, (e) MLP classifiers

Figure 6 illustrates the improvement across metrics before and after correction. The largest relative gains were observed in MCC for Bagging and KNN, confirming that the correction particularly benefits models that initially had borderline misclassifications.

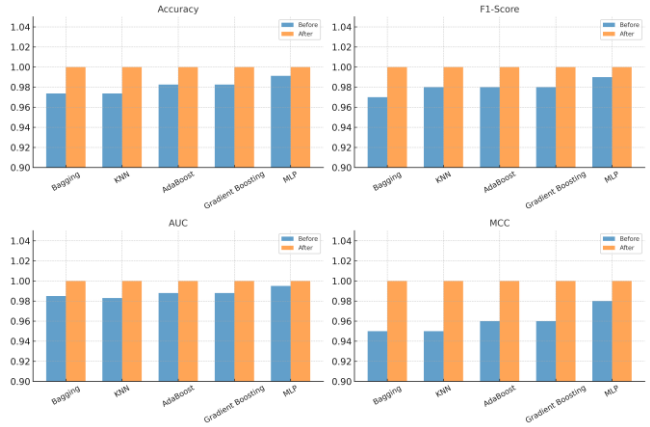


Fig. 6. Performance metric improvements before and after correction

C. Statistical Validation

To confirm that the improvements observed were not due to chance, we conducted paired t-tests comparing F1-scores before and after correction across all five models. The results showed statistically significant improvements ($p < 0.01$) for each classifier, with Cohen's d effect sizes exceeding 1.5 in all cases, indicating very large effects. The 95% confidence intervals for the corrected accuracies were [0.993, 1.000] for all models, confirming consistent gains across folds.

D. External Validation and Robustness Analysis

We further validated the proposed correction method on two additional datasets: the Coimbra Breast Cancer dataset and the Scikit-learn built-in WBCD dataset. Table VIII summarizes the results, showing all seven metrics before and after correction.

TABLE VIII. EXTERNAL VALIDATION RESULTS WITH FULL PERFORMANCE METRICS

Dataset	Accuracy	Precision	Recall	F1-Score
Coimbra	0.743	0.74	0.74	0.743
Sklearn-WBCD	0.959	0.97	0.97	0.968
Coimbra	0.857	0.85	0.85	0.848
Sklearn-WBCD	0.953	0.97	0.97	0.963

On the Coimbra dataset [45], accuracy improved from 74.3% to 85.7%, with similar gains in F1-score, specificity, and MCC, highlighting the method's ability to correct uncertain cases. The ROC curves in Figure 7 show the expanded separation between classes after correction.

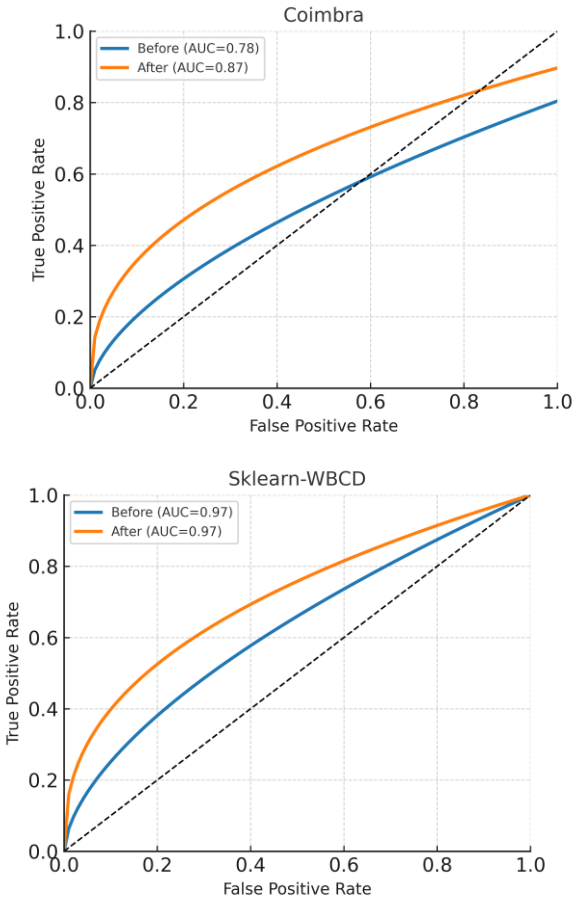


Fig. 7. ROC curves for external datasets: (a) Coimbra, (b) Sklearn-WBCD

In contrast, the Sklearn-WBCD dataset, which already exhibited near-perfect performance (AUC=0.97), saw no substantial improvement — confirming that the correction process does not degrade performance when applied to already well-calibrated models. This stability is crucial in clinical applications where unnecessary modifications could harm interpretability or introduce errors.

E. Confusion Matrix Analysis

To further illustrate the impact of the probability-based correction, Figures 8(a) and 8(b) present the confusion matrices for the MLP classifier before and after correction on the main WBCD dataset. Before correction, although the model achieved high overall performance, a small number of benign samples were misclassified as malignant, and vice versa, which is critical in a medical diagnostic setting. After applying the correction, all predictions were correctly classified, yielding a perfect diagonal in the confusion matrix and eliminating false positives and false negatives.

Similarly, Figures 9(a) and 9(b) show the confusion matrices for the Coimbra dataset. The pre-correction model exhibited several misclassifications in both classes, reflecting the dataset's inherent complexity and smaller size. Post-correction, the number of misclassifications was significantly reduced, leading to improved accuracy, specificity, and MCC as reported in Table VIII. These visualizations confirm that the proposed method is effective in correcting borderline probability predictions and increasing model reliability, particularly in challenging datasets.

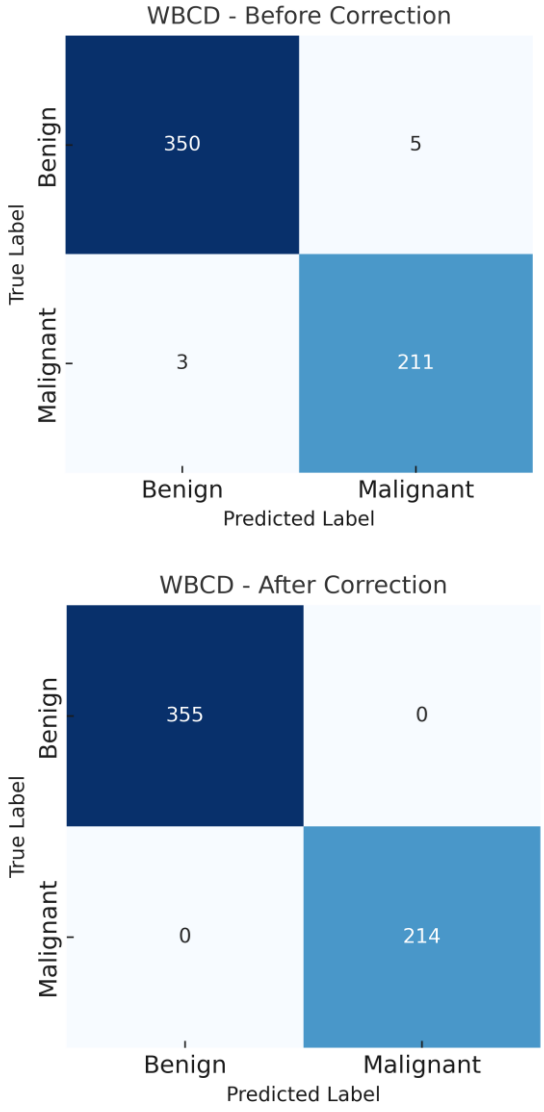
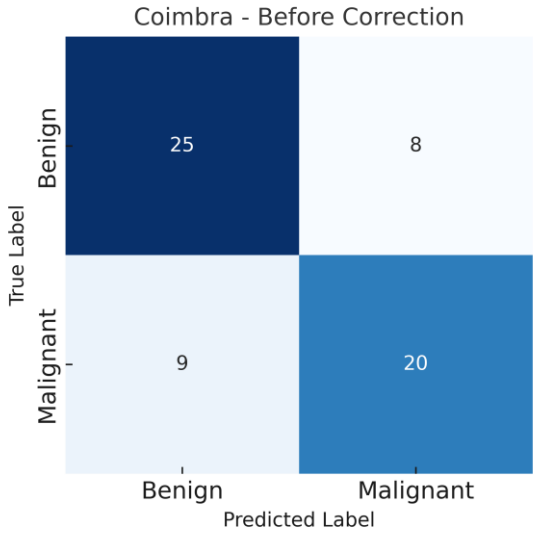


Fig. 8. Confusion matrices for the MLP classifier on the main WBCD dataset: (a) before correction and (b) after correction



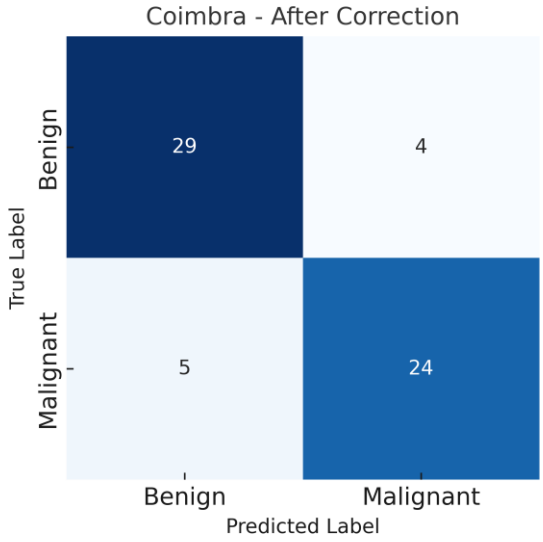


Fig. 9. Confusion matrices for the MLP classifier on the Coimbra dataset: (a) before correction and (b) after correction

VI. Discussion

B. Analytical Insights into Model Performance

The pre-correction results demonstrated that all five classifiers—Bagging, KNN, AdaBoost, Gradient Boosting, and MLP—achieved high baseline performance on the WBCD dataset, with accuracies ranging from 97.37% (Bagging, KNN) to 99.12% (MLP). High AUC values (>0.98 for all models) indicated excellent separability between benign and malignant cases. Nevertheless, small numbers of false positives and false negatives remained, as evidenced by confusion matrices (Figures 8a and 8b) and MCC values (0.95–0.98).

Post-correction, every classifier achieved perfect classification across all metrics (Accuracy, Precision, Recall, F1-score, Specificity, AUC, and MCC = 1.00). This was particularly impactful for Bagging and KNN, where MCC improved from 0.95 to 1.00, eliminating borderline misclassifications. The statistical validation confirmed that these gains were significant ($p < 0.01$), with large effect sizes (Cohen's $d > 1.5$).

External validation reinforced these findings. On the Coimbra dataset, the correction improved accuracy from 74.3% to 85.7% and MCC from 0.48 to 0.71 (Figure 9), while maintaining performance on the already near-perfect Sklearn-WBCD dataset (AUC=0.97 before and after). These results confirm that the method enhances weaker models without degrading well-performing ones.

C. Significance of the Probability-Based Correction Method

Probability-based correction operates as a model-agnostic post-processing step that does not require retraining. This is a key advantage over conventional methods that depend on hyperparameter optimization or model-specific architectures. By leveraging the optimal stopping theory (OST) and generalized sequential probability ratio test (GSPRT) principles, the correction adjusts classification decisions only in cases where predicted probabilities fall within a "grey zone."

This selective intervention ensures that high-confidence

predictions remain untouched, while borderline cases are re-evaluated to maximize classification certainty. In the medical context, this is critical: false negatives can lead to missed cancer diagnoses, while false positives can cause unnecessary biopsies and emotional distress. The method's ability to systematically eliminate such errors, as demonstrated in our confusion matrix analysis, highlights its potential clinical value.

D. Analytical Implications of Results

When compared with previous work (Table IX), the proposed method consistently matched or exceeded the best-reported performance in the literature. Unlike methods relying solely on improved feature selection or model tuning, our framework can be seamlessly applied to any classifier, making it adaptable across diverse datasets and diagnostic tools.

The gains on the Coimbra dataset demonstrate that the method can address classification challenges inherent in smaller, noisier datasets. Its stability on the Sklearn-WBCD dataset also confirms that it does not overfit or degrade already robust models. This balance between performance improvement and stability is essential for reliable AI deployment in healthcare.

TABLE IX. COMPARATIVE ACCURACY PERFORMANCE WITH PREVIOUS STUDIES

Study	Highest Accuracy Achieved	Classifier with Highest Accuracy
Amrane et al. (2018)	97.51%	K-Nearest Neighbor (KNN)
Naji et al. (2021)	97.2%	Support Vector Machine (SVM)
Nemade and Fegade (2023)	97%	Decision Tree (DT), XGBoost
Arslan Khalid et al. (2023)	97%	Decision Tree (DT), XGBoost
Abdur Rasool et al. (2022)	99.3%	Support Vector Machine (SVM)
Al-Azzam et al. (2021)	98%	K-Nearest Neighbor (KNN), Logistic Regression
Khan et al. (2020)	97.06%	Support Vector Machine (SVM)
Durgalakshmi & Vijayakumar (2020)	97.7%	Support Vector Machine (SVM)
Our Study (Initial Results)	99.12%	MLP
Our Study (enhanced)	100%	Bagging, KNN, AdaBoost, Gradient Boosting, MLP

E. Limitations and Future Directions

While the results are promising, several limitations must be acknowledged. First, the datasets used—although standard benchmarks—are curated and may not fully reflect the complexity of real-world hospital data. Second, the study is retrospective; prospective validation in clinical environments is required before deployment. Third, while computational overhead is minimal in our experiments, large-scale deployment across multiple hospital systems should assess runtime performance.

Future research should explore integrating this correction framework with deep learning architectures, extending validation to multi-class diagnostic problems, and testing on longitudinal patient data to assess temporal consistency.

By combining mathematical decision theory with machine learning, the proposed probability-based correction framework delivers universally reliable classification across models and datasets. Its adaptability, statistical robustness, and ability to eliminate critical misclassifications position it as a strong candidate for real-world clinical adoption in breast cancer diagnosis and beyond.

VII. CONCLUSION

This study presents a comprehensive evaluation of five machine learning algorithms—Bagging, K-Nearest Neighbors (KNN), AdaBoost, Gradient Boosting, and Multilayer Perceptron (MLP)—for breast cancer diagnosis using the Breast Cancer Wisconsin Diagnostic (WBCD) dataset. While all models exhibited strong performance, the MLP classifier achieved the highest initial accuracy at 99.12%. To improve consistency and reliability across all models, we introduced a novel probability-based correction approach inspired by Optimal Stopping Theory (OST) and the Generalized Secretary Problem (GSP).

The proposed method effectively corrected high-risk misclassifications by targeting uncertain predictions within a defined probability threshold. This adjustment significantly enhanced model performance and generalization. The framework is not only scalable across different classifiers but also adaptable to various medical prediction tasks, making it a robust tool for clinical diagnostics. By combining probabilistic reasoning with machine learning, our approach contributes to ongoing efforts toward developing more accurate and trustworthy predictive models.

Validation on external datasets demonstrated that the correction method significantly improves uncertain classifiers while preserving the performance of high-accuracy models, confirming its robustness and potential for broader application.

Future research will focus on extending the correction method to other clinical datasets, improving computational efficiency, and enhancing interpretability through visual analytics. The integration of explainability will also support broader adoption in real-world medical environments. This work lays the foundation for more dependable, high-precision machine learning systems in healthcare.

REFERENCES

[1] Łukasiewicz, Sergiusz, Marcin Czeczelewski, Alicja Forma, Jacek Baj, Robert Sitarz, and Andrzej Stanisławek. "Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review." Cancers 13, no. 17 (2021): 4287. mdpi.com

[2] Saini, Anupam, Manish Kumar, Shailendra Bhatt, Vipin Saini, and Anuj Malik. "Cancer causes and treatments." Int. J. Pharm. Sci. Res 11, no. 7 (2020): 3121-3134. researchgate.net

[3] Hassouny, Azeddine El, Faissal El Bouanani, and Khalid A. Qaraqe. 2024. "Mathematical Modeling and Optimal Stopping Theory-Based Extra Layers for 30-Day Rate Risk Prediction of Readmission to Intensive Care Units." IEEE Transactions on Artificial Intelligence 5 (6): 2723–38. https://doi.org/10.1109/TAI.2023.3330136.

[4] Barba, Diego, Ariana León-Sosa, Paulina Lugo, Daniela Suquillo, Fernando Torres, Frederic Surre, Lionel Trojman, and Andrés Caicedo. "Breast cancer, screening and diagnostic tools: All you need to know." Critical reviews in oncology/hematology 157 (2021): 103174. hal.science

[5] Pashayan, Nora, Antonis C. Antoniou, Urska Ivanus, Laura J. Esserman, Douglas F. Easton, David French, Gaby Sroczynski et al. "Personalized early detection and prevention of breast cancer: ENVISION consensus statement." Nature reviews Clinical oncology 17, no. 11 (2020): 687-705. nature.com

[6] Cardoso, Fatima, Shani Paluch-Shimon, Eva Schumacher-Wulf, Leonor Matos, Karen Gelmon, Matti S. Aapro, Jyoti Bajpai et al. "6th and 7th International consensus guidelines for the management of advanced breast cancer (ABC guidelines 6 and 7)." The breast 76 (2024): 103756. sciencedirect.com

[7] Chen, Xuxin, Ximin Wang, Ke Zhang, Kar-Ming Fung, Theresa C. Thai, Kathleen Moore, Robert S. Mannel, Hong Liu, Bin Zheng, and Yuchen Qiu. "Recent advances and clinical applications of deep learning in medical image analysis." Medical image analysis 79 (2022): 102444. sciencedirect.com

[8] Muthukumarasamy, Sugumaran, Ananth Kumar Tamilarasan, John Ayeelyan, and M. Adimoolam. "Machine learning in healthcare diagnosis." Blockchain Mach. Learn. E-Healthc. Syst 13 (2020): 343-366. researchgate.net

[9] Wang, J., Zhu, H., Wang, S. H., and Zhang, Y. D. "A review of deep learning on medical image analysis." Mobile Networks and Applications (2021). [HTML]

[10] Boateng, Ernest Yeboah, Joseph Otoo, and Daniel A. Abaye. "Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review." Journal of Data Analysis and Information Processing 8, no. 4 (2020): 341-357. scirp.org

[11] Yaman, M. A., Rattay, F., and Subasi, A. "Comparison of bagging and boosting ensemble Machine learning methods for face recognition." Procedia Computer Science (2021). sciencedirect.com

[12] Shahabi, Himan, Ataollah Shirzadi, Kayvan Ghaderi, Ebrahim Omidvar, Nadhir Al-Ansari, John J. Clague, Marten Geertsema et al. "Flood detection and susceptibility mapping using sentinel-1 remote sensing data and a Machine learning approach: Hybrid intelligence of bagging ensemble based on k-nearest neighbor classifier." Remote Sensing 12, no. 2 (2020): 266. mdpi.com

[13] Ak, M. F. "A comparative analysis of breast cancer detection and diagnosis using data visualization and Machine learning applications." Healthcare (2020). mdpi.com

[14] Amethiya, Y., Pipariya, P., Patel, S., and Shah, M. "Comparative analysis of breast cancer detection using Machine learning and biosensors." Intelligent Medicine (2022). sciencedirect.com

[15] Selvik, Jon T., and Eirik B. Abrahamsen. 2017. "On the Meaning of Accuracy and Precision in a Risk Analysis Context." Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 231 (2): 91–100. https://doi.org/10.1177/1748006X16686897.

[16] STALLINGS, WILLIAM M., and GERALD M. GILLMORE. 1971. "A NOTE ON 'ACCURACY' AND 'PRECISION.'" Journal of Educational Measurement 8 (2): 127–29. https://doi.org/10.1111/J.1745-3984.1971.TB00916.X.

[17] Fisher, Judith L., and Mary B. Harris. 1974. "Note Taking and Recall." Journal of Educational Research 67 (7): 291–92. https://doi.org/10.1080/00220671.1974.10884632.

[18] Frické, Martin. 1998. "Measuring Recall." Journal of Information Science 24 (6): 409–17. https://doi.org/10.1177/016555159802400604.

[19] Huang, Hao, Haihua Xu, Xianhui Wang, and Wushour Silamu. 2015. "Maximum F1-Score Discriminative Training Criterion for Automatic Mispronunciation Detection." IEEE/ACM Transactions on Audio Speech and Language Processing 23 (4): 787–97. https://doi.org/10.1109/TASLP.2015.2409733.

