Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 4, October 2025, pp: 1355-1366 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Breast Cancer Classification Using z-score Thresholding and Machine Learning

Mustafa E. Yildirim[1,2], Yucel B. Salman[3]

[1] Electronics and Communications Engineering, American University of Malta, Bormla, Malta
[2] Department of Electrical & Electronics Engineering, Bahcesehir University, Istanbul, Turkey
[3] Department of Software Engineering, Bahcesehir University, Istanbul, Turkey

Corresponding author: Mustafa E. Yildirim (e-mail: mustafaeren.yildirim@bau.edu.tr, eren.yildirim@aum.edu.mt),
Author(s) Email: batu.salman@bau.edu.tr

Abstract. Image processing and machine learning are being used in biomedical applications as supporting tools for the detection and diagnosis of certain diseases. Breast cancer is one of these diseases that researchers have devoted great effort to for decades. To accomplish this task, image-based and feature-based public datasets are available for use. Due to several factors such as hardware limitations or preprocessing, images can become noisy. The noise in images, which can lead to anomalies or outliers in the dataset, may decrease detection accuracy and mislead medical staff during the diagnostic stage. Therefore, this study aims to present the effect of removing outliers from the dataset on the detection accuracy of breast cancer. The proposed method removes outliers detected through z-score analysis. The remaining data are normalized, and the classification accuracies of ten methods are obtained through direct implementation. The methods include XGBoost, Neural Network, CNN, RNN, AdaBoost, LSTM, GRU, Random Forest, SVM, and Logistic Regression. The public dataset Wisconsin Diagnostic Breast Cancer (WDBC) was used in this study. An ablation study was conducted by fine-tuning the threshold value of the z-score method. The results showed that the best accuracy was obtained when the threshold value was set to 3. Additionally, a comparison was made between the results obtained using the entire dataset and the dataset after outlier removal. The results showed that the average accuracy of all classifiers was 98.08%. In conclusion, the findings indicate that removing outliers from the dataset increases the overall accuracy of breast cancer detection.

Keywords: Breast cancer; Outlier detection; z-score; WDBC

I. Introduction

Breast cancer continues to be one of the most prevalent and life-threatening cancers among women worldwide. According to the World Health Organization (WHO), breast cancer accounts for approximately 25% of all cancer diagnoses in women, with over 2.3 million new cases reported annually as of 2020. Despite advancements in treatment modalities, early detection and accurate diagnosis remain the most effective strategies for improving survival rates and reducing breast cancer-related mortality. Early identification of malignant tumors allows for timely intervention, significantly improving the prognosis and quality of life for patients [1], [2].

In recent years, the integration of Artificial Intelligence (AI) into healthcare has revolutionized diagnostic methods, enabling automated, efficient, and highly accurate solutions for disease classification. Among AI-driven technologies, Machine Learning (ML) and Deep Learning (DL) have emerged as transformative tools in the medical domain. These approaches leverage data-driven algorithms to identify complex patterns and relationships in medical data, providing actionable insights that aid clinicians in decision-making. Their ability to process large datasets and extract meaningful information has made them particularly valuable in breast cancer diagnosis, where accurate classification of tumors as benign or malignant is critical [3], [4].

The Wisconsin Diagnostic Breast Cancer (WDBC) dataset has become a benchmark dataset in breast cancer research and is widely used for developing and evaluating predictive models. This dataset contains detailed measurements of tumor characteristics such as cell radius, texture, perimeter, area, and smoothness, making it an ideal resource for training and testing machine learning algorithms. The structured nature and accessibility of the WDBC have allowed researchers to explore a wide range of classification techniques, from traditional statistical models to advanced deep learning architectures. The primary objective of these studies is to achieve high classification accuracy while ensuring the robustness and generalizability of the models [5], [6].

The Mammographic Mass Dataset (MMD) [7], another dataset frequently used in breast cancer research, contains 961 records with features such as

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

tumor shape, margin, density, and patient age. The dataset is labeled to indicate the severity of breast masses (benign or malignant) and is often used for feature-based classification tasks. Studies have shown that Decision Trees achieve 83.5% accuracy on this dataset, but preprocessing techniques such as feature scaling have improved this performance to 88.2% [8]. Logistic Regression (LR) has also performed well, achieving 87% accuracy with appropriate preprocessing steps [9].

Achieving robustness and generalizability is crucial for the practical application of machine learning models in clinical settings. Robustness refers to a model's ability to perform consistently across different datasets and under varying conditions, while generalizability ensures that the model can handle unseen data effectively. However, several challenges, such as the presence of outliers, imbalanced datasets, and overfitting, can hinder the performance of predictive models. Addressing these challenges is essential to ensure that machine learning models transition successfully from research to real-world clinical practice.

Outlier detection and removal are particularly important preprocessing steps in machine learning pipelines, as outliers can introduce noise, distort model training, and lead to biased predictions. Outliers are data points that deviate significantly from most of the dataset, and their presence can adversely affect the performance of both traditional and deep learning algorithms. Statistical methods, such as the z-score outlier detection method [10], have been widely adopted for detecting and eliminating outliers. By removing these anomalies, researchers can improve dataset quality, reduce the risk of overfitting, and enhance the model's ability to generalize to new data [11], [12].

The application of machine learning techniques to the WDBC dataset has yielded promising results in breast cancer classification. Traditional algorithms, such as Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR), have demonstrated strong performance due to their ability to handle structured data and identify meaningful patterns. For instance, studies have shown that SVM achieves high classification accuracy when combined with appropriate feature selection and preprocessing techniques [13], [14]. Similarly, ensemble learning methods, such as Random Forest and boosting algorithms like XGBoost and AdaBoost, have been employed to further improve classification accuracy by aggregating predictions from multiple weak learners [15], [16].

In addition to traditional models, the advent of deep learning has opened new possibilities for breast cancer diagnosis. DL models, including Convolutional Neural Networks (CNN) [17], Recurrent Neural Networks (RNN) [18], Long Short-Term Memory (LSTM) networks [19], and Gated Recurrent Units (GRU) [20], have demonstrated superior capabilities in capturing complex, non-linear relationships within data. CNN, for example, have been adapted for tabular datasets such as WDBC, leveraging their ability to automatically extract high-level features from raw data [21], [22]. RNN and their variants, on the other hand, are particularly effective in sequential data analysis and have been used to model temporal dependencies in medical datasets [23], [24].

Despite their high accuracy, deep learning models often face challenges such as overfitting and the need for large labeled datasets. Hybrid approaches that combine traditional machine learning algorithms with deep learning frameworks have been proposed to address these limitations. Additionally, preprocessing techniques such as outlier removal, feature scaling, and dimensionality reduction have been shown to significantly enhance model performance by improving data quality and optimizing feature representation [25], [26].

The importance of preprocessing in machine learning cannot be overstated, as it directly impacts the reliability and interpretability of predictive models. Studies have demonstrated that removing outliers and balancing datasets can lead to substantial improvements in classification accuracy and robustness. These preprocessing steps are particularly relevant in medical applications, where the cost of misclassification can be high. Moreover, integrating explainable AI (XAI) techniques into machine learning pipelines has gained traction in recent years, as it provides transparency and interpretability to model predictions. By understanding the features and patterns that drive a model's decisions, clinicians can gain confidence in its recommendations and integrate it into their diagnostic workflows [27], [28]. Despite the large number of studies utilizing machine learning in breast cancer detection, limited attention has been given to the impact of outliers. Few studies have addressed this issue [29], [30], [31], although they show that outliers can decrease model performance. This study addresses this gap in the literature through a systematic investigation into the effect of outlier removal on model accuracy. We aim to highlight the potential for improved performance in ML-based diagnostic systems.

In this study, we evaluate the impact of outlier removal on the performance of various machine learning and deep learning models applied to the WDBC dataset. Using the z-score method for outlier detection, we preprocess the dataset to eliminate anomalies and compare the results with conventional approaches. The goal is to highlight the importance of data preprocessing in improving classification accuracy, robustness, and generalizability. The

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

expected outcome of outlier removal from the dataset is that the feature-wise interclass difference will become more significant, thereby increasing detection accuracy. Our findings provide valuable insights into the role of preprocessing techniques in enhancing the reliability of breast cancer diagnostic models, paving the way for their potential integration into clinical practice.

The remainder of this paper is organized as follows: Section II describes the dataset, preprocessing methods, and machine learning models used in this study. Section III presents the experimental results and compares them with prior studies. Section IV discusses the findings, compares them with the state-of-the-art (SOTA), and outlines the limitations of the study. Section V concludes the paper and provides suggestions for potential future research.

II. Method

This study focuses on improving the accuracy of breast cancer classification through outlier removal and the application of machine learning methods. The proposed method consists of several steps: dataset collection, z-score filtering, normalization, data splitting, classification, and comparison with previous studies. Fig. 1 illustrates the workflow of the model used in this study.

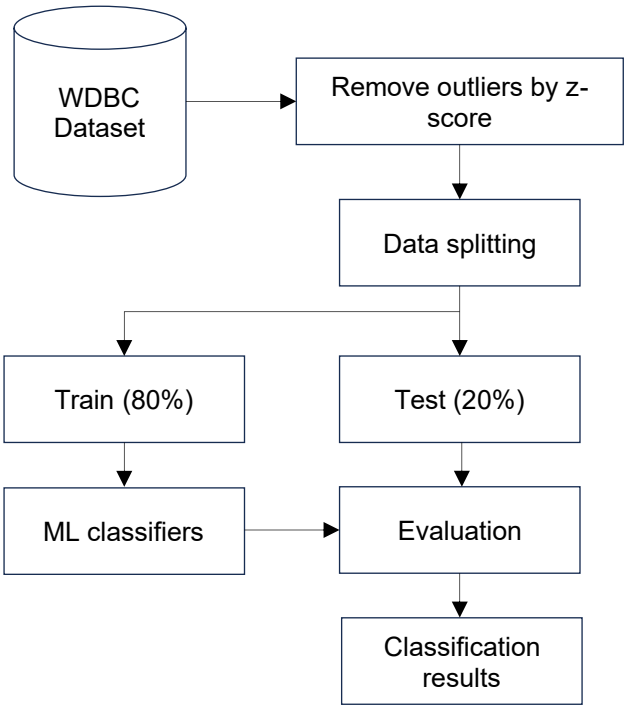

Fig. 1. Framework of the study.

A. Dataset

The study was implemented and tested using the WDBC dataset for breast cancer detection. This dataset was obtained from the University of Wisconsin Hospitals. It consists of features computed from digitized images of fine needle aspirates (FNA) of breast masses. Example images for each class of the WDBC dataset are illustrated in Fig. 2.

The dataset contains 569 unique samples, of which 212 are malignant and 357 are benign. The dataset can be accessed at https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic. Ten distinct features were extracted from each cell nucleus: radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. For every feature, three statistical values were computed—mean, standard error, and worst—resulting in a total of 30 features per sample. There are no missing values in

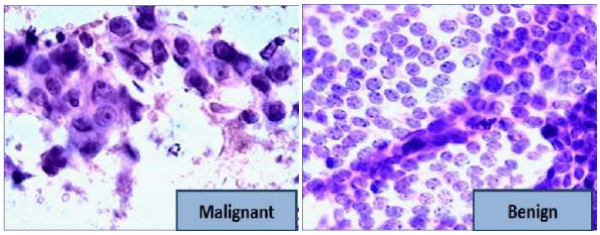

Fig. 2. Sample images from WDBC dataset, (a) Malignant class, (b) Benign class.

the dataset, which makes it reliable for researchers. All features are numerical and have different ranges. Each feature was normalized within its specific range using the StandardScaler module as part of this study.

B. Outlier Detection and Z-score

In any dataset, outliers can significantly affect statistical predictions as well as model parameter estimates. They may distort the distribution of variables in the dataset. These values are located far from the general population of the distribution and can be detected using outlier detection methods. Therefore, outlier detection was applied to the dataset to identify and remove outliers. Several outlier detection methods are discussed in the literature. Isolation Forest (IF) is a widely used method; however, it contains randomness and depends on multiple parameters. Another method, the Interquartile Range (IQR), suffers from dependency on dataset size and tends to underperform with small datasets. Therefore, the z-score method was chosen for this study. It is a statistical measure that indicates how many standard deviations a data point deviates from the mean of the distribution [32], [33]. The calculation for each point x is shown in Eq. (1):

$$z_x = \frac{x-\mu}{\sigma} \quad (1)$$

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

$$\begin{cases} x: outlier\ and\ removed & if\ z_x \geq threshold \\ x: normal\ data\ sample & else \end{cases} \quad (2)$$

where, z_x is the z-score, or the distance of point x from the mean; μ and σ are the mean and the standard deviation of the sample set, respectively. Samples with a z-score greater than a predetermined threshold were labeled as outliers and removed from the dataset. This operation is given in Eq. (2). In the literature, the typical threshold value is ±3 [34], as approximately 99.6% of samples in a normally distributed population fall within ±3 standard deviations. The remaining data were used for training and testing. The dataset was split into training and testing subsets in an 80:20 ratio. Both subsets were normalized using standard scaling before the training stage was initiated. This procedure was repeated for each iteration of the 5-fold cross-validation. Data splitting was performed in a random manner.

C. ML Classifiers

This section explains the ML methods we used in this study for breast cancer classification. The dataset after outlier removal was trained and tested using ten ML methods: XGBoost, NN (Neural Network), CNN, RNN, GRU, LSTM, SVM, RF, and LR. XGBoost, RF, and AdaBoost are tree-based methods. NN, CNN, RNN and LSTM are neural network–based architectures. SVM is a margin-based classifier, while LR is a linear classifier.

1. XGBoost

This method is a tree-based classifier defined by the characteristic equation in Eq. (3) [35], where $l(y_i, \hat{y}_i)$ is the loss function, y_i and $\hat{y}_i$ are the actual and predicted output values for the i^{th} sample, γ is penalty factor, λ is the regularization parameter, T is the number of leaves and w is the leaf weight.

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \gamma T + 0.5\lambda \sum_{j=1}^{T} w_j^2 \quad (3)$$

2. NN

This method is a fully connected and layer-based method with the characteristic equation in Eq. (4) [36], where the final decision, weight matrix, bias vector, activation function of l^{th} layer is a^l, W^l, b^l and σ, respectively.

$$a^l = \sigma\left(W^l a^{(l-1)} + b^l\right) \quad (4)$$

3. CNN

This method uses convolutions, activation functions and pooling steps to extract low- and high-level futures from an input image and generates an output label. Its general formula is as shown in Eq. (5) [37].

$$a_{i,j} = \sigma \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{n=1}^{N} I_c(i+m, j+n)\, K_c(m,n) + b \quad (5)$$

In Eq. (5), $a_{i,j}$ is the output, σ is the activation function, I is the input image with the dimension $m \times n$, K is the kernel function and b is the bias.

4. RNN

This method uses recurrent computation for hidden states during each layer of the network. Its equations are as shown in Eq. (6) and Eq. (7) [38] with h_t and y_t are the hidden states and output, W is the weight matrix, b and c are bias coefficients. The parameters σ and ϕ are activation functions.

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b) \quad (6)$$

$$y_t = \phi\left(W_y h_t + c\right) \quad (7)$$

5. LSTM

This method is a type of RNN that incorporates gates and cell states, as represented by Eq. (8), Eq. (9), Eq. (10), Eq. (11), Eq. (12) and Eq. (13) [38]. In the following equations, σ is the activation function, b is the bias constant, W is the weight matrix, x is input, h is hidden state, f_t is forget gate, o is the output gate and C is the cell state.

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \quad (8)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (9)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (10)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (11)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (12)$$

$$h_t = o_t \odot \tanh C_t \quad (13)$$

6. AdaBoost

This method constructs a strong classifier by combining multiple weak classifiers using adaptive weighting. It is calculated as shown in Eq. (14) [39], x is the input, α is the weight of the weak classifier h.

$$F(x) = sign(\sum_{m=1}^{M} \alpha_m h_m(x)) \quad (14)$$

7. SVM

This method classifies samples by finding the optimal hyperplane that separates data points into distinct classes with minimal classification error. Eq. