

Research Article

A Unifying Paradigm in Predictive Oncology Through Hybrid Voting Ensembles for Optimising Breast Cancer Detection

Raz M. Hama Salih^{1,*} , Saman Hussein Mahmood¹ 

¹ Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil, Erbil, 44001, Iraq

*Corresponding Author: Raz M. Hama Salih, E-mail: raz.hamasalih@su.edu.krd

Article Info	Abstract
Article History	Machine learning (ML) offers substantial promise for improving cancer diagnosis, yet many existing approaches rely on single-classifier models. This study proposes and evaluates a hybrid weighted voting ensemble to optimise predictive accuracy for breast cancer classification by combining a diverse set of ML classifiers. Using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, five base learners were developed: a multi-layer perceptron (MLP), a Classification and Regression Tree (CART), and three support vector machine (SVM) models with linear, radial basis function (RBF), and polynomial kernels. A comparative analysis of individual model performance was conducted against the proposed ensemble. The weighted voting model achieved the best overall performance, with an accuracy of 0.9737, F1-score of 0.9641, and AUC of 0.9952, surpassing the strongest individual classifiers (MLP and SVM-RBF, each with an accuracy of 0.9591). Permutation feature importance highlighted "worst" measurements, such as concave points_worst and radius_worst, as the most informative predictors, consistent with established malignancy indicators. The proposed hybrid voting ensemble provides a robust and reliable framework for breast cancer prediction, demonstrating clear advantages over single-classifier models and supporting more precise decision-making in oncology.
Received Oct 20, 2025	
Revised Dec 07, 2025	
Accepted Dec 12, 2025	
Keywords	
Diagnostic accuracy	
Machine-learning	
Ensemble-learning	
Breast cancer	
Predictive modeling	
 Copyright: © 2026 Raz M. Hama Salih, and Saman Hussein Mahmood. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license.	

1. Introduction

Breast cancer is a major global health challenge, being one of the most commonly diagnosed cancers and a leading cause of cancer-related deaths among women [1, 2]. According to the Global Cancer Observatory 2022 estimates, female breast cancer accounted for 11.6% of new cancer cases, with nearly 2.3 million diagnoses, and 6.9% of all cancer deaths [1, 3, 4]. The impact of this malignancy is expected to increase, and it has been forecasted to reach as high as 35 million new cases per year by 2050, placing a significant strain on global health systems [3, 4]. Precision in diagnosis is vital for optimising patient outcomes and survival rates, and thus is an axis of clinical practice and study [5].

To address healthcare challenges, machine learning (ML) is a revolutionary approach in the global healthcare sector [6]. Incorporating ML into clinical processes enhances diagnostic precision, streamlines processes, and supports decision-making by identifying intricate patterns in medical data [7, 8]. It is increasingly used for disease forecasting, classification, and treatment planning, thereby enabling more personalised patient care [9, 10].

For breast malignancy, ML techniques have emerged as highly promising tools to support clinicians in making earlier and more accurate predictions [5, 11]. By characterising cytological attributes, these models can distinguish between benign and malignant tumours with high accuracy [8, 9]. In doing so, they aim to enhance diagnostic accuracy, reduce unnecessary invasive tests, and improve patient outcomes while optimising healthcare resources [8].

This study aims to develop and evaluate a hybrid voting ensemble algorithm specifically designed to optimise the collective predictive performance of MLP, multi-kernel SVM, and CART models for breast cancer disease prediction. By integrating these diverse classifiers, this research aims to develop a more accurate, robust, and reliable diagnostic tool than can be achieved with individual models or conventional ensemble approaches. This paper makes three main contributions:

- 1) To develop and implement a new hybrid voting ensemble algorithm for breast cancer classification.
- 2) To conduct a comprehensive performance analysis of five distinct base learners: MLP, SVM-Linear, SVM-Polynomial, SVM-RBF, and CART, establishing their individual predictive capabilities on a benchmark dataset.
- 3) To provide a comparative evaluation demonstrating the superiority of the proposed optimised ensemble model against its individual components, highlighting the synergistic benefits of the hybrid approach.

2. Literature Review

Applications of machine learning to breast cancer prediction have been extensively studied, with a focus on developing the best-performing classification models. Neural Networks, particularly Multi-Layer Perceptron (MLPs), demonstrate the capacity to model complex, non-linear relationships and achieve high classification accuracy [12-14]. Support Vector Machines (SVMs) are also widely used with kernel functions (linear, polynomial, and radial basis), which are crucial for achieving optimal performance [15-17]. Another widely used approach is the decision tree, particularly the Classification and Regression Tree

(CART) algorithm, which provides interpretable and concise classification rules, even when slightly lower predictive accuracy is tolerated [18-20].

To overcome the limitations inherent in single-classifier systems, ensemble learning has gained significant traction in the field. The fundamental principle of ensemble methods is to combine the predictions of multiple base learners to produce a single, more robust outcome [21, 22]. Techniques such as bagging, boosting, and stacking have been successfully applied to breast cancer datasets, often yielding superior performance [23, 24]. More specifically, voting classifiers, which aggregate the predictions of multiple models, have proven particularly effective and straightforward ensemble strategies for enhancing diagnostic precision in breast cancer [2, 25, 26].

Despite these advancements, an important research gap persists. While individual models have shown promise and standard ensembles provide clear benefits, there is a lack of research on hybrid voting strategies that systematically optimise a diverse set of powerful classifiers (MLP, multiple SVM kernels, and CART) specifically for breast cancer prediction. Many studies focus on comparing individual models or implementing homogeneous or standard heterogeneous ensembles. The potential to create a synergistic, optimised hybrid model that leverages the unique strengths of MLPs' nonlinearity, the varied decision boundaries of multi-kernel SVMs, and the transparent logic of CART remains underexplored.

This study manages the research gap by proposing a hybrid weighted voting ensemble that combines predictions from a Multi-Layer Perceptron (MLP), a multi-kernel Support Vector Machine (SVM), and Classification and Regression Trees (CART), with weights specified by their AUC (Area Under the Curve) scores. This approach will improve diagnostic accuracy, F1-score, and AUC compared to the individual classifiers.

3. Materials and Methods

This section outlines the data source, data preprocessing, machine learning models, the proposed voting ensemble algorithm, and the evaluation metrics used to assess model performance. The methodology ensures replicability and provides a solid foundation for experimental analysis.

3.1. Data Source and Description

The study utilises the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, available in the UCI Machine Learning Repository. This dataset is a widely used benchmark for research on breast cancer

diagnosis. It comprises 569 patient records, each with 30 numeric attributes derived from digitised images of fine-needle aspirates (FNAs) of breast tumours. Each record includes measurements for radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension, with three recorded values: mean, standard error, and worst (largest) measurement. The target variable comprises 212 malignant cases (37.3%) and 357 benign cases (62.7%), yielding a slightly imbalanced but realistic medical classification problem. Furthermore, this dataset was selected as a public example from the UCI Machine Learning Repository to enable reproducible performance comparisons and to validate classification algorithms, as its numerical features are well-suited for this task.

3.2. Data Preprocessing

Before model training, the following preprocessing steps were performed:

- 1) Missing Values: The dataset contains no missing entries; thus, no imputation was necessary.
- 2) To ensure that features with larger numeric ranges do not disproportionately affect model training, z-score standardisation was applied to normalise the dataset, setting features' means to 0 and standard deviations to 1, which is crucial for algorithms like Support Vector Machines (SVMs) and Neural Networks. While z-score standardisation can be sensitive to outliers, the high-quality WDBC dataset, which consists of clinical measurements, reduces the likelihood of significant errors. Additionally, the ensemble model used in this study helps mitigate the impact of small data variations, making z-score standardisation a suitable choice.
- 3) Data Splitting: The dataset was divided into two subsets, 80% for training and 20% for testing.

3.3. Base Machine Learning Models

Five machine-learning classifiers were employed as base learners: Decision Tree (CART), Multi-Layer Perceptron (MLP), and Support Vector Machine (with linear, polynomial, and radial basis function) kernels.

3.3.1. MLP

A Multi-layer Perceptron (MLP) is a feed-forward neural network with an input layer of 30 neurons, a single layer or multiple hidden layers using activation functions, and an output layer employing a sigmoid function to handle tasks of binary classification. It is a design that is efficient at learning non-linear decision areas, thus suitable for complex medical datasets [13].

$$y_j^{(l)} = f\left(\sum_{i=1}^n w_{ij}^{(l)} x_i^{(l-1)} + b_j^{(l)}\right)$$

where w_{ij} is the weight connecting neurons i and j , and b_j is the bias term

3.3.2. SVM

Support Vector Machines (SVMs) find the best hyperplane for optimal class separation and reduce classification error. They are typically utilised in healthcare applications for their reliability in high-dimensional spaces [16]. There are several kernel functions of SVM:

1) Linear kernel Function

The linear kernel function is efficient for linearly separable data. It detects a hyperplane that is maximally separable and minimally misclassifying, and the allowance for misclassification determines the hyperplane's accuracy [16].

$$K(x, y) = x \cdot z + C \quad (1)$$

where C is a parameter that balances the margin of the decision boundary and the training data accuracy.

2) Polynomial kernel

The Polynomial kernel transforms input data into a higher-dimensional space using a polynomial function, making it valid for data that cannot be separated by a straight line [19].

$$K(x, y) = (\gamma x \cdot z + C)^d \quad (2)$$

where γ : scaling parameter that involves the inner product $x \cdot z$, and d is polynomial's degree.

3) Radial Basis Function

It is one of the most popular kernels. RBF kernel measures similarity based on the radial distance between data points, using a gamma parameter that controls the influence of each data point [20, 21].

$$K(x, y) = \text{EXP}(-\gamma \|x - z\|^2) \quad (3)$$

where γ is a positive gamma parameter that determines the effect of each data point.

3.3.3. CART

The Classification and Regression Tree (CART) algorithm partitions data into subsets using a tree-like structure.

- Splitting Criterion: The Gini index is used to evaluate split quality.
- Structure: Recursive partitioning continues until terminal nodes (leaves) represent class labels.

- Strengths: Offers interpretability and transparency, providing decision rules that are easy for medical practitioners to follow.
- For CART using the Gini index as a splitting criterion:

$$Gini(S) = 1 - \sum_{i=1}^c (p_i)^2 \quad (4)$$

To create a consistent and reproducible baseline for evaluating the proposed ensemble method, we used the default hyperparameters from scikit-learn (version 1.7.2) for each base learner. This study primarily focused on showing the performance improvement from our creative weighted voting strategy rather than on optimising individual models. While parameter tuning could improve each base learner's performance, using the default settings ensures that the ensemble's superior results are attributable to the ensemble technique itself. This approach establishes a clear baseline for comparison and highlights the hybrid model's strengths.

3.4. Proposed Hybrid Voting Ensemble Algorithm

Ensemble methods combine multiple classifiers to enhance classification accuracy and robustness by reducing variance and bias [27]. In this research, we employ the Voting method, a straightforward yet powerful ensemble approach. Voting combines predictions from several classifiers using majority or weighted methods. The final classification decision is the one with the most votes. Weighted voting can be used to enhance ensemble precision and assign greater influence to classifiers based on their performance [28].

$$y_{\text{ensemble}} = \underset{j}{\operatorname{argmax}} \sum_{m=1}^M \delta(y_m = j) \quad (5)$$

- $\delta(\cdot)$: indicator function
- y_m : classifier's prediction
- M : total number of classifiers.

In the proposed Weighted Voting ensemble, weights for each base classifier were measured by the area Under the Curve (AUC) in such a way that the classifiers with higher AUC scores, such as MLP and SVM-RBF, had a greater influence on the final prediction. In contrast, those with lower scores, like SVM-Poly, contributed less. This technique influenced the strengths of effective models while benefiting from the diversity of weaker classifiers, and the weight selection process was transparent and based on experimental results [29].

Algorithm 1. Pseudocode of the Proposed Weighted Voting Ensemble for Breast Cancer Classification

```

Start
For seed = 0 to N do
    Set random seed
    Split the dataset: training (80%) and testing (20%)
    Standardise features (using z-score)
    Classifiers methods:
        C1: SVM (Linear)
        C2: SVM (RBF)
        C3: SVM (Polynomial)
        C4: Decision Tree (CART)
        C5: Multi-Layer Perceptron (MLP)
    For each classifier Ci do
        Train Ci on the training set
        Predict probabilities on the test set
        Calculate performance metrics: AUC, Accuracy, and F1-score.
        Store AUCi
    End For
    Assign weights:
        wi = AUCi for each classifier Ci
    Construct Weighted Voting Ensemble:
        Combine {C1...C5} using soft voting with weights {w1...w5}
    Evaluate ensemble:
        Compute Accuracy, F1-score, AUC
        Generate a confusion matrix and an ROC under the curve
    Record results for this seed
End For
Apply 10-fold cross-validation for robustness
Report mean ± standard deviation of all metrics
Stop

```

In the hybrid ensemble, the weights for every base model were selected based on AUC (Area Under the Curve) instead of equal or accuracy-based techniques because the equal weighting strategy can ignore differences in performance of the robust classifiers, for example, MLP and SVM-RBF, and weaker ones, for example, SVM (Poly). Moreover, accuracy-based methods are misleading in slightly imbalanced data. Conversely, the AUC is an accurate measure of a model's capacity to discriminate between benign and

malignant cases for all thresholds, which is appropriate for medical diagnosis applications where minimising false negatives is crucial. A model with a high AUC achieves a high True Positive Rate while maintaining a low False Positive Rate, allowing us to prioritise models that effectively differentiate between tumour types [30].

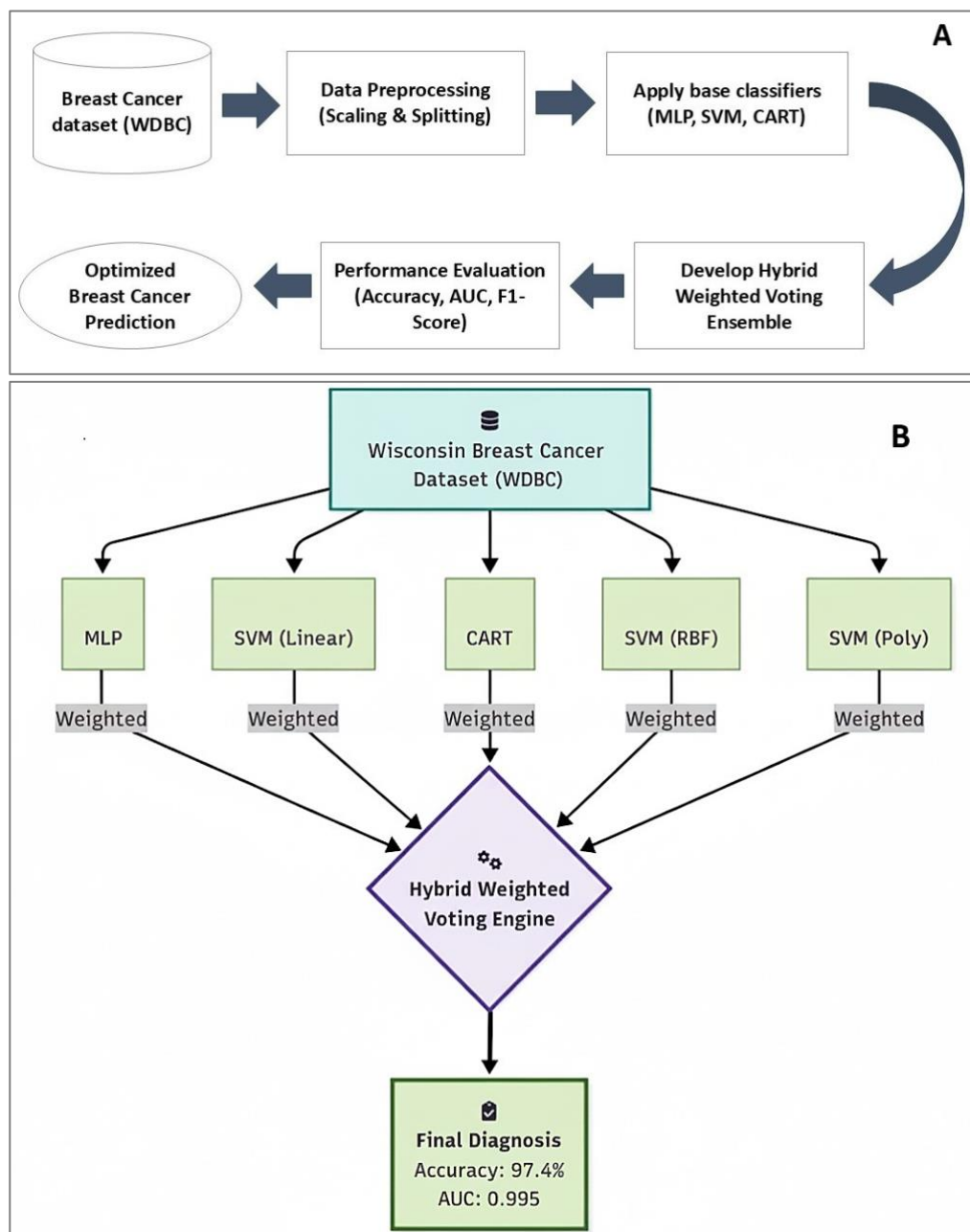


Figure 1. A. Research workflow diagram for optimising breast cancer prediction. B. Architecture of the Hybrid Weighted Voting Ensemble Model

3.5. Experimental Setup and Performance Evaluation Metrics

Experiments were conducted using Python 3.10, scikit-learn, pandas, and NumPy. To ensure robustness, a cross-validation (10-fold) was used on the training set, and the reported metrics were averaged

across folds. Furthermore, Area Under the ROC Curve (AUC), Accuracy, and F1-Score are used as evaluation metrics.

4. Results and Disucssions

Figure 2 presents a heatmap of the Pearson correlation coefficients for the 30 features in the WDBC dataset. Colour intensity reflects the strength of correlation, with deep red indicating a strong positive correlation (coefficient near +1.0) and blue indicating a negative correlation. As observed, many features are highly correlated. For instance, the means of radius, perimeter, and area are strongly correlated, as they are geometrically related to tumour measurements. This high degree of multicollinearity highlights the importance of using robust models, such as SVMs and ensemble methods, that can effectively handle such relationships.

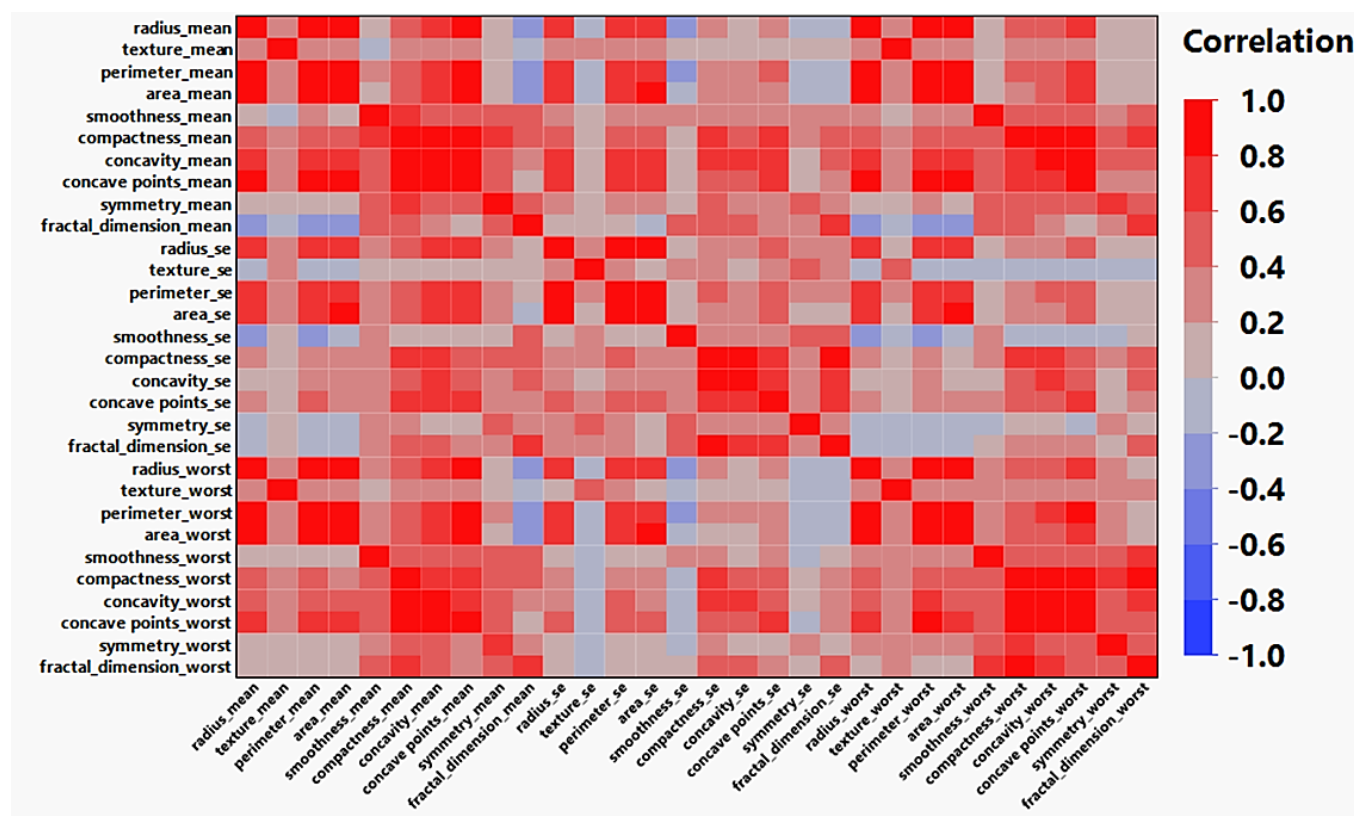


Figure 2. Feature Correlation Matrix to show the relationship between each pair of features

Table 1 compares the five base classifiers and the proposed Weighted Voting ensemble model across five key metrics, with one standard deviation included for each metric: Accuracy, F1-score, and AUC. The results show that the Neural Network (MLP) and SVM (RBF) models both achieved an accuracy of 0.9591 (standard deviation 0.0083), indicating high and comparable performance to the SVM (Linear) model. In contrast, the Decision Tree (CART) model shows moderate performance with an accuracy of 0.9152

(0.0207), while the SVM (Poly) performs the poorest with an accuracy of 0.8801 (0.0124). Notably, the SVM-Poly model achieves very high precision (0.9780 ± 0.0010) but the lowest recall (0.6953 ± 0.0331), suggesting it is overly conservative in identifying malignant cases and resulting in a high number of false negatives.

The standout performer is the Weighted Voting ensemble model, which surpasses all individual models across nearly every metric. It achieves the highest accuracy of 0.9737 (0.0041), F1-score of 0.9641 (0.0058), and AUC of 0.9952 (0.0004). This superior performance highlights the synergistic effect of combining diverse classifiers, in which the strengths of individual models offset their weaknesses, yielding a more robust and accurate final prediction.

Table 1. Comparison of the performance of the individual and ensemble machine learning models

Model	Accuracy $\pm$ SD	F1-score $\pm$ SD	Precision $\pm$ SD	Recall $\pm$ SD	AUC $\pm$ SD
Decision Tree (CART)	0.9152 ± 0.0207	0.8902 ± 0.0228	0.8686 ± 0.0532	0.9141 ± 0.0110	0.9150 ± 0.0143
Neural Network (MLP)	0.9591 ± 0.0083	0.9446 ± 0.0100	0.9604 ± 0.0324	0.9297 ± 0.0110	0.9945 ± 0.0001
SVM Linear	0.9561 ± 0.0041	0.9403 ± 0.0053	0.9594 ± 0.0110	0.9219 ± 0.0000	0.9940 ± 0.0014
SVM Poly	0.8801 ± 0.0124	0.8126 ± 0.0230	0.9780 ± 0.0010	0.6953 ± 0.0331	0.9850 ± 0.0050
SVM RBF	0.9591 ± 0.0000	0.9444 ± 0.0006	0.9598 ± 0.0105	0.9297 ± 0.0110	0.9930 ± 0.0006
Weighted Voting	0.9737 ± 0.0041	0.9641 ± 0.0058	0.9837 ± 0.0002	0.9453 ± 0.0110	0.9952 ± 0.0004

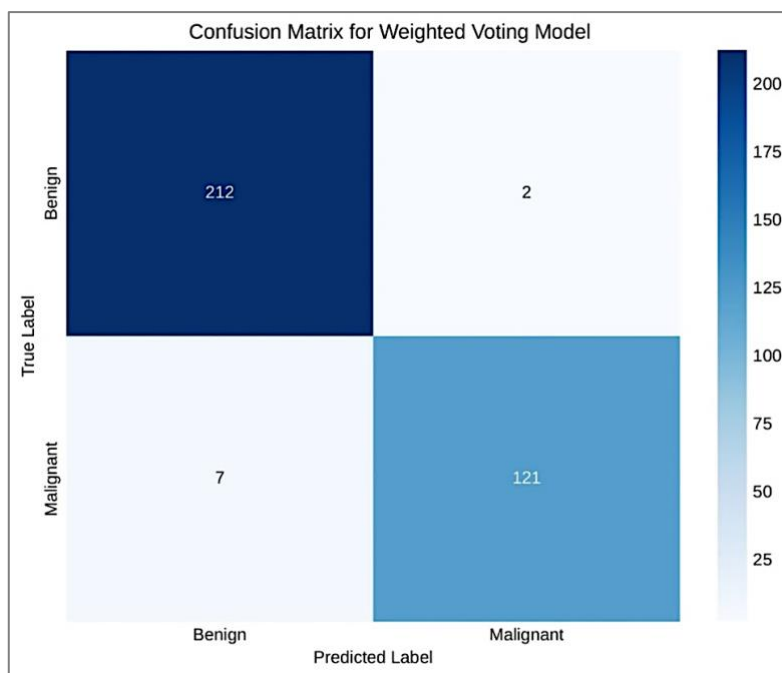


Figure 3. Confusion Matrix for Weighted Voting Model

Figure 3 presents the confusion matrix for the Weighted Voting ensemble model, offering a granular view of its classification performance. The matrix summarises the predictions against the actual labels.

- True Negatives: 212 cases were correctly identified as benign.
- False Positives: 2 benign cases were misclassified as malignant.
- True Positives: 121 cases were accurately identified as malignant.
- False Negatives: 7 malignant cases were misclassified as benign.

These results demonstrate the model's strong ability to distinguish between benign and malignant tumours. The extremely low number of false positives is crucial for avoiding unnecessary patient anxiety and invasive procedures. Similarly, the very low count of false negatives, the most critical type of error in cancer diagnosis, indicates the model's high reliability in detecting malignancy.

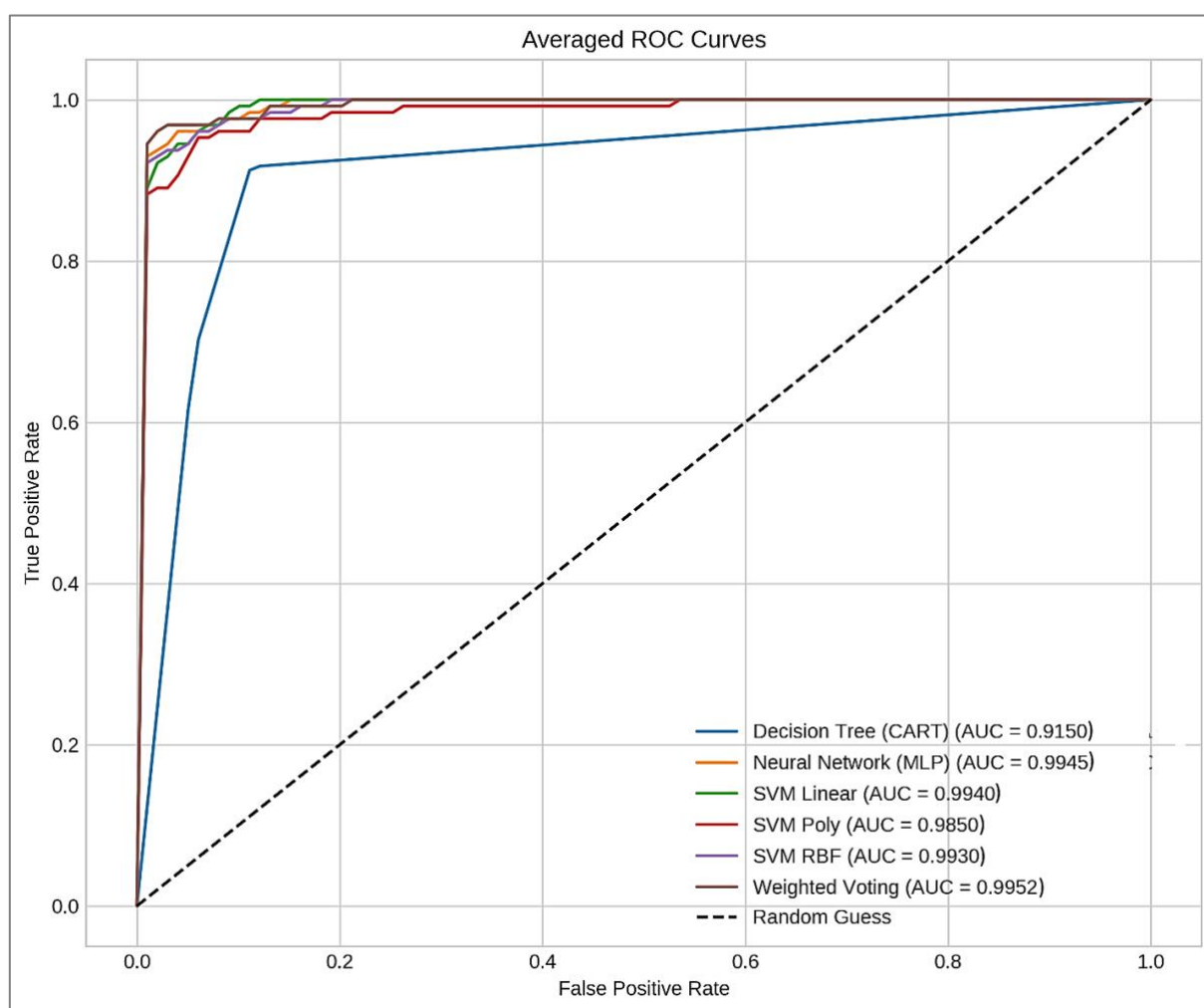


Figure 4. Averaged ROC Curves show the performance of classification models

Figure 4 shows the ROC curves for each model, averaged over 10-fold cross-validation. The AUC indicates performance; a value of 1 indicates a perfect classifier. All models significantly exceed random guessing, as indicated by the dashed line at an AUC of 0.5. The curves for the MLP and SVM variants are clustered closely together in the top-left corner, indicating excellent discrimination capability. The Decision

Tree (CART) curve is visibly lower, corresponding to its lower AUC of 0.9150. Critically, the red line representing the Weighted Voting model envelops the others, consistently demonstrating a higher true positive rate for any given false positive rate. This visual evidence is confirmed by its highest average AUC score of 0.9952, which reinforces its superiority as the most effective classifier among those tested.

The boxplots in Figure 5 depict the distribution of Accuracy and F1-scores for each model across the 10 cross-validation folds. This visualisation reveals not only the median performance but also the models' stability and consistency. For both Accuracy and F1-score, the Weighted Voting model exhibits the highest median value and the tightest distribution (a small interquartile range). This indicates that the ensemble model is not only more accurate on average but also more reliable, with its performance showing little variation across different subsets of the data. Conversely, the Decision Tree and SVM Poly models display wider distributions, suggesting their performance is less stable and more dependent on the specific training data split.

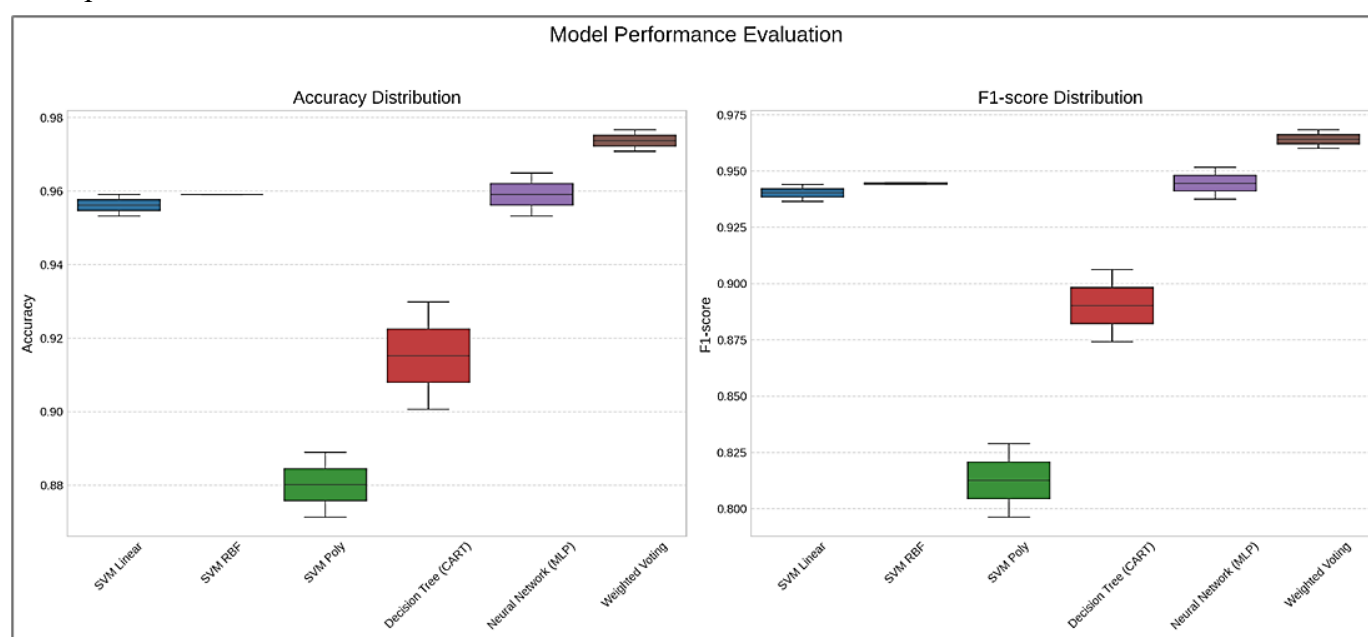


Figure 5. Boxplot for Accuracy and F1-score Distribution for each model

Figure 6 displays the results of the permutation feature importance analysis for the best-performing model, the Weighted Voting ensemble. This technique assesses a feature's importance by measuring the decrease in model performance (e.g., AUC) when the feature's values are randomly shuffled. A larger decrease implies a greater importance.

The analysis reveals that features associated with the "worst" measurements of cell nuclei are the most influential predictors. The top five most important features are concave points_worst, radius_worst,

texture_worst, area_se, and perimeter_worst. The high importance of "worst" concave points, radius, and perimeter suggests that the most extreme irregularities in a tumour's shape and size are strong indicators of malignancy. Features such as smoothness_se, symmetry_mean, and concavity_se have a minimal impact on predictions and provide limited information for classification.

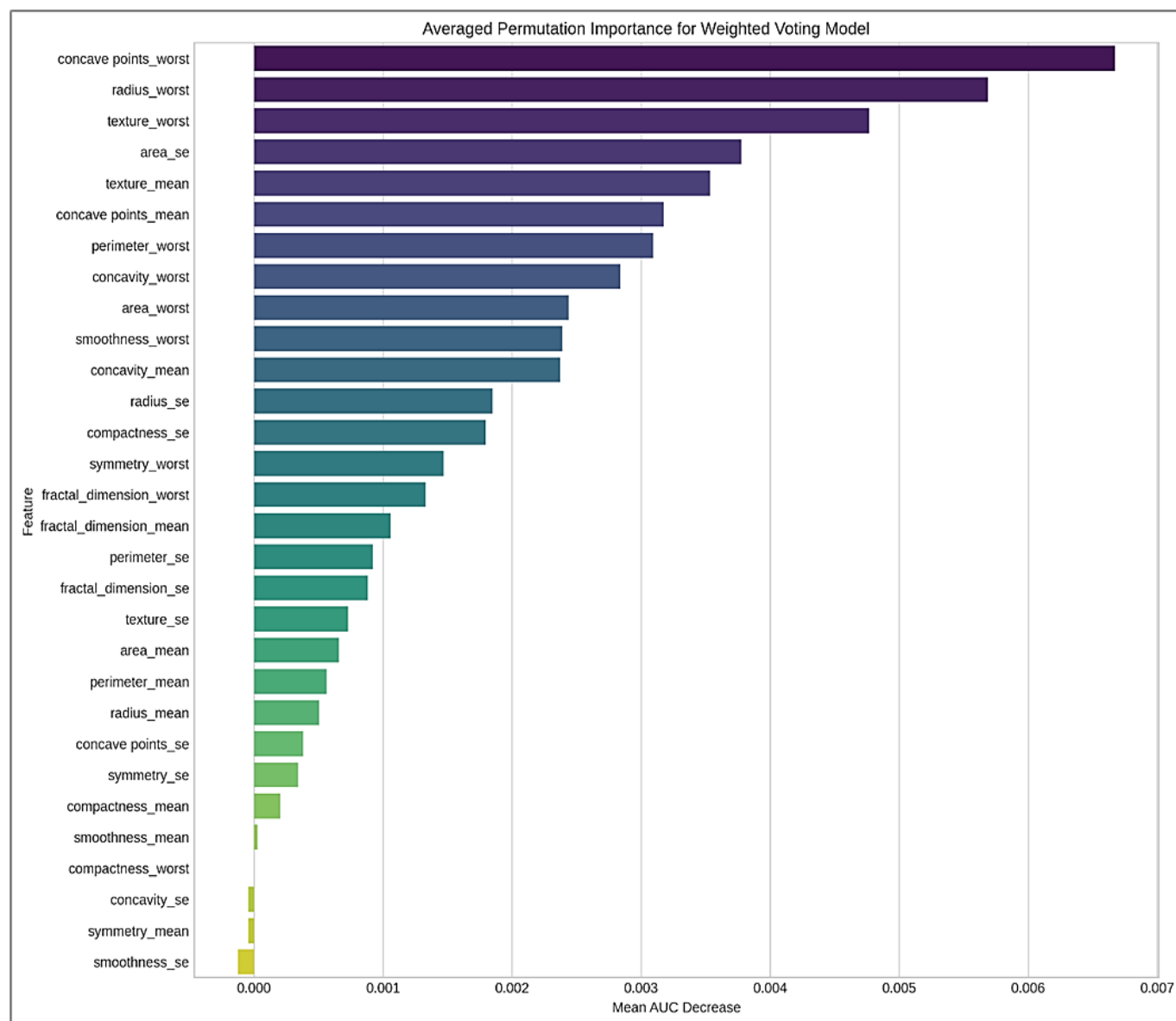


Figure 6. Averaged permutation importance for the weighted voting model

This study developed a hybrid voting ensemble algorithm for breast cancer prediction and demonstrated superior performance compared to individual machine-learning models. The results indicate the promise of collaborative ensemble approaches to improve oncological diagnostic accuracy [1, 2].

The best-performing model in this study was the Weighted Voting ensemble model, and it surpassed all five base learners of Decision Tree (CART), Neural Network (MLP), SVM Linear, SVM Polynomial, and SVM RBF in measures such as accuracy (0.9737), F1-score (0.9641), and AUC (0.9952). It

demonstrates the benefit of combining classifiers: the ensemble model leverages their strengths and mitigates their weaknesses to produce more accurate estimates. Although the MLP and RBF-SVM estimators achieved high individual accuracies of 0.9591, the ensemble method further improved performance.

Conversely, the SVM Poly model achieves the lowest accuracy of 0.8801, but the highest precision is 0.9780, indicating its conservative nature in predicting malignancy cases. This trade-off between precision and recall is a challenge in medical diagnosis, but the ensemble model offers a viable solution. Although the decision tree model's accuracy is lower, its explanatory power remains helpful in clinical practice. Overall, integrating interpretable and high-performing models provides a more effective and efficient diagnostic solution.

The permutation feature importance analysis further revealed that the "worst" measurements of cell nuclei, such as `concave points_worst`, `radius_worst`, and `perimeter_worst`, were the most influential predictors. This aligns with established histopathological criteria in cancer diagnosis, where the most extreme irregularities in a tumour's size and shape are strong indicators of malignancy. The high importance of these "worst" features suggests that the models are learning clinically relevant patterns from the data.

The application of machine learning to breast cancer diagnosis is a well-established research area, with numerous studies demonstrating the efficacy of various algorithms. The performance of the individual models in this study is consistent with previous findings on the WDBC dataset, which show the utility of SVMs, neural networks, and decision trees [3, 4].

However, this study's primary contribution is the implementation and evaluation of a hybrid voting ensemble that systematically optimises a diverse set of powerful classifiers. While other studies have explored ensemble methods, they often focus on homogeneous ensembles (e.g., Random Forest) or standard heterogeneous ensembles. The tailored approach that integrates MLP's ability to model non-linear relationships, the varied decision boundaries of multi-kernel SVMs, and the transparency of CART within an optimised weighted-voting framework constitutes a novel contribution. The superior results achieved by this hybrid model advance the field by demonstrating a practical and effective strategy for combining the predictive power of disparate models.

The high accuracy, precision, and recall of the proposed hybrid voting ensemble have significant clinical implications. An accurate and reliable diagnostic tool can aid clinicians in the following ways:

- **Enhanced Diagnostic Accuracy:** The model's capability to differentiate between benign and malignant tumours with high accuracy can serve as a valuable second opinion for pathologists and radiologists, potentially reducing the rate of misdiagnosis [5, 6].
- **Reduction of Unnecessary Procedures:** The extremely low number of false positives (2 out of 214 benign cases) is crucial for avoiding unnecessary patient anxiety and invasive procedures like biopsies [7].
- **Enhanced Clinical Decision-Making:** By providing a reliable prediction based on a comprehensive analysis of cytological features, the model can support clinicians in making more informed decisions about patient management and treatment planning.
- **Potential for Early Detection:** While this study focused on diagnosis, the underlying principles can be extended to develop models for early detection, which is paramount for improving patient prognosis and survival rates [8].

This study has some limitations despite its promising results. First, the models were developed and validated using a single dataset (WDBC), limiting their generalizability. Further testing on larger, diverse clinical datasets is needed. Second, while the selected base learners are effective, the study did not consider other advanced models, such as gradient boosting machines or deep learning architectures, which could enhance performance. Future research should focus on real-world validation and on expanding the ensemble to include a broader range of classifiers.

5. Conclusions

This study demonstrated the design, implementation, and superior performance of a novel hybrid voting ensemble algorithm for breast cancer prediction. The findings unequivocally demonstrate that the weighted voting ensemble outperforms the predictive capabilities of five individual base machine-learning models (MLP, SVM-Linear, SVM-Poly, SVM-RBF, and CART) across all key metrics, including accuracy, F1-score, and AUC.

The primary strength of the proposed model lies in its synergistic integration of diverse classifiers. By combining the capacity of MLP to model non-linear relationships, the varied decision boundaries of multi-kernel SVMs, and the interpretability of CART, the ensemble effectively leverages the individual strengths of each model while mitigating their respective weaknesses. This leads to a final prediction that is not only more accurate but also more stable and reliable, as evidenced by a tighter performance

distribution across cross-validation folds.

Furthermore, the feature importance analysis confirmed the model's clinical relevance, identifying the "worst" measurements of cell-nuclear characteristics as the most significant predictors of malignancy. This alignment with established histopathological criteria validates that the model is learning clinically meaningful patterns from the data.

Data Availability Statement: The dataset analysed in this study is publicly available in the University of California, Irvine (UCI) Machine Learning Repository. Additional processed data and experimental results are available from the corresponding author upon reasonable request.

Funding: The authors declare that no funding was received for this research.

Declaration of Competing Interest: The authors declare that they have no known competing interests.

Research Involving Human/Animal Participants: This article does not involve any studies conducted by the authors on animals or human participants.

References

- [1] A. S. Assiri, S. Nazir, and S. A. Velastin, "Breast tumor classification using an ensemble machine learning method," *Journal of Imaging*, vol. 6, no. 6, p. 39, 2020, doi: <https://doi.org/10.3390/jimaging6060039>.
- [2] F. Bray *et al.*, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 74, no. 3, pp. 229-263, 2024, doi: <https://doi.org/10.3322/caac.21834>.
- [3] O. Iparraguirre-Villanueva, A. Epifanía-Huerta, C. Torres-Ceclén, J. Ruiz-Alvarado, and M. Cabanillas-Carbonell, "Breast cancer prediction using machine learning models," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, 2023, doi: <https://doi.org/10.14569/IJACSA.2023.0140272>.
- [4] J. Ahmad *et al.*, "Deep learning empowered breast cancer diagnosis: Advancements in detection and classification," *PLOS ONE*, vol. 19, no. 7, p. e0304757, 2024, doi: <https://doi.org/10.1371/journal.pone.0304757>.
- [5] A. Miranda-Filho *et al.*, "The GLOBOCAN 2022 cancer estimates: Data sources, methods, and a snapshot of the cancer burden worldwide," *International Journal of Cancer*, vol. 156, no. 7, pp. 1336-1346, 2025, doi: <https://doi.org/10.1002/ijc.35278>.
- [6] L. Liao, "Inequality in breast cancer: Global statistics from 2022 to 2050," *The Breast*, vol. 79, p. 103851, 2025, doi: <https://doi.org/10.1016/j.breast.2024.103851>.
- [7] A. Singh and K. S. Kaswan, "Breast cancer diagnosis using soft voting classifier approach," presented at the 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP), 2024, doi: <https://doi.org/10.1109/INNOCOMP63224.2024.00055>.
- [8] A. Batool and Y. Byun, "Toward improving breast cancer classification using an adaptive voting ensemble learning algorithm," *IEEE Access*, vol. 12, pp. 12869-12882, 2024, doi: <https://doi.org/10.1109/ACCESS.2024.3356602>.
- [9] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Medicine*, vol. 17, no. 1, p. 195, 2019, doi: <https://doi.org/10.1186/s12916-019-1426-2>.
- [10] R. M. Hama Salih and S. H. Mahmood, "Application of support vector machines kernel functions for breast cancer detection

- and classification: A comparative study," *Cihan University-Erbil Scientific Journal*, vol. 9, no. 1, pp. 41-46, 2025, doi: <https://doi.org/10.24086/cuesj.v9n1y2025.pp41-46>.
- [11] M. M. Rahman, K. H. Kobir, S. Akther, and M. A. H. Kallol, "Ensemble machine learning for enhanced breast cancer prediction: A comparative study," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 7, 2024, doi: <https://doi.org/10.14569/IJACSA.2024.0150792>.
- [12] U. Naseem *et al.*, "An automatic detection of breast cancer diagnosis and prognosis based on machine learning using ensemble of classifiers," *IEEE Access*, vol. 10, pp. 78242-78252, 2022, doi: <https://doi.org/10.1109/ACCESS.2022.3174599>.
- [13] A. Rezaeipanah, R. Syah, S. Wulandari, and Arbansyah, "Design of ensemble classifier model based on MLP neural network for breast cancer diagnosis," *Inteligencia Artificial*, vol. 24, no. 67, pp. 147-156, 2021, doi: <https://doi.org/10.4114/intartif.vol24iss67pp147-156>.
- [14] S. H. Hasanah, "Classification support vector machine in breast cancer patients," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 16, no. 1, pp. 129-136, 2022, doi: <https://doi.org/10.30598/barekengvol16iss1pp129-136>.
- [15] M. Al-Raei, "The role of artificial intelligence in enhancing breast disease management: Early detection and prognostic innovations," *Archives of Medicine and Health Sciences*, vol. 13, no. 3, pp. 475-479, 2025, doi: https://doi.org/10.4103/amhs.amhs_351_24.
- [16] A. Tuerhong, M. Silamujiang, Y. Xianmuxiding, L. Wu, and M. Mojarad, "An ensemble classifier method based on teaching-learning-based optimisation for breast cancer diagnosis," *Journal of Cancer Research and Clinical Oncology*, vol. 149, no. 11, pp. 9337-9348, 2023, doi: <https://doi.org/10.1007/s00432-023-04861-5>.
- [17] F. K. Nasser and S. F. Behadili, "Breast cancer detection using decision tree and k-nearest neighbour classifiers," *Iraqi Journal of Science*, vol. 63, no. 11, pp. 4987-5003, 2022, doi: <https://doi.org/10.24996/ij.s.2022.63.11.34>.
- [18] S. H. Murad, N. B. Tayfor, N. H. Mahmood, and L. Arman, "Hybrid genetic algorithms-driven optimisation of machine learning models for heart disease prediction," *MethodsX*, vol. 15, p. 103510, 2025, doi: <https://doi.org/10.1016/j.mex.2025.103510>.
- [19] A. Bohr and K. Memarzadeh, "The rise of artificial intelligence in healthcare applications," in *Artificial Intelligence in Healthcare*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 25-60, doi: <https://doi.org/10.1016/B978-0-12-818438-7.00002-2>.
- [20] B. D. Shivahare *et al.*, "Delving into machine learning's influence on disease diagnosis and prediction," *The Open Public Health Journal*, vol. 17, no. 1, p. e18749445297804, 2024, doi: <https://doi.org/10.2174/0118749445297804240401061128>.
- [21] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209-249, 2021, doi: <https://doi.org/10.3322/caac.21660>.
- [22] S. Talatian Azad, G. Ahmadi, and A. Rezaeipanah, "An intelligent ensemble classification method based on multi-layer perceptron neural network and evolutionary algorithms for breast cancer diagnosis," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 34, no. 6, pp. 949-969, 2022, doi: <https://doi.org/10.1080/0952813X.2021.1938698>.
- [23] L. Aissaoui Ferhi, M. Ben Amar, F. Choubani, and R. Bouallegue, "Enhancing diagnostic accuracy in symptom-based health checkers: A comprehensive machine learning approach with clinical vignettes and benchmarking," *Frontiers in Artificial Intelligence*, vol. 7, p. 1397388, 2024, doi: <https://doi.org/10.3389/frai.2024.1397388>.
- [24] M. Ebrahim, A. A. H. Sedky, and S. Mesbah, "Accuracy assessment of machine learning algorithms used to predict breast cancer," *Data*, vol. 8, no. 2, p. 35, 2023, doi: <https://doi.org/10.3390/data8020035>.
- [25] T. Islam *et al.*, "Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI," *Scientific Reports*, vol. 14, no. 1, p. 8487, 2024, doi: <https://doi.org/10.1038/s41598-024-57740-5>.

-
- [26] S. Das, S. Koley, and T. Saha, "Machine learning approaches for investigating breast cancer," *Biosciences Biotechnology Research Asia*, vol. 20, no. 4, pp. 1109-1131, 2023, doi: <https://doi.org/10.13005/bbra/3163>.
- [27] G. Mallinis, N. Verde, S. Siachalou, D. Latinopoulos, C. Akratos, and I. Kagalou, "Evaluation of multiple classifier systems for mapping different hierarchical levels of forest ecosystems in the Mediterranean region using Sentinel-2, Sentinel-1, and ICESat-2 data," *Forests*, vol. 14, no. 11, p. 2224, 2023, doi: <https://doi.org/10.3390/f14112224>.
- [28] J. Kotary, V. D. Vito, and F. Fioretto, "Differentiable model selection for ensemble learning," presented at the Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, 2023, pp. 1954–1962, doi: <https://doi.org/10.24963/ijcai.2023/217>
- [29] N. K. Chauhan and K. Singh, "Performance assessment of machine learning classifiers using selective feature approaches for cervical cancer detection," *Wireless Personal Communications*, vol. 124, no. 3, pp. 2335-2366, 2022, doi: <https://doi.org/10.1007/s11277-022-09467-7>.
- [30] P. Thölke *et al.*, "Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data," *NeuroImage*, vol. 277, p. 120253, 2023, doi: <https://doi.org/10.1016/j.neuroimage.2023.120253>.