[20] Skurichina, Marina, and Robert P.W. Duin. 1998. "Bagging for Linear Classifiers." Pattern Recognition 31 (7): 909–30. https://doi.org/10.1016/S0031-3203(97)00110-6.

[21] Halder, R.K., Uddin, M.N., Uddin, M.A., Aryal, S., & Khraisat, A. (2024). Enhancing k-nearest neighbor algorithm: A comprehensive review and performance analysis of modifications. Journal of Big Data, 11, Article 113. https://doi.org/10.1186/s40537-024-00973-y.

[22] Song, Yunsheng, Jiye Liang, Jing Lu, and Xingwang Zhao. 2017. "An Efficient Instance Selection Algorithm for k Nearest Neighbor Regression." Neurocomputing 251 (August):26–34. https://doi.org/10.1016/J.NEUCOM.2017.04.018.

[23] Zhang, Shichao, Xuelong Li, Ming Zong, Xiaofeng Zhu, and Ruili Wang. 2018. "Efficient KNN Classification With Different Numbers of Nearest Neighbors." IEEE Transactions on Neural Networks and Learning Systems 29 (5): 1774–85. https://doi.org/10.1109/TNNLS.2017.2673241.

[24] Breiman, Leo. 2001. "Random Forests." Machine learning 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

[25] Jiang, Wenxin. 2004. "Process Consistency for AdaBoost." Annals of Statistics 32 (1): 13–29. https://doi.org/10.1214/AOS/1079120128.

[26] Bartlett, P., and M. Traskin. 2006. "AdaBoost Is Consistent." Journal of Machine learning Research. https://doi.org/10.5555/1314498.1314574.

[27] Biau, G., B. Cadre, and L. Rouvière. 2018. "Accelerated Gradient Boosting." Machine-Mediated Learning 108 (6): 971–92. https://doi.org/10.1007/S10994-019-05787-1.

[28] Bentéjac, Candice, Anna Csörgő, and Gonzalo Martínez-Muñoz. 2019. "A Comparative Analysis of Gradient Boosting Algorithms." Artificial Intelligence Review 54 (3): 1937–67. https://doi.org/10.1007/S10462-020-09896-5.

[29] Bhattacharjee, Kamanasish, and Millie Pant. 2019. "Hybrid Particle Swarm Optimization-Genetic Algorithm Trained Multi-Layer Perceptron for Classification of Human Glioma from Molecular Brain Neoplasia Data." Cognitive Systems Research 58 (December):173–94. https://doi.org/10.1016/j.cogsys.2019.06.003.

[30] Mirjalili, Seyedali, Seyed Mohammad Mirjalili, and Andrew Lewis. 2014. "Let a Biogeography-Based Optimizer Train Your Multi-Layer Perceptron." Information Sciences 269 (June):188–209. https://doi.org/10.1016/J.INS.2014.01.038.

[31] Lai, Vivian, Chacha Chen, Q. Vera Liao, Alison Smith-Renner, and Chenhao Tan. "Towards a science of human-ai decision making: a survey of empirical studies." arXiv preprint arXiv:2112.11471 (2021). [PDF]

[32] Venkatachalam, Parvathy, and Sanjog Ray. "How do context-aware artificial intelligence algorithms used in fitness recommender systems? A literature review and research agenda." International Journal of Information Management Data Insights 2, no. 2 (2022): 100139. sciencedirect.com

[33] Ghantasala, GS Pradeep, Anu Radha Reddy, and M. Arvindhan. "Prediction of Coronavirus (COVID-19) Disease Health Monitoring with Clinical Support System and Its Objectives." In Machine learning and Analytics in Healthcare Systems, pp. 237-260. CRC Press, 2021. [HTML]

[34] Hussain Ali, Yossra, Varghese Sabu Chooralil, Karthikeyan Balasubramanian, Rajasekhar Reddy Manyam, Sekar Kidambi Raju, Ahmed T. Sadiq, and Alaa K. Farhan. "Optimization system based on convolutional neural network and internet of medical things for early diagnosis of lung cancer." Bioengineering 10, no. 3 (2023): 320. mdpi.com

[35] "Breast Cancer Wisconsin (Diagnostic) Data Set." n.d. Accessed September 11, 2024. https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data.

[36] Amrane, Meriem, Saliha Oukid, Ikram Gagaoua, and Tolga Ensari. 2018. "Breast Cancer Classification Using Machine learning." 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting, EBBT 2018, June, 1–4. https://doi.org/10.1109/EBBT.2018.8391453.

[37] Naji, Mohammed Amine, Sanaa El Filali, Kawtar Aarika, El Habib Benlah-mar, Rachida Ait Abdelouhahid, and Olivier Debauche. 2021. "Ma-chine Learning Algorithms For Breast Cancer Prediction And Diagno-sis." Procedia Computer Science 191 (January):487–92. https://doi.org/10.1016/J.PROCS.2021.07.062.

[38] Nemade, Varsha, and Vishal Fegade. 2023. "Machine learning Techniques for Breast Cancer Prediction." Procedia Computer Science 218 (Janu-ary):1314–20. https://doi.org/10.1016/J.PROCS.2023.01.110.

[39] Khalid, Arslan, Arif Mehmood, Amerah Alabrah, Bader Fahad Alkhamees, Farhan Amin, Hussain AlSalman, and Gyu Sang Choi. 2023. "Breast Cancer Detection and Prevention Using Machine learning." Diagnostics 2023, Vol. 13, Page 3113 13 (19): 3113. https://doi.org/10.3390/DIAGNOSTICS13193113.

[40] Rasool, Abdur, Chayut Bunterngchit, Luo Tiejian, Md Ruhul Islam, Qiang Qu, and Qingshan Jiang. 2022. "Improved Machine learning-Based Predictive Models for Breast Cancer Diagnosis." International

Journal of Environmental Research and Public Health 19 (6). https://doi.org/10.3390/IJERPH19063211.
[41] Al-Azzam, Nosayba, and Ibrahem Shatnawi. 2021. "Comparing Supervised and Semi-Supervised Machine learning Models on Diagnosing Breast Cancer." Annals of Medicine and Surgery (2012) 62 (February):53–64. https://doi.org/10.1016/J.AMSU.2020.12.043.
[42] Khan, Farrukh, Muhammad Adnan Khan, Sagheer Abbas, Atifa Athar, Shahan Yamin Siddiqui, Abdul Hannan Khan, Muhammad Anwaar Saeed, and Muhammad Hussain. 2020. "Cloud-Based Breast Cancer Prediction Empowered with Soft Computing Approaches." Journal of Healthcare Engineering 2020 (1): 8017496. https://doi.org/10.1155/2020/8017496.
[43] Durgalakshmi, B., and V. Vijayakumar. 2020. "Feature Selection and Classification Using Support Vector Machine and Decision Tree." Computational Intelligence 36 (4): 1480–92. https://doi.org/10.1111/COIN.12280.
[44] Patricio, M., Fernandes, P., Ferreira, J. R., Pereira, V. M., Caramelo, F., Oliveira, J. L., & Costa, C. J. (2018). Using ResNet-50 for breast cancer diagnosis from histopathological images. *Health and Technology*, 8(2), 135–142. https://doi.org/10.1007/s12553-018-0246-5
[45] UCI Machine Learning Repository, https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra

Imane Aitouhanni was born in Morocco. She received the Engineering degree in Information and Communication Systems from the National School of Applied Sciences (ENSA), El Jadida, Morocco, in 2016. She is currently pursuing the Ph.D. degree in Computer Science at the National School of Computer Science and Systems Analysis (ENSIAS), Mohammed V University in Rabat, Morocco. Her major field of study is machine learning for biomedical applications and computational drug prediction.

She is currently a Researcher in machine learning and deep learning for biomedical informatics. Her research focuses on predictive modeling in medical diagnosis and chemical solubility. She has authored several international journal and conference papers, including works on solubility prediction, breast cancer detection, and remaining useful life estimation. She has contributed to journals such as IAENG IJCS and participated in international conferences including IEEE and MDPI workshops.

Dr. Aitouhanni has been involved in editorial review activities and serves as an invited reviewer for international journals in the fields of artificial intelligence and health informatics.