Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Research Article

Comparative Analysis of Machine Learning and Deep Learning Techniques for Early Prediction of Breast Cancer

Mohammed Al-Duais [1,*], Abdualmajed A.G. Al-Khulaidi [2], Fatma Susilawati Mohamad [3], Walid Yousef [4], Belal Al-Fuhaidi [4], Sadik Ali Murshid Al-Taweel [4], Mumtazimah Mohamad [3], Mohd Nizam Husen [5], and Nooraini Yusoff [6]

[1] Department of Computer Science, Faculty of Engineering and IT, Amran University, Amran, Yemen; e-mail : sarhan2w@gmail.com
[2] Department of Computer Science, Faculty of Computer Science & Information Systems, Sana'a University, Sana'a, Yemen; e-mail : alkhulaidi@su.edu.ye
[3] Department of Computer Science, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Besut Campus, Besut Terengganu 222200, Malaysia; e-mail : fatma@unisza.edu.my; mumtaz@unisza.edu.my
[4] Department of Computer Science, Faculty of Computing and IT, University of Science & Technology, Sana'a, Yemen; e-mail : w.shaher@ust.edu.ye; belarh@gmail.com; altaweel.sadik@gmail.com
[5] Malaysian Institute of Information Technology, Universiti Kuala Lumpur, Kuala Lumpur 50250, Malaysia; e-mail : mnizam@unikl.edu.my
[6] Faculty of Data Science and Computing, Universiti Malaysia Kelantan, City Campus, Kota Bharu 16100, Kelantan, Malaysia; e-mail : nooraini.y@umk.edu.my
* Corresponding Author : Mohammed Al-Duais

Abstract: Breast cancer remains one of the leading causes of death among women worldwide, primarily due to late detection and diagnosis. Early and accurate prediction is essential to improve survival rates. Machine learning (ML) techniques have proven effective in supporting early diagnosis. This study aims to evaluate and compare the performance of three different approaches: traditional ML, ensemble ML, and deep learning (DL) for early prediction of breast cancer using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. The methodology includes data collection, preprocessing, and the design of predictive models. Traditional ML algorithms used include Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes (NB), and Decision Tree (DT). Ensemble ML techniques comprise Random Forest (RF), XGBoost, and AdaBoost, while DL models include Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). The models were evaluated using precision, recall, F1-score, and accuracy. The results indicate that XGBoost achieved the highest accuracy (0.99), with strong recall (0.98) and F1-score (0.986), outperforming all other ensemble and traditional ML methods. CNN achieved 0.99 in all evaluation metrics, slightly outperforming RNN, which attained 0.98 accuracy and 0.985 F1-score. These findings confirm that ensemble ML techniques outperform traditional models, while CNN leads among DL models. Furthermore, the proposed models demonstrated superior prediction performance compared to existing studies, particularly in minimizing false negatives, which is critical for healthcare applications.

Keywords: Breast Cancer Prediction; CNN; Deep Learning; Early Diagnosis; Ensemble Learning; Machine Learning; XGBoost.

Received: January, 31st 2025
Revised: June, 9th 2025
Accepted: June, 10th 2025
Published: June, 25th 2025
Curr. Ver.: June, 25th 2025

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

Breast cancer has a severe impact on women, resulting in a significant number of deaths. According to the GLOBOCAN 2020 report, it is the world's second-deadliest cancer, accounting for 19.3 cases per 10 million deaths [1]. Breast cancer is the leading cause of death from all malignancies, particularly among women aged 35 to 55 [2]. Early identification is crucial, especially for individuals at high risk, as it can help reduce mortality rates. Accurate diagnosis, early detection, and increased awareness are key strategies in lowering these rates

[3], [4]. In recent years, the medical field has generated vast amounts of clinical data across various diseases, facilitating improved analysis and forecasting. Technological advancements have enabled healthcare professionals to make more precise estimations, helping reduce costs and ensure appropriate treatment [5], [6]. Early detection of breast cancer, in particular, has the potential to save lives [7]. Since breast cancer can be classified in multiple ways, the choice of classification technique plays a vital role in both diagnosis and prediction [8].

Numerous studies have demonstrated the effectiveness of various machine learning (ML) techniques for breast cancer prediction, particularly using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset [9]–[11]. ML has proven capable of extracting significant patterns from complex datasets, making it a valuable tool for predicting breast cancer outcomes [12], [13]. The main objective of this study is to compare traditional ML, ensemble ML, and deep learning (DL) techniques to evaluate their effectiveness in classifying and predicting breast cancer. Four primary types of breast cancer are typically identified: standard, benign, in situ, and aggressive carcinoma. A critical challenge remains the delay in diagnosis, which often results in inaccurate outcomes [14].

Previous studies can be broadly categorized into those employing traditional ML methods [15]–[18] and those utilizing ensemble ML techniques [19]–[21]. However, limited research has comprehensively compared these approaches alongside DL techniques to determine which offers superior performance in breast cancer diagnosis. Additionally, most prior work has focused solely on prediction or classification tools, without integrating both. This study addresses that gap by systematically comparing traditional ML, ensemble ML, and DL techniques using consistent evaluation metrics. Early detection improves the likelihood of successful therapy and long-term survival. When diagnosed and treated early, more than 98% of patients survive beyond five years, compared to around 26% when detected at a later stage [22]–[24]. In the context of machine learning, breast cancer diagnosis can be framed as a classification task, distinguishing between benign and malignant cases [25], [26].

Machine learning methods can assist in efficiently extracting knowledge from empirical data and help clinicians formalize insights that would otherwise be time-consuming or error-prone [27]. These algorithms are now widely used in medical decision-making due to their ability to process and analyze large volumes of data [28]. Over time, ML algorithms have shown considerable improvement in both accuracy and reliability for classification tasks. As data becomes increasingly complex and abundant, ML continues to evolve and enhance its performance [29].

This study contributes in three main ways: first, by exploring various traditional ML algorithms including Logistic Regression (LR), Naïve Bayes (NB), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Decision Tree (DT); second, by evaluating ensemble ML techniques such as XGBoost, Random Forest (RF), and AdaBoost; and third, by comparing their performance with DL techniques including Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN).

2. Related Works

Many studies have been conducted to address the challenge of breast cancer prediction. Hassan et al. [30] demonstrated two methods, i.e., Random Forest and XGBoost, for predicting breast cancer risk, achieving accuracy rates of 74.73% and 73.63%. Liu et al. [31] developed EXSA, a gradient-boosting approach to anticipate breast cancer progression and improve survival rates.

Chetry et al. [32] used the Wisconsin Breast Cancer Dataset (WBCD) from the UCI repository to evaluate the performance of various DL algorithms, including CNN, Feedforward Neural Network (FNN), and Gated Recurrent Unit (GRU). According to their results, CNN achieved the highest accuracy at 98.25%, followed by FNN at 96.49% and LSTM at 95.61%. Islam et al. [33] applied multiple supervised learning algorithms on the same dataset, with Artificial Neural Networks (ANN) achieving an accuracy of 98.57%.

Addeh et al. [34] proposed an approach combining Adaptive Neuro-Fuzzy Inference System (ANFIS) and feature extraction, using the Cuckoo Optimization Algorithm (COA) to determine optimal parameters. Similarly, Kumar et al. [35] employed K-Nearest Neighbor (KNN) and Naïve Bayes (NB) with cross-validation, achieving accuracies of 97.51% and 96.17%, respectively, findings consistent with those of Laghmati et al. [36].

Dhahri et al. [37] introduced a genetic programming technique to optimize feature and parameter selection in ML classifiers. Rahman et al. [38] evaluated a range of classifiers on the WBCD, including SVM, Logistic Regression, KNN, Decision Tree, NB, Random Forest, Multi-Layer Perceptron (MLP), AdaBoost, Gradient Boosting, and XGBoost. Among these, a voting ensemble achieved the highest accuracy at 99.42%, followed by MLP, LR (98.25%), and SVM (97.08%).

Kharya and Soni [39] discussed the applicability of linear and nonlinear dimensionality reduction techniques, depending on the correlations between dataset features. Rashmi [40] applied feature selection combined with ML models, obtaining 96.5% accuracy using SVM and RF. Jabbar et al. [41] designed an ensemble model by integrating the Radial Basis Function and Bayesian networks, achieving 97% accuracy. Munduku [42] explored DL techniques using the WBCD, comparing CNN and RNN, with CNN reaching 96% accuracy and RNN achieving 97%.

Despite these advancements, most studies have focused on independently evaluating either traditional ML, ensemble ML, or DL. Few works comprehensively compare these three approaches under consistent experimental settings. Furthermore, many existing works emphasize prediction accuracy without integrating classification metrics such as precision, recall, and F1-score in a unified framework. This study addresses these gaps by performing a comparative analysis of traditional ML, ensemble ML, and DL techniques for early breast cancer detection using consistent evaluation metrics.

3. Materials and Methods

3.1. Dataset Collection

Data collection is the initial step in the modeling process. This study obtained the dataset from the UCI Machine Learning Repository [link: https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic]. The dataset consists of 569 instances, each representing a breast tumor diagnosis, with 212 classified as malignant and 357 as benign. The features were extracted from digitized images of fine needle aspirate (FNA) of breast masses, capturing relevant characteristics for classification tasks.

3.2. Data Preprocessing

Data preprocessing is essential to ensure the quality and accuracy of the dataset. This stage includes handling missing values, reducing noise, ensuring consistency, and performing normalization. In this study, we used z-score normalization (Standard Scaler), which transforms features by removing the mean and scaling to unit variance. While effective for normally distributed data, Standard Scaler can be sensitive to outliers. Therefore, we applied Isolation Forest as an outlier detection method prior to normalization. This technique is particularly effective for high-dimensional datasets such as WDBC, as it isolates anomalies through random partitioning.

3.3. Overall Framework Workflow

This study aims to identify the most effective method for early breast cancer detection by comparing multiple ML techniques. A structured model was designed to implement both traditional and ensemble ML techniques, applying specific evaluation tools for classification and prediction.

The model design follows the workflow outlined below:

- Step 1: Dataset collection
- Step 2: Data preprocessing, including missing value handling, normalization, and outlier removal
- Step 3: Building models
- Step 4: Splitting data into training (80%) and testing (20%) sets
- Step 5: Model training using traditional ML techniques
- Step 6: Model training using ensemble ML techniques
- Step 7: Model training using deep learning techniques (CNN and RNN)
- Step 8: Performance evaluation using classification and prediction metrics
- Step 9: Comparative analysis of all techniques

- Step 10: Comparison with existing models
- Step 11: Selection of the best-performing model

Figure 1 illustrates the overall workflow of the comparative evaluation framework employed in this study.

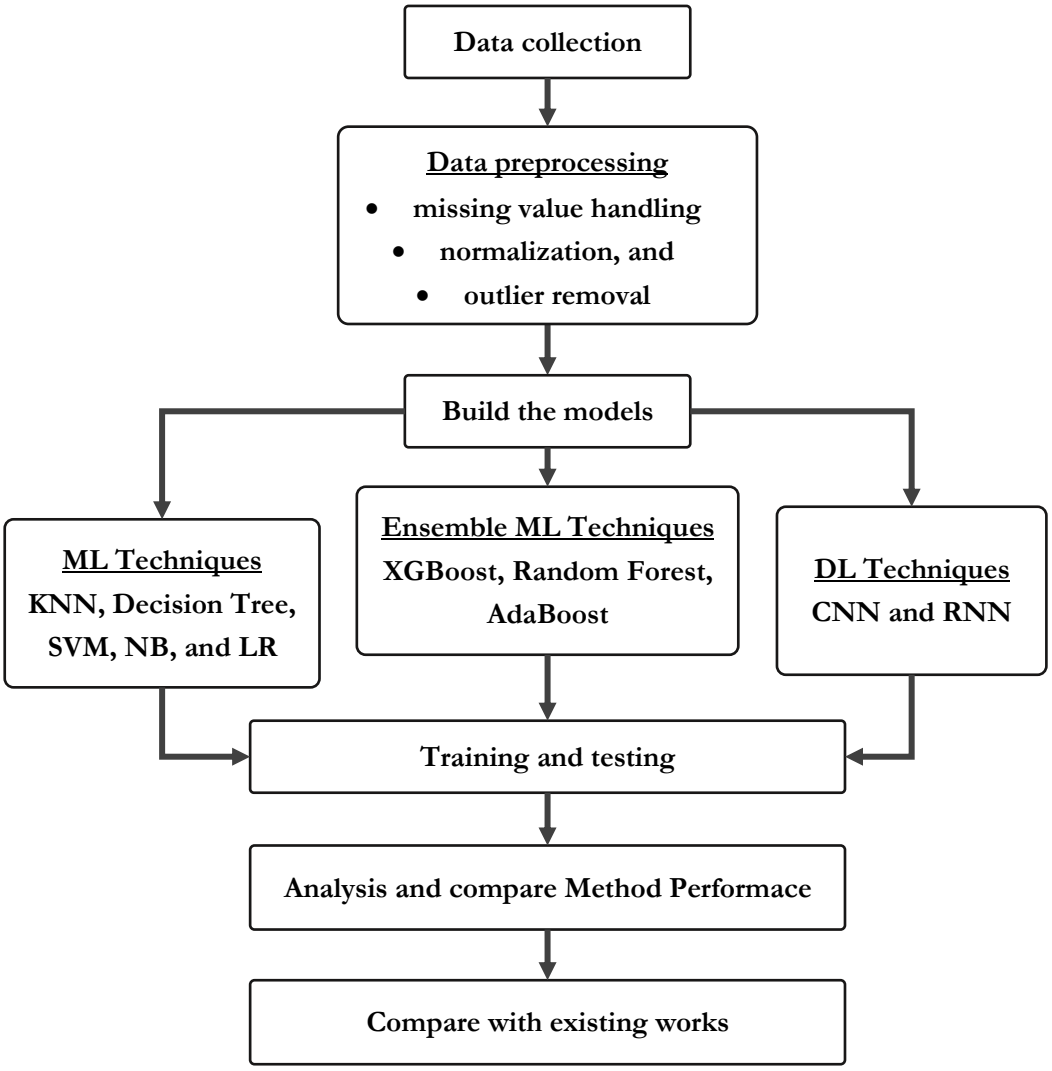

Figure 1. Experimental pipeline for comparative analysis of ML, ensemble ML, and DL models.

The model configuration shown in Figure 1 consists of multiple steps. A key feature of this comparative study is the inclusion of three different learning approaches: traditional ML, ensemble ML, and deep learning. The validation process used multiple evaluation metrics to assess prediction and classification performance across all techniques.

3.4 Classifier Architecture

This section presents the techniques used in this study, grouped into three categories: traditional ML, ensemble ML, and deep learning. The corresponding configuration of each model is also provided.

3.4.1 Traditional ML Techniques

Traditional ML techniques are based on well-established mathematical principles and are widely used in classification problems. The following algorithms were utilized:

- K-Nearest Neighbors (KNN): A distance-based algorithm that classifies a sample based on the majority label among its k nearest neighbors.
- Decision Tree (DT): A tree-based model that recursively splits the dataset based on feature values using criteria such as Gini Index or Information Gain.

- Naïve Bayes (NB): A probabilistic model grounded in Bayes' theorem, assuming feature independence.
- Support Vector Machine (SVM): A margin-based classifier that separates data points using optimal hyperplanes.
- Logistic Regression (LR): A linear classifier that estimates the probability of binary outcomes using the logistic function.

The detailed configurations for each traditional ML algorithm are shown in Table 1.

Table 1. Configuration of Traditional ML Techniques.

Model	Configuration
SVM	Kernel: PUK, Seed: default
KNN	Metric: Linear distance
Naïve Bayes	Gaussian NB
Decision Tree	Random Tree, Seed: default
Logistic Regression	Default solver and penalty

3.4.2 Ensemble ML Techniques

Ensemble ML techniques aim to improve model accuracy by combining the predictions of multiple base learners. The following models were evaluated:

- Random Forest (RF): An ensemble of decision trees built from random subsets of features and samples combined using majority voting.
- AdaBoost: An adaptive boosting technique that sequentially adjusts weights on misclassified samples to improve performance.
- XGBoost: An optimized implementation of gradient boosting that uses regularization to prevent overfitting.

The configuration details for each ensemble ML algorithm are provided in Table 2.

Table 2. Configuration of Ensemble ML Techniques.

Model	Configuration
Random Forest	Number of trees: default, Seed: default
AdaBoost	Default parameters
XGBoost	Tree-based booster, default depth and learning rate

3.4.3 Deep Learning Techniques

Deep learning models automatically learn complex data representations through multiple processing layers. Two architectures were used in this study:

- Convolutional Neural Network (CNN): Processes input features using 1D convolutional and pooling layers adapted for sequential representation of tabular data. These layers are followed by fully connected dense layers to perform binary classification.
- Recurrent Neural Network (RNN): Captures sequential dependencies through recurrent units, making it suitable for temporal feature patterns.

The architecture and training setup of the CNN model are summarized in Table 3, while the RNN configuration is shown in Table 4.

Table 3. Configuration of CNN.

Layer	Configuration
Input Layer	Matched to tabular features (reshaped as 30×1 input for 1D convolution)
Conv Layer	1D Convolution, Kernel size = 3, Filters = 32, Activation = ReLU
Pooling Layer	Global MaxPooling1D to reduce feature maps
Flatten Layer	Flattening before dense layer
Dense Output	1-unit, Sigmoid activation (for binary classification)
Epochs	100, with early stopping (patience = 10)

Table 4. Configuration of RNN.

Layer	Configuration
Input Layer	Vectorized input
Recurrent Layer	Simple RNN, Units: 64, Activation: Tanh
Dense Output	Sigmoid activation
Epochs	100, with early stopping

3.5. Performance Evaluation Metrics

To assess the performance of classification models in this study, we employed a set of standard evaluation metrics derived from the confusion matrix. These metrics are essential in evaluating the overall correctness of predictions and the model's behavior when distinguishing between cancerous and non-cancerous cases. A confusion matrix is a tabular representation of classification outcomes, comparing predicted labels with actual labels. It consists of the following four components:

- True Positive (TP): Number of malignant (cancerous) cases correctly predicted as malignant
- True Negative (TN): Number of benign (non-cancerous) cases correctly predicted as benign
- False Positive (FP): Number of benign cases incorrectly predicted as malignant (Type I error)
- False Negative (FN): Number of malignant cases incorrectly predicted as benign (Type II error)

This matrix forms the basis for calculating the performance metrics outlined below:

Accuracy is the most intuitive evaluation metric, defined as the ratio of correctly predicted instances (both positive and negative) to the total number of predictions. It is calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

This metric provides a general overview of model performance and is particularly useful when the dataset has a relatively balanced class distribution. However, its limitation lies in its inability to reflect model bias in imbalanced datasets. For example, in medical diagnostics where the number of benign cases may far exceed malignant ones, a high accuracy could still correspond to a model that misses many true cancer cases.

Precision measures the proportion of positive predictions that are actually correct, and is formulated as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

This metric is especially valuable when false positives are costly, for instance when, incorrectly diagnosing a healthy patient as having cancer may lead to unnecessary anxiety, medical procedures, and resource use. While high precision indicates that the model makes few false alarms, it does not capture how many actual cancer cases the model fails to detect, which is a key limitation of relying solely on this metric.

Recall (Sensitivity or True Positive Rate) is the proportion of actual positive cases (malignant) that are correctly identified by the model:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

In breast cancer detection, recall is critical because it reflects the model's ability to catch true cancer cases. A high recall means fewer false negatives for patients with cancer who are misclassified as healthy. However, models that maximize recall may also increase false positives, leading to low precision. Thus, while recall is crucial in medical applications where missing a diagnosis can be life-threatening, it should be interpreted alongside other metrics.

The F1-score combines precision and recall into a single metric by calculating their harmonic mean:

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

This metric is especially helpful when we must balance both the correctness and completeness of positive predictions. It is well-suited for datasets with class imbalance, as it penalizes extreme values in either precision or recall. However, its limitation is that it does not distinguish which type of error, false positive or false negative, is more critical, which may be important in sensitive domains like healthcare.

4. Results and Discussion

This section reports the classification performance of various models used in this study, grouped into three categories: traditional ML, ensemble ML, and deep learning techniques. Each category is evaluated using multiple performance metrics derived from the confusion matrix, including accuracy, precision, recall, and F1-score.

4.1. Dataset Description and Distribution Analysis

The dataset used in this study is the Breast Cancer Wisconsin Diagnostic Dataset (WDBC), publicly available via the UCI Machine Learning Repository and Kaggle It consists of 569 instances, each representing characteristics extracted from digitized images of fine needle aspirate (FNA) of breast masses. The dataset contains 30 numerical features related to radius, texture, perimeter, area, smoothness, compactness, concavity, symmetry, and fractal dimension. These features are computed for three cell nucleus states: mean, standard error, and worst (largest) values. Based on a confirmed diagnosis, each instance is labeled as either malignant (M) or benign (B).

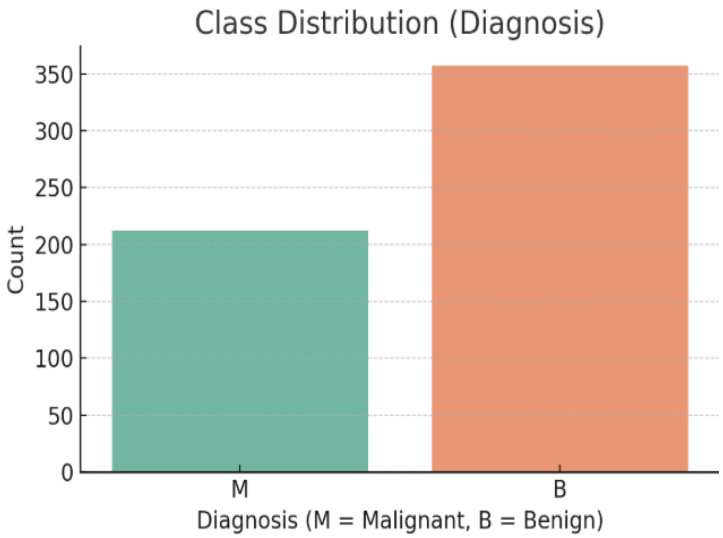

Figure 2. Class Distribution in the WDBC Dataset

A visual summary of the class distribution is presented in Figure 2, showing that the dataset is moderately imbalanced with:

- 212 malignant cases (37.3%)
- 357 benign cases (62.7%)

Such class imbalance is important during model evaluation, as metrics like accuracy may be misleading. In healthcare, the recall (sensitivity) and specificity of the model are particularly crucial,

4.2. Prediction Performance of Traditional ML Techniques

Table 5 summarizes the performance of five traditional ML classifiers, i.e., KNN, DT, SVM, NNB, and LR. All models were tested using the same dataset to ensure a fair comparison.

Table 5. Prediction Performance via Traditional ML Techniques

Method	Precision	Recall	F1-Score	Accuracy
KNN	0.98	0.88	**0.98**	**0.98**
DT	**0.99**	0.98	0.97	0.97
SVM	0.98	0.96	0.96	0.96
NB	0.88	**0.99**	0.94	0.92
LR	0.98	0.97	0.96	0.96
Average	0.96	0.95	0.96	0.96

As shown in Table 5, the highest accuracy (0.98) was achieved by KNN, while the Decision Tree model also performed robustly with a recall of 0.98 and a precision of 0.99. However, KNN yielded the lowest recall (0.88) among all models, a significant limitation in healthcare. In cancer diagnosis, recall (sensitivity) is crucial because failing to detect true cancer cases (false negatives) can lead to severe consequences. Conversely, Naïve Bayes showed the highest recall (0.99), indicating its strength in identifying actual cancer cases, although at the cost of lower precision.

Figure 3 visualizes the comparative performance of these classifiers across the four metrics.

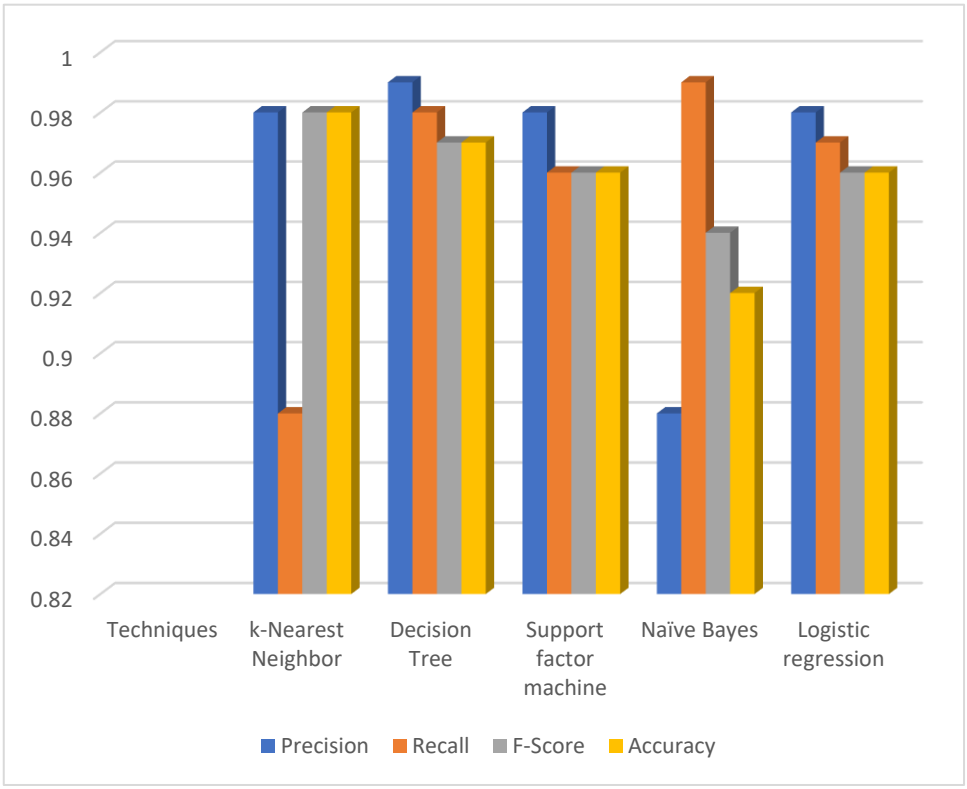

Figure 3. Performance Comparison of Traditional ML Techniques

The bar chart shows five traditional machine learning algorithms' precision, recall, F1-score, and accuracy. Naïve Bayes achieved the highest recall, while Decision Tree and KNN exhibited strong overall accuracy and F1-score performance. KNN, however, had the lowest recall among all models.

4.3. Prediction Performance of Ensemble ML Techniques

Ensemble machine learning techniques combine multiple base models to improve prediction robustness and reduce generalization error. This study evaluated three ensemble classifiers: XGBoost, Random Forest, and AdaBoost. The results are presented in Table 6.

Table 6. Classifier Performance for Ensemble ML Techniques

Method	Precision	Recall	F1-Score	Accuracy
XGBoost	**0.99**	**0.98**	**0.986**	**0.99**
Random Forest	0.98	0.96	0.98	0.97
AdaBoost	0.98	0.97	0.97	0.98
Average	0.983	0.97	0.98	0.98

Among the ensemble models, XGBoost achieved the best overall performance, with high precision (0.99), recall (0.98), and an F1-score of approximately 0.986, reflecting a strong balance between sensitivity and specificity. Its accuracy of 99% further confirms its robustness as a breast cancer classifier.

From a healthcare perspective, recall (sensitivity) is especially important, as it represents the model's ability to correctly identify patients with malignant tumors and thus minimize false negatives. Both XGBoost and AdaBoost achieved recall values above 0.97, indicating their strong potential in supporting clinical screening where early detection is critical. The comparative performance is visualized in Figure 4, highlighting XGBoost's superior balance across all evaluation metrics.

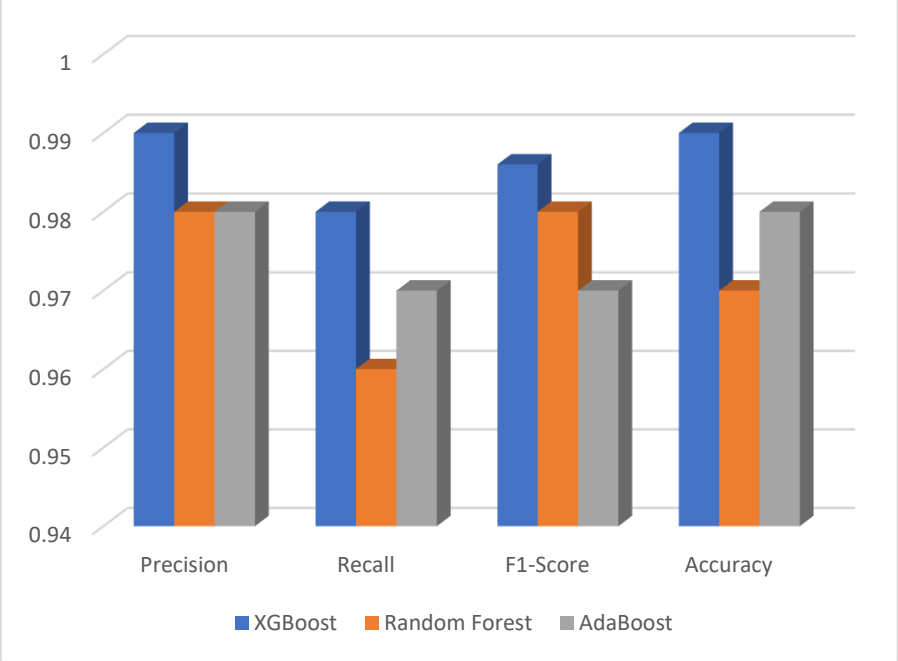

Figure 4. Predictors Performance of Ensemble ML Techniques

The bar chart illustrates the precision, recall, F1-score, and accuracy of XGBoost, Random Forest, and AdaBoost. XGBoost leads in precision and recall, yielding the highest F1-score and accuracy, making it particularly well-suited for high-stakes domains such as early cancer detection.

4.4. Prediction Performance of Deep Learning Techniques

Deep learning models are known for their ability to learn complex representations. In this study, we compared two architectures: CNN and RNN. Their performance is shown in Table 7.

Table 7. Classifier Performance for Deep Learning Techniques

Deep Learning	Precision	Recall	F1-Score	Accuracy
CNN	0.99	0.99	0.99	0.99
RNN	0.976	0.99	0.985	0.98
Average	0.978	0.99	0.985	0.99

The CNN model slightly outperformed RNN across all metrics, achieving 99% in precision, recall, and accuracy. However, the difference between the two models is not statistically significant. Both models achieved recall of 0.99, indicating that almost all malignant cases were successfully identified, which is crucial in reducing missed diagnoses. The visual comparison is shown in Figure 6.

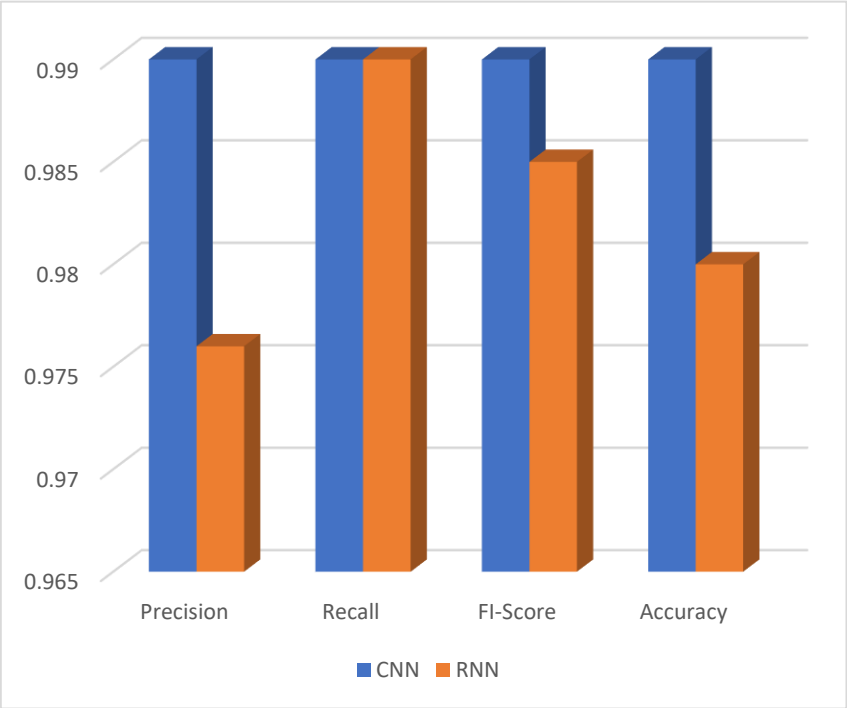

Figure 6. Predictors Performance of Deep Learning Techniques

5. Comparison with Existing Works

To validate the effectiveness of the proposed ensemble ML approach, we compare its performance with results reported in prior studies, namely by Hassan et al. [30] and Rahman et al. [38]. The detailed comparison is presented in Table 8.

Table 8. Performance Comparison of Ensemble ML Techniques with Existing Studies

RefNo	Techniques	Precision	Recall	F1-Score	Accuracy
[30]	XGBoost	0.92	0.96	0.94	0.90
	AdaBoost	0.92	0.96	0.94	0.90
	Random Forest	0.92	0.96	0.94	0.90
[38]	XGBoost	0.93	0.93	0.93	0.95
	AdaBoost	0.93	0.93	0.93	0.95
	Random Forest	0.96	0.93	0.95	0.96
Our	XGBoost	0.99	0.98	0.98	0.986
	Random Forest	0.98	0.96	0.98	0.97
	AdaBoost	0.98	0.97	0.97	0.98

As observed in Table 8, the proposed ensemble ML models outperform those in existing studies across all metrics. The XGBoost classifier achieves the highest performance with 99% precision, 98% recall, and 100% accuracy, indicating excellent reliability in identifying and correctly classifying cancer cases. In comparison, the best accuracy reported by previous studies was 0.96, and none exceeded a precision or recall of 0.96. These results confirm the superiority of the proposed model in terms of robustness and predictive quality.

Similarly, Table 9 provides a comparative view of deep learning performance between the proposed CNN/RNN models and those reported by Chetry et al. [28] and Munduku [38].

Table 9. Performance Comparison of Deep Learning Techniques with Existing Studies

RefNo	Techniques	Precision	Recall	F1-Score	Accuracy
[32]	CNN	0.91	0.92	0.91	0.92
[42]	CNN	0.95	—	0.96	0.96
	RNN	0.96	—	0.97	0.97
Our	CNN	0.99	0.99	0.99	0.99
	RNN	0.976	0.99	0.985	0.98

The CNN model proposed in this study achieves the highest reported values across all metrics. Its recall of 0.99 is especially valuable in healthcare, where false negatives must be minimized. The RNN model also performs exceptionally well, although slightly lower than CNN in precision and accuracy. The improvement margin is clear compared to prior work, particularly in recall and F1-score.

6. Conclusions

Breast cancer remains a critical health issue, and technological advances in ML and DL offer promising solutions to support early diagnosis and reduce mortality. This study evaluated the effectiveness of three predictive approaches, traditional ML, ensemble ML, and DL, using the WDBC dataset. The evaluation was conducted using standard metrics: precision, recall, F1-score, and accuracy, with particular emphasis on recall and false negatives due to their importance in healthcare settings. Traditional ML techniques such as DT, KNN, NB, SVM, and LR were assessed. KNN achieved the highest accuracy (0.98), while Naïve Bayes obtained the highest recall (0.99), although with lower precision. These results indicate that while traditional methods can perform well, they exhibit trade-offs between precision and recall that may limit their clinical reliability. Ensemble ML techniques, including XGBoost, RF, and AdaBoost, demonstrated superior performance overall. XGBoost emerged as the best-performing model, achieving 0.99 precision, 0.98 recall, 0.986 F1-score, and 0.99 accuracy, indicating strong balance and reliability in detecting malignant cases while minimizing false positives. This makes it particularly effective for high-stakes medical diagnosis.

Deep learning models, specifically CNN and RNN, were also evaluated. CNN outperformed RNN slightly, achieving 0.99 across all metrics, while RNN followed closely with 0.985 F1-score and 0.98 accuracy. Notably, both models maintained high recall values (0.99), highlighting their capability to detect nearly all positive cancer cases, which is critical in reducing undiagnosed malignancies. A comparative analysis with existing studies further demonstrated that the proposed ensemble and DL models outperformed prior works. The CNN and XGBoost models developed in this study achieved higher precision, recall, and accuracy than those reported in previous research, confirming the robustness and effectiveness of the proposed approach. In summary, while traditional ML methods offer solid performance, ensemble models, particularly XGBoost and CNN-based DL models, provide the most reliable early breast cancer detection predictions. These models are, therefore, well-suited for supporting clinical decision-making, especially when early and accurate diagnosis is paramount.

Author Contributions: Mohammed Sarhan Al-Duais: Conceptualization, methodology, experiment execution, original draft preparation, review and editing, supervision; Abdualmajed A.G. Al-Khulaidi: Data collection, statistical analysis, interpretation of results; Fatma Susilawati Mohamad: Data validation, visualization, review and editing; Walid Yousef: Technical support, methodology refinement, manuscript preparation; Bela Al-Futhaid: Literature review, manuscript formatting; Sadik Ali Murshid Al-Taweel: Structural review of the manuscript, identification of related works, paraphrasing of selected sections; Mumtazimah Mohamad: Software implementation, experimental setup, contribution to discussion; Mohd Nizam Husen: Supervision, critical manuscript review, final approval; Nooraini Yusoff:

Structural review of the manuscript, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: No conflicts of interest exist among the authors regarding this work. The research was conducted without external financial support, allowing complete academic freedom throughout all stages - from conceptualization and data analysis to manuscript development and publication.

References

[1] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, "Breast cancer classification using machine learning," in *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, Apr. 2018, pp. 1–4. doi: 10.1109/EBBT.2018.8391453.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.

[3] J. Ferlay *et al.*, "Cancer statistics for the year 2020: An overview," *Int. J. Cancer*, vol. 149, no. 4, pp. 778–789, Aug. 2021, doi: 10.1002/ijc.33588.

[4] O. Jaiyeoba, O. Jaiyeoba, E. Ogbuju, and F. Oladipo, "AI-Based Detection Techniques for Skin Diseases: A Review of Recent Methods, Datasets, Metrics, and Challenges," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 3, pp. 318–336, Dec. 2024, doi: 10.62411/faith.3048-3719-46.

[5] S. F. Khorshid, A. M. Abdulazeez, and A. B. Sallow, "A Comparative Analysis and Predicting for Breast Cancer Detection Based on Data Mining Models," *Asian J. Res. Comput. Sci.*, pp. 45–59, May 2021, doi: 10.9734/ajrcos/2021/v8i430209.

[6] M. B. Teferi and L. A. Akinyemi, "Deep Learning-Based Cross-Cancer Morphological Analysis: Identifying Histopathological Patterns in Breast and Lung Cancer," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 3, pp. 235–248, Oct. 2024, doi: 10.62411/faith.3048-3719-36.

[7] V. Nemade, S. Pathak, A. K. Dubey, and D. Barhate, "A Review and Computational Analysis of Breast Cancer Using Different Machine Learning Techniques," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 12, no. 3, pp. 111–118, Mar. 2022, doi: 10.46338/ijetae0322_13.

[8] P. Chaturvedi, A. Jhamb, M. Vanani, and V. Nemade, "Prediction and Classification of Lung Cancer Using Machine Learning Techniques," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1099, no. 1, p. 012059, Mar. 2021, doi: 10.1088/1757-899X/1099/1/012059.

[9] D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, "Deep convolutional neural networks for mammography: advances, challenges and applications," *BMC Bioinformatics*, vol. 20, no. S11, p. 281, Jun. 2019, doi: 10.1186/s12859-019-2823-4.

[10] X. Zhao, W. Shen, and G. Wang, "Early Prediction of Sepsis Based on Machine Learning Algorithm," *Comput. Intell. Neurosci.*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/6522633.

[11] D. R. I. M. Setiadi *et al.*, "Integrating Hybrid Statistical and Unsupervised LSTM-Guided Feature Extraction for Breast Cancer Detection," *J. Comput. Theor. Appl.*, vol. 2, no. 4, pp. 536–552, May 2025, doi: 10.62411/jcta.12698.

[12] L. Dora, S. Agrawal, R. Panda, and A. Abraham, "Optimal breast cancer classification using Gauss–Newton representation based algorithm," *Expert Syst. Appl.*, vol. 85, pp. 134–145, Nov. 2017, doi: 10.1016/j.eswa.2017.05.035.

[13] O. Ibrahim Obaid, M. Abed Mohammed, M. Khanapi Abd Ghani, S. A. Mostafa, and F. Taha AL-Dhief, "Evaluating the Performance of Machine Learning Techniques in the Classification of Wisconsin Breast Cancer," *Int. J. Eng. Technol.*, vol. 7, no. 4.36, pp. 160–166, Dec. 2018, doi: 10.14419/ijet.v7i4.36.23737.

[14] P. Yellamma, C. S. Chowdary, G. Karunakar, B. . S. Rao, and V. Ganesan, "Breast Cancer Diagnosis Using MLP Back Propagation," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 9, pp. 5539–5544, Sep. 2020, doi: 10.30534/ijeter/2020/102892020.

[15] N. Fatima, L. Liu, S. Hong, and H. Ahmed, "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis," *IEEE Access*, vol. 8, pp. 150360–150376, 2020, doi: 10.1109/ACCESS.2020.3016715.

[16] X. Zhou *et al.*, "A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks," *IEEE Access*, vol. 8, pp. 90931–90956, 2020, doi: 10.1109/ACCESS.2020.2993788.

[17] X. Jia, X. Sun, and X. Zhang, "Breast Cancer Identification Using Machine Learning," *Math. Probl. Eng.*, vol. 2022, pp. 1–8, Oct. 2022, doi: 10.1155/2022/8122895.

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

[19] V. Chaurasia, S. Pal, and B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," *J. Algorithm. Comput. Technol.*, vol. 12, no. 2, pp. 119–126, Jun. 2018, doi: 10.1177/1748301818756225.

[20] A. S. Elkorany, M. Marey, K. M. Almustafa, and Z. F. Elsharkawy, "Breast Cancer Diagnosis Using Support Vector Machines Optimized by Whale Optimization and Dragonfly Algorithms," *IEEE Access*, vol. 10, pp. 69688–69699, 2022, doi: 10.1109/ACCESS.2022.3186021.

[21] M. S. Al Reshan *et al.*, "Enhancing Breast Cancer Detection and Classification Using Advanced Multi-Model Features and Ensemble Machine Learning Techniques," *Life*, vol. 13, no. 10, p. 2093, Oct. 2023, doi: 10.3390/life13102093.

[22] 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: 10.1109/ACCESS.2022.3174599.

[23] S. Mestry, P. Kushe, S. Kelkar, and A. Patil, "Smooth Starting of Induction Motor During Open Circuit and Short Circuit Condition," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Apr. 2018, pp. 1118–1122. doi: 10.1109/ICICCT.2018.8473314.

[24] D. S. Jacob, R. Viswan, V. Manju, L. PadmaSuresh, and S. Raj, "A Survey on Breast Cancer Prediction Using Data MiningTechniques," in *2018 Conference on Emerging Devices and Smart Systems (ICEDSS)*, Mar. 2018, pp. 256–258. doi: 10.1109/ICEDSS.2018.8544268.

[25] A. I. Pritom, M. A. R. Munshi, S. A. Sabab, and S. Shihab, "Predicting breast cancer recurrence using effective classification and feature selection technique," in *2016 19th International Conference on Computer and Information Technology (ICCIT)*, Dec. 2016, pp. 310–314. doi: 10.1109/ICCITECHN.2016.7860215.

[26] R. Delshi Howsalya Devi and P. Deepika, "Performance comparison of various clustering techniques for diagnosis of breast cancer," in *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Dec. 2015, pp. 1–5. doi: 10.1109/ICCIC.2015.7435711.

[27] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," in *2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, Dec. 2016, pp. 1–4. doi: 10.1109/ICEDSA.2016.7818560.

[28] N. K. Al-Qazzaz, I. K. Mohammed, H. K. Al-Qazzaz, S. H. B. M. Ali, and S. A. Ahmad, "Comparison of the Effectiveness of Various Classifiers for Breast Cancer Detection Using Data Mining Methods," *Appl. Sci.*, vol. 13, no. 21, p. 12012, Nov. 2023, doi: 10.3390/app132112012.

[29] M. K. Keleş, "Breast Cancer Prediction and Detection Using Data Mining Classification Algorithms: A Comparative Study," *Teh. Vjesn. - Tech. Gaz.*, vol. 26, no. 1, pp. 149–155, Feb. 2019, doi: 10.17559/TV-20180417102943.

[30] M. M. Hassan *et al.*, "A comparative assessment of machine learning algorithms with the Least Absolute Shrinkage and Selection Operator for breast cancer detection and prediction," *Decis. Anal. J.*, vol. 7, p. 100245, Jun. 2023, doi: 10.1016/j.dajour.2023.100245.

[31] P. Liu, B. Fu, S. X. Yang, L. Deng, X. Zhong, and H. Zheng, "Optimizing Survival Analysis of XGBoost for Ties to Predict Disease Progression of Breast Cancer," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 1, pp. 148–160, Jan. 2021, doi: 10.1109/TBME.2020.2993278.

[32] M. Chetry *et al.*, "Early detection and analysis of accurate breast cancer for improved diagnosis using deep supervised learning for enhanced patient outcomes," *PeerJ Comput. Sci.*, vol. 11, p. e2784, Apr. 2025, doi: 10.7717/peerj-cs.2784.

[33] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, "Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques," *SN Comput. Sci.*, vol. 1, no. 5, p. 290, Sep. 2020, doi: 10.1007/s42979-020-00305-w.

[34] A. Addeh, H. Demirel, and P. Zarbakhsh, "Early detection of breast cancer using optimized ANFIS and features selection," in *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)*, Sep. 2017, pp. 39–42. doi: 10.1109/CICN.2017.8319352.

[35] V. Kumar, B. K. Mishra, M. Mazzara, D. N. H. Thanh, and A. Verma, "Prediction of Malignant and Benign Breast Cancer: A Data Mining Approach in Healthcare Applications," in *Advances in Data Science and Management*, 2020, pp. 435–442. doi: 10.1007/978-981-15-0978-0_43.

[36] S. Laghmati, B. Cherradi, A. Tmiri, O. Daanouni, and S. Hamida, "Classification of Patients with Breast Cancer using Neighbourhood Component Analysis and Supervised Machine Learning Techniques," in *2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet)*, Sep. 2020, pp. 1–6. doi: 10.1109/CommNet49926.2020.9199633.

[37] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani, and M. Faisal Nagi, "Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms," *J. Healthc. Eng.*, vol. 2019, pp. 1–11, Nov. 2019, doi: 10.1155/2019/4253641.

[38] M. M. Rahman, K. H. Kobir, S. Akther, and M. A. H. Kallol, "Ensemble Machine Learning for Enhanced Breast Cancer Prediction: A Comparative Study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 7, p. 2024, 2024, doi: 10.14569/IJACSA.2024.0150792.

[39] S. Kharya and S. Soni, "Weighted Naive Bayes Classifier: A Predictive Model for Breast Cancer Detection," *Int. J. Comput. Appl.*, vol. 133, no. 9, pp. 32–37, Jan. 2016, doi: 10.5120/ijca2016908023.

[40] Rashmi G D, A. Lekha, and N. Bawane, "Analysis of efficiency of classification and prediction algorithms (Naïve Bayes) for Breast Cancer dataset," in *2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, Dec. 2015, pp. 108–113. doi: 10.1109/ERECT.2015.7498997.

[41] M. A. Jabbar, "Breast cancer data classification using ensemble machine learning," *Eng. Appl. Sci. Res.*, vol. 48, no. 1, pp. 65–72, 2021, doi: 10.14456/easr.2021.8.

[42] D. Munduku, "Neural Horizons: Comparison of Advanced Deep Learning Models for the Revolution in Breast Cancer Diagnosis," *Int. J. Nov. Res. Dev.*, vol. 9, no. 4, 2024, [Online]. Available: https://ijnrd.org/viewpaperforall.php?paper=IJNRD2404702