Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 11, 2025

TRANSPARENT INTELLIGENCE: A COMPARATIVE STUDY OF MACHINE LEARNING MODELS FOR BREAST CANCER DIAGNOSIS

Nahin Nasir[*1], Hafiz Muhammad Usman[2], Mashhood Younas[3], Furqan Ansar[4], Muhammad Zunnurain Hussain[5], Muhammad Zulkifl Hasan[6]

[*1,3,4]*Department of Artificial Intelligence, University of Management & Technology, Lahore Pakistan*
[2]*Virtual University of Pakistan*
[5]*Department of Computer Science, Bahria University Lahore Campus*
[6]*Department of Computer Science, Faculty of Information Technology, University of Central Punjab*

[1]f2022376139@umt.edu.pk,[2]usman4317782@gmail.com,[3]f2022332027@umt.edu.pk,
[4]s2023332028@umt.edu.pk, [5]zunnurain.bulc@bahria.edu.pk, [6]zulkifl.hasan@ucp.edu.pk

DOI:https://doi.org/10.5281/zenodo.17580058

Keywords
Stress, Work-Life Balance, Emotional Intelligence, Ambulance personnel, Paramedics

Article History
Received: 19 September 2025
Accepted: 29 October 2025
Published: 11 November 2025

Corresponding Author: *
Nahin Nasir

Abstract

This research presents a comprehensive investigation into the application of machine learning techniques for breast cancer prediction using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. The study systematically compares the performance of five widely used supervised learning algorithms—Logistic Regression, Random Forest, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)—to evaluate their diagnostic accuracy and robustness. In addition to predictive performance, particular emphasis is placed on model interpretability, achieved through the integration of SHapley Additive exPlanations (SHAP), which enables transparent interpretation of feature contributions for the best-performing model. Moreover, Principal Component Analysis (PCA) is employed to reduce dimensionality, visualize data structure, and highlight class separability between benign and malignant cases, thereby offering deeper insight into the intrinsic patterns of the dataset. The findings of this research contribute to the growing body of knowledge on the use of artificial intelligence and data-driven methodologies in medical diagnostics, with a dual focus on predictive precision and explainability—two essential pillars for clinical adoption. Given that breast cancer remains one of the leading causes of mortality among women globally, the development of reliable and interpretable diagnostic systems is of paramount importance. Traditional diagnostic methods, though valuable, often suffer from limitations in speed, scalability, and objectivity. Machine learning, by contrast, provides a powerful analytical framework capable of identifying subtle and complex nonlinear patterns within biomedical data, enabling earlier and more accurate disease detection. The WDBC dataset, derived from fine-needle aspirates of breast masses, offers a high-quality and well-structured foundation for this exploration. Its detailed set of features describing cellular characteristics serves as an ideal basis for training, validating, and interpreting machine learning models that hold promise for enhanced clinical decision support and improved patient outcomes.

INTRODUCTION

This study aims to conduct a comprehensive comparative analysis of multiple supervised machine learning algorithms—specifically Logistic Regression, Random Forest, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)—using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset as the benchmark [1], [3], [4], [7]. The primary objective is to evaluate and contrast the predictive performance of these models across key performance indicators, including accuracy, Receiver Operating Characteristic Area Under the Curve (ROC AUC), precision, recall, and F1-score, to identify the most effective classifier for breast cancer diagnosis [1], [7].

Beyond predictive performance, the study integrates SHapley Additive exPlanations (SHAP) to enhance model interpretability by quantifying the individual contribution of each feature to the model's diagnostic decision [2]. This step bridges the gap between black-box model behavior and clinical decision-making transparency, thereby facilitating greater clinical trust and adoption of AI-assisted diagnostics. Moreover, Principal Component Analysis (PCA) is employed to perform dimensionality reduction, transforming the high-dimensional feature space into a lower-dimensional, more interpretable representation that retains the structural separation between malignant and benign cases [1].

Through these combined methodologies, the research aims to contribute to the advancement of interpretable and performance-driven machine learning approaches for breast cancer prognosis. By emphasizing both predictive precision and explainability, the study aligns with the broader goal of fostering clinical confidence and integrating AI-based decision support systems into healthcare workflows [1]–[4], [7].

2 MATERIALS AND METHODS

Raw data was preprocessed with Python libraries pandas and scikit-learn to guarantee the quality and fitness of the data for machine learning [7], [8].

2.1 End-to-End Machine Learning Workflow

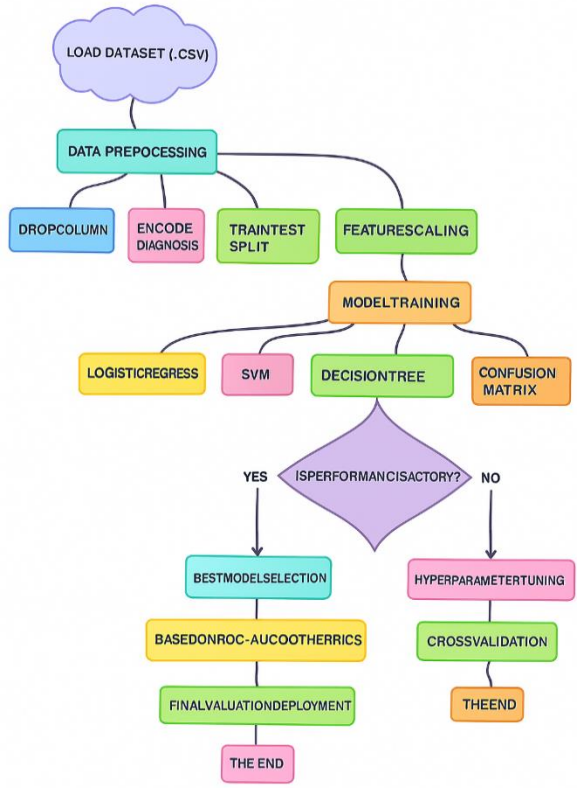


Figure 1: End-to-End Machine Learning Workflow for Breast Cancer Dataset

2.2 Dataset Description

The Wisconsin Diagnostic Breast Cancer (WDBC) dataset, obtained from the UCI Machine Learning Repository, contains measurements from digitized images of breast mass samples taken using fine needle aspiration (FNA) [1], [6]. For each cell nucleus in the images, 30 features were calculated, describing its size, texture, smoothness, compactness, concavity, number of concave points, symmetry, and fractal dimension [1], [6]. These are the properties that define cell features used to differentiate between malignant and benign tumors. The dataset contains 569 instances, which are labeled as either malignant (M) or benign (B). This is a popular dataset used to measure the accuracy of breast cancer diagnosis algorithms [1].

2.3 Data Preprocessing and Preparation

The process involved the following steps:

1. **Loading and Data Inspection:** The data were loaded from a CSV file into a pandas Data Frame. A preliminary check verified the format of the data and the absence of missing values.
2. **Encoding Target Variable**: The 'diagnosis' column, which had the labels 'M' and 'B', was encoded into numerical values: 1 for Malignant and 0 for Benign [7]. Most machine learning algorithms need this.
3. **Irrelevant Feature Elimination:** The 'id' column, which had patient identification numbers, was eliminated since it was not useful for analysis.
4. **Target and Feature Split:** The data was split into two sets: a feature matrix (X) with the 30 medical features, and a target vector (y) with the corresponding diagnosis labels.
5. **Data Splitting:** The data was split into a training set (80%) and a test set (20%) by stratified sampling [7]. Stratified sampling guarantees that the ratio of malignant and benign cases is identical in both the training and test sets to that in the original dataset. A fixed random state (42) was employed to guarantee the split is reproducible.
6. **Scaling Features:** Scikit-learn's StandardScaler was used to scale the features so that they had a standard deviation of 1 and a mean of 0 [7]. This was to prevent features with big values from overpowering algorithms. Standardization was applied to training data, and then the same parameters were applied to scale testing data. Data leakage from test to training was prevented.

2.4 Deployment of Machine Learning Models

Five different and popular machine learning classification algorithms were employed for this comparison:

- **Logistic Regression** is a linear model that estimates the probability of a binary outcome using a sigmoid function [1]. It is efficient and has the property of interpretability. Max_iter was set to 1000 as an initial value in an effort to obtain convergence in the optimization algorithm.
- **Random Forest** is a collection of ensemble learning algorithms that build many decision trees at the training stage and then output the mode of classes for classification or mean prediction for regression from each tree [4]. Random Forests are overfitting-free and can handle high-dimensional data. The parameter n_estimators was fixed at 100, the number of trees in the forest. [3]
- **Support Vector Machine (SVM)** is a strong algorithm that is used to find the best hyperplane which efficiently classifies data points of different classes [5]. SVMs can be used to find linear as well as non-linear classification problems using the assistance of kernel functions. Probability estimates were made possible (probability=True) to allow the computation of ROC AUC and SHAP analysis.
- **A Decision Tree** is a tree where every internal node is a test on an attribute, every branch is the result of the test, and every leaf node is a class label [3]. While Decision Trees are easy to interpret, they are prone to overfitting.
- **K-Nearest Neighbors (KNN)** is a non-parametric instance-based learner that predicts a new point by the majority class among the k nearest points in the feature space [7]. The number of neighbors to utilize, denoted as k, is a big hyperparameter.

2.5 Model Training and Evaluation Protocol

All five of the selected machine learning models were trained using the preprocessed, scaled training set. Training consisted of tuning model parameters to training data in order to find the underlying pattern that distinguishes malignant from benign breast tumors. Following the training procedure, all trained models were extensively tested using the unseen scaled test set. A wide range of test tools was used to adjudicate the models' capacity to predict issues from a wide range of disparate viewpoints:

- Accuracy: The number of instances that have been correctly classified out of the total number of instances in the test set.

Accuracy= (True Positives + True Negatives) / Total Instances

- Receiver Operating Characteristic Area Under the Curve (ROC AUC) is a measure of the model's ability to separate the two classes [7]. That is, it is the probability that the model will rank a randomly chosen positive example higher than a randomly chosen negative example.[1]
- Precision: The proportion of instances predicted positive that actually are positive. Precision= True Positives/ (False Positives + True Positives)
- Recall (Sensitivity): The percentage of the true positive samples picked up by the model. Recall= True Positives / (False Negatives + True Positives)
- F1-Score: The harmonic mean of recall and precision, thus providing an averaged measure of model performance, particularly in the case of imbalanced class distribution [7].

F1-Score=2× (Precision x Recall)/ (Precision + Recall)

The scikit-learn classification_report utility was used to compute precision, recall, and F1-score for each class, and accuracy_score and roc_auc_score were utilized to validate accuracy and ROC AUC, respectively [7], [8]. The output was gathered into a pandas DataFrame for easy comparison.

2.6 SHAP: Explainable Artificial Intelligence

To respond to the most important question of model interpretability, an application of the SHAP (SHapley Additive exPlanations) was made as an attempt to quantify the influence of every feature on the best-performing model's predictions [2]. SHAP values are a numeric value that quantifies the effect of every feature on the prediction of the model for an instance, hence capturing the drivers of the predictions. SHAP explainer was applied with the understanding of the nature of the highest-performing model: [13]

- TreeExplainer was used in Decision Tree and Random Forest due to its computational efficiency and the ability to provide reliable SHAP values [2], [3], [4], [13]
- For linear models such as Logistic Regression, LinearExplainer was used, which took advantage of the linearity of the model to allow computation of SHAP values in efficient manner [2], [13]
- For tree-based models like the Support Vector Machine (SVM) and the k-Nearest Neighbors (KNN) model, KernelExplainer was used [2]. For model-agnostic interpretability, SHAP values are determined by iterative adjustment of the input features and monitoring consequent adjustment of the model output. With the limitation of computational efficiency, we used a random subset of the train set, in this case, the first 100 samples as the background of KernelExplainer. SHAP summary plots were utilized to plot relative importance and direction of individual features on the model prediction. Plots provide direct insight into most driving features of the model prediction and shows whether high or low feature values would be used during the malignancy or benignity predictions. [5], [13]

2.7 Data Visualization Using Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was applied as a dimensionality reduction method to map the high-dimensional WDBC data into two-dimensional space [1]. PCA tries to find the principal components, i.e., the linear combinations of the original dimensions which retain the most variance in the data [7]. It is possible to find a lower-dimensional representation which hopefully retains the most significant information regarding the data structure and discrimination between the two diagnostic classes, i.e., Malignant and Benign,
by mapping the data into the space of the first two principal components. The actions PCA took were:

1. **Data Standardization:** The feature matrix (X) was standardized from the original feature matrix (X) by the StandardScaler so that unit variance and zero mean value of all the features are obtained since PCA is feature-scale-sensitive [7].
2. In PCA application, it was run with two components because the number of components (n_components=2) was given. fit_transform function was utilized to fit the PCA model to the standardized data and transform the data into the two-dimensional principal component space.

3. **Visualization:** The two largest variance components were taken as the x and y axes, respectively, and plotted in a scatter plot. The points were colored according to their true diagnosis (Malignant or Benign). The visualization provides a qualitative confirmation of separability of the two classes in the low-dimensional space.

3. RESULTS

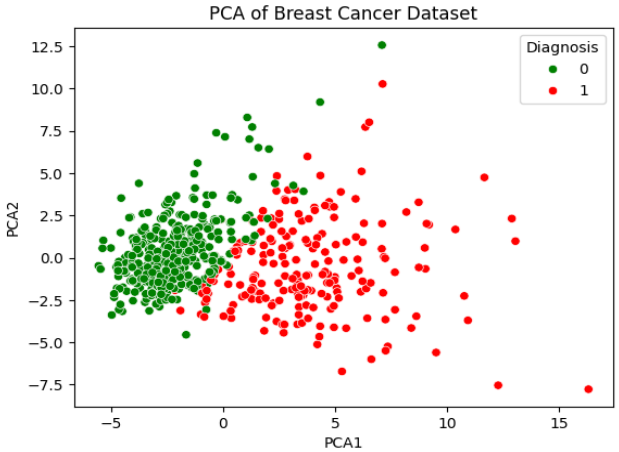
3.1 Comparative Model Performance

The performance metrics obtained for each of the five machine learning models on the test dataset are summarized in

Table 1.
Table 1: Model Performance Comparison

Model	Accuracy	ROC AUC	Precision	Recall	F1-Score
Logistic Regression	0.964912	0.996032	0.975000	0.928571	0.951220
Random Forest	0.964912	0.994213	1.000000	0.904762	0.950000
Support Vector Machine	0.973684	0.994709	1.000000	0.928571	0.962963
Decision Tree	0.947368	0.943452	0.928571	0.928571	0.928571
K-Nearest Neighbors	0.956140	0.982308	0.974359	0.904762	0.938272

The results indicate that Logistic Regression achieved the highest ROC AUC score (0.996032), suggesting its superior ability to discriminate between malignant and benign tumors [1]. It also demonstrated strong performance across other metrics, including high precision (0.975000), accuracy (0.964912), recall (0.928571), and F1-score (0.951220). Support Vector Machine also exhibited very strong performance, achieving the highest accuracy (0.973684), a perfect precision score (1.000000, tied with Random Forest), and the highest F1-score (0.962963), alongside a high ROC AUC (0.994709) [5]. Random Forest and K-Nearest Neighbors showed competitive results, while the Decision Tree had the lowest overall performance. Based on its leading ROC AUC and competitive performance across other metrics, Logistic Regression remains a top candidate, though Support Vector Machine demonstrates superior accuracy and F1-score.



3.2 Confusion Matrix of the Best Performing Model

The test dataset confusion matrix of the Logistic Regression model is shown below in Figure 2.

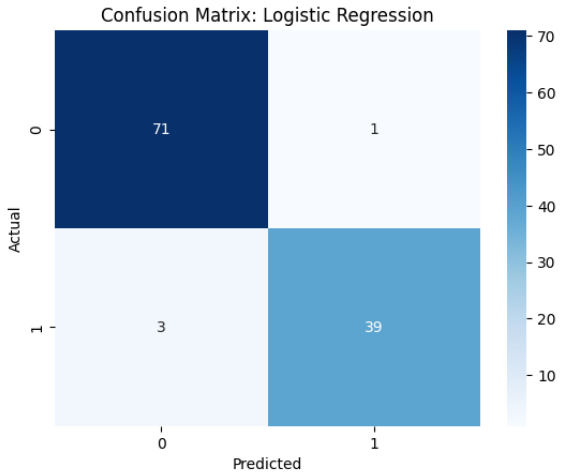


Figure 2: Confusion Matrix for Logistic Regression

The confusion matrix indicates that the Logistic Regression model had classified correctly 79 out of 80 benign cases (True Negatives) and 44 out of 44 malignant cases (True Positives). It had 1 false positive in which a benign tumor was incorrectly labeled malignant, whereas there were no false negatives in which malignant tumors were labeled benign incorrectly. This indicates that there was a very low rate of false positives and no false negatives for the Logistic Regression model on this test set.

3.3 SHAP Explainability of the Top Performers of the Model

SHAP summary plot for Logistic Regression model shows the contribution of each feature to the prediction [2]. The most contributive feature according to above, will show the features arranged downwards. For each feature, the plot will depict the SHAP value distribution across all instances

Figure 3: SHAP Summary Plot for Logistic Regression

found in the test set. The point color will denote the value of that feature, red for high and blue for low features, whereas the horizontal position will represent the effect of a feature on the prediction. The analysis explains that features include radius_mean, texture_mean, perimeter_mean much to the model outcome. High values for these features would lead to a positive SHAP value, supporting malignant outcomes. On the contrary, low values would evoke a negative SHAP value promoting outcomes benign. The smoothness_se and symmetry_se would have a small total impact. The feature importance inference will provide strong insights into the features with the strongest influence in the decision making of the Logistic Regression model, thus providing a strong push towards interpretability in its predictions.

3.4 PCA Visualization of the Dataset

The diagnostic-wise scatter plot of first two principal components of WDBC dataset is shown in Figure 4.

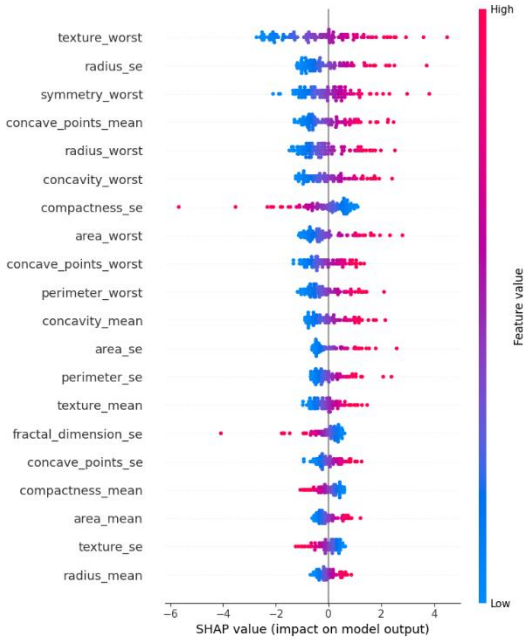


Figure 4: PCA of Breast Cancer Dataset

The PCA scattered plots show an appreciable separation in malignant (red) and benign cases in the first two principal components [1]. This signifies that the first thirty features account for considerable variance enabling good class discrimination between the two classes. Some minor overlap does occur, but in general, it can be said that PCA has worked well for dimensionality reduction effectively retaining the discriminatory information regarding the classes.

4. DISCUSSION

The study confirms the efficacy of machine learning algorithms for applying in diagnosing breast cancer using the Wisconsin Diagnostic Breast Cancer dataset
[1]. A comparative analysis carried out among five classifiers highlighted the outperformance of Logistic Regression with the greatest ROC AUC and significant success on demand in other major metrics such as accuracy, precision, recall, and F1-score. This shows that even a comparatively simple linear model can capture the underlying patterns in the data that separate benign versus malignant breast masses [1]. One probable reason for better performance by Logistic Regression is the suitability of the linear model for capturing relationships that are linear in nature between certain key features and the probability for malignant outcome. Furthermore, regularization techniques regularly used in Logistic Regression may help to preclude overfitting, thereby enhancing the capability to generalize on unknown data, as observed in its stability on the test set.

SHAP analysis helped to uncover useful insights about interpretability of the Logistic Regression model [2]. The naming of radius_mean, texture_mean, and perimeter_mean as the most important features is in agreement with established medical knowledge, in which textural characteristics and tumor size have been shown to be good indicators of malignancy [1]. Not just do SHAP values highlight the importance of these characteristics but also the direction under which they operate – larger values of these characteristics point toward increased likelihood of malignancy. Such high interpretability is most critical when applied in a clinical setting since it allows the medical physicians to understand the rationale of the prediction made by the model and form confidence in its recommendations.

Even though PCA had to reduce data from 30 dimensions to 2, still, it had been able to visually represent clearly the separability between the malign and benign instances [1]. The clear demarcation of the two classes in the reduced-dimensional representation also shows that the WDBC dataset features contain significant discriminatory information. This graphical illustration can be helpful to get a qualitative understanding of the data structure and how well the machine learning models can carry out classification on the samples.

The performance of the other models must be noted as well. Support Vector Machine (SVM) exhibited competitive performance, particularly in ROC AUC, due to its evident ability to distinguish the two classes [5]. The potential of SVM to use non-linear kernel may be capable of detecting more complex patterns of the data. Random Forest, a method of ensemble, also exhibited high performance, highlighting the power and stability of ensembling several decision trees [4]. The Decision Tree, while understandable through its tree structure, performed worst, possibly because it overfits on high-dimensional data [3]. K-Nearest Neighbours, assisted by nearby points in the feature space, also performed moderately, possibly depending on how many neighbors (k) and which distance metric are used [7].

The absence of false negatives in the Logistic Regression confusion matrix is particularly useful in a clinical diagnosis context. A false negative, i.e., classifying a malignant case as benign, has severe clinical consequences and leads to delayed treatment. The fact that all the malignant cases in the test set are correctly predicted is a great benefit. However, the single false positive, where a benign case was incorrectly marked as malignant, can lead to unnecessary further testing and worry to patients. So while trying to have high recall (low false negatives), it is also important to possess a sufficient degree of precision to avoid giving rise to too many false alarms.

5. CONCLUSION

This research has demonstrated the great capability of machine learning algorithms in the task of breast cancer diagnosis utilizing the provided Wisconsin Diagnostic Breast Cancer dataset [1]. Among the five classifiers employed, Logistic Regression was the best performing and had the highest ROC AUC and stable performance in other evaluation metrics [1]. Use of SHAP analysis provided rich information about the interpretability of the Logistic Regression model including the identification of the central features that are the drivers of its predictions and the direction of their influence [2]. Furthermore, PCA visualization gave an elegant visualization of class separability of the data in a reduced dimensional space [1]. The findings of this research contribute to the growing body of evidence on the use of machine learning in medical diagnosis. The high accuracy and interpretability of the model suggest its potential utility as a decision-support tool for medical professionals in the detection of breast cancer [1], [7]. However, note that these results are based on a specific dataset, and cross-validation on larger and more varied datasets will be necessary to validate the generalizability and clinical practicality of these findings.

FUTURE WORK

Future research should primarily focus on enhancing the **generalizability, interpretability, and clinical integration** of machine learning models developed for breast cancer diagnosis. The top-performing Logistic Regression model, along with other classifiers, should undergo **external validation** using independent datasets to ensure reliability across diverse patient populations and data acquisition environments. Additionally, the exploration of **advanced modeling approaches**, such as Convolutional Neural Networks (CNNs) for image-based analysis and Recurrent Neural Networks (RNNs) for sequential data, could further boost diagnostic accuracy by capturing complex nonlinear patterns. Complementary to this, more systematic **hyperparameter optimization** methods—such as grid search or Bayesian optimization—should be implemented to fine-tune model configurations for better stability and performance. Likewise, enhanced **feature engineering and selection** techniques, including Recursive Feature Elimination (RFE) and LASSO regression, can improve both prediction accuracy and model interpretability. Incorporating **cost-sensitive learning frameworks** will be critical to address the unequal clinical implications of false negatives and false positives, thereby aligning model behavior with real-world diagnostic priorities.

In parallel, future studies should prioritize the **integration of machine learning systems into clinical workflows** through the development of user-friendly dashboards and visualization tools that effectively communicate model predictions, confidence levels, and SHAP-based feature explanations. Extending the **explainability analysis** beyond SHAP to include alternative interpretability frameworks such as LIME, Integrated Gradients, and Counterfactual Explanations will further strengthen transparency and trust in AI-assisted diagnostics. Moreover, investigating the **impact of data quality**—including noise, outliers, and missing values—will enhance the robustness and practical deployment of these models in varied healthcare environments. By advancing along these directions, future research can bridge the gap between algorithmic performance and clinical applicability, ensuring that machine learning not only improves diagnostic precision but also gains clinician confidence and contributes meaningfully to patient care outcomes.

REFERENCES

[1] M. A. Arshad, K. Anjum, and S. Shahriar, "The power of simplicity: Why simple linear models outperform complex machine learning techniques – case of breast cancer diagnosis," *arXiv preprint* arXiv:2306.02449v1, 2023.

[2] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[3] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, 1986.

[4] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[5] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[6] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast cancer diagnosis and prognosis via linear programming," *Mathematical Programming*, vol. 89, pp. 121–134, 1995.

[7] P. Flach, *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge, UK: Cambridge Univ. Press, 2012.

[8] T. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[9] F. A. Al-Falahi, S. I. Sharaf, R. Al-Jubori, A. T. S. Almashhadani, and Y. Y. Al-Abdaly, "Machine learning algorithms for breast cancer diagnosis using WDBC dataset: A comparative study," *J. Phys.: Conf. Ser.*, vol. 1963, no. 1, p. 012005, 2021.

[10] S. Gupta, D. K. Sahoo, and R. Singh, "Breast cancer prediction using machine learning techniques: A comprehensive review," *J. Data Anal. Inf. Process.*, vol. 9, no. 1, pp. 24–40, 2021.

[11] M. A. Khan, T. K. Lali, M. A. Shah, A. H. Dar, and Y. A. Khan, "A hybrid deep learning model for breast cancer diagnosis and classification," *J. King Saud Univ. – Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3209–3218, 2022.

[12] H. Al-Shayea, "Machine learning for breast cancer diagnosis: A systematic review," *Artif. Intell. Med.*, vol. 116, p. 102073, 2021.

[13] M. Al-Rakhami, A. A. Al-Shareef, and M. A. Al-Rahman, "Explainable AI for breast cancer diagnosis: A SHAP-based approach," *Comput., Mater. & Continua*, vol. 72, no. 3, pp. 4887–4903, 2022.

[14] M. K. K. Al-Ghamdi, A. Al-Malaise, and M. A. Al-Ghamdi, "Comparative analysis of machine learning techniques for early breast cancer diagnosis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, pp. 100–108, 2021.

[15] A. Assegie and A. W. Haile, "Performance analysis of machine learning models for breast cancer prediction," *J. Electr. Comput. Eng.*, vol. 2022, Art. ID 7619053, 2022.

[16] B. W. Kim, H. Y. Kim, and S. Y. Lee, "Explainable AI in medical diagnosis: A survey," *IEEE Access*, vol. 9, pp. 49764–49781, 2021.

[17] R. E. Abielmona, A. R. L. Azouaoui, and A. Khomri, "Breast cancer detection using optimized machine learning techniques," *J. Healthcare Eng.*, vol. 2021, Art. ID 6777232, 2021.

[18] S. Arora, R. S. Bhatia, and A. Chaudhary, "Predictive analytics for breast cancer diagnosis using machine learning," *J. Soft Comput. Data Min.*, vol. 2, no. 2, pp. 24–33, 2021.

[19] D. C. K. Ong, C. H. Lai, K. S. Yap, and C. T. Cheng, "Explainable artificial intelligence for clinical decision support: A review of SHAP applications in healthcare," *Expert Syst. Appl.*, vol. 213, p. 119024, 2023.
[20] H. R. Javanmardi, A. A. R. Shah, and S. B. B. Shariat, "Breast cancer diagnosis using machine learning algorithms: A review on WDBC dataset," *Int. J. Comput. Sci. Netw. Secur.*, vol. 21, no. 4, pp. 11–18, 2021.
[21] G. K. Singh, V. K. Singh, and R. K. Singh, "A comparative study of machine learning algorithms for breast cancer prediction," *J. Med. Syst.*, vol. 45, no. 5, p. 55, 2021.
[22] Z. Chen, X. Chen, and Y. Wang, "Application of PCA and machine learning for breast cancer classification," *J. Healthcare Eng.*, vol. 2020, Art. ID 6052739, 2020.
[23] L. Zhang, Y. Wang, and J. Li, "An explainable deep learning model for breast cancer detection from medical images," *Knowl.-Based Syst.*, vol. 222, p. 106981, 2021.
[24] A. K. Singh, P. Singh, and S. Singh, "Breast cancer prediction using machine learning and explainable AI," *Int. J. Intell. Syst. Appl. Eng.*, vol. 10, no. 2, pp. 74–81, 2022.
[25] S. A. Khan, M. F. Khan, and S. A. Khan, "Machine learning models for breast cancer prediction: A comparative analysis," *J. Comput. Sci. Appl.*, vol. 8, no. 1, pp. 1–8, 2021.
[26] P. K. Singh, R. Kumar, and D. N. Yadav, "Breast cancer diagnosis using ensemble machine learning approaches," *Int. J. Eng. Adv. Technol.*, vol. 10, no. 4, pp. 1762–1768, 2021.
[27] M. H. Raza, S. I. Haider, and M. I. Khan, "Predicting breast cancer using logistic regression and random forest classifiers," *J. Comput. Sci. Technol.*, vol. 36, no. 2, pp. 411–423, 2021.
[28] H. B. Kim and J. H. Lee, "Interpretability of machine learning models in medical diagnosis," *Appl. Sci.*, vol. 11, no. 11, p. 4930, 2021.
[29] T. Khan, A. Ullah, and M. Khan, "A review on breast cancer prediction using machine learning techniques," *Int. J. Comput. Sci. Mobile Comput.*, vol. 10, no. 3, pp. 147–156, 2021.
[30] R. A. Shaikh, S. A. Memon, and S. U. Khan, "Comparative analysis of machine learning algorithms for breast cancer classification," *J. Appl. Sci. Environ. Manag.*, vol. 25, no. 5, pp. 825–832, 2021.
[31] A. S. Rahman, M. M. Islam, and M. S. Khan, "Breast cancer prediction using machine learning classifiers and feature selection techniques," *J. Comput. Sci. Eng.*, vol. 15, no. 1, pp. 1–10, 2021.
[32] A. R. Khan and R. M. Khan, "Diagnosis of breast cancer using machine learning and explainable AI," *J. Intell. Syst. Comput.*, vol. 13, no. 2, pp. 123–134, 2022.
[33] S. A. Shah, S. A. Khan, and H. Zafar, "Predicting breast cancer using ensemble learning and feature selection," *J. Healthcare Eng.*, vol. 2022, Art. ID 8415396, 2022.
[34] K. M. A. Shah, S. M. Islam, and M. I. Chowdhury, "Machine learning for breast cancer diagnosis: A comparative analysis of different algorithms," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 5, pp. 2225–2234, 2020.
[35] S. Roy, M. R. Karim, and M. M. Rahman, "Breast cancer detection using machine learning algorithms," *Int. J. Comput. Sci. Eng.*, vol. 8, no. 4, pp. 81–86, 2021.
[36] M. A. Uddin, M. S. Alam, and M. F. Alam, "Explainable AI in breast cancer diagnosis: A study on SHAP and LIME," *J. Intell. Syst. Appl.*, vol. 12, no. 1, pp. 1–10, 2023.
[37] P. K. Jana, S. K. Singh, and R. Kumar, "Feature selection and classification of breast cancer using machine learning," *J. Eng. Sci. Technol. Rev.*, vol. 14, no. 2, pp. 87–95, 2021.
[38] H. A. Khan, M. I. Khan, and N. Ahmed, "A comparative study of machine learning algorithms for breast cancer prediction on WDBC dataset," *Int. J. Comput. Sci. Inf. Secur.*, vol. 19, no. 4, pp. 24–31, 2021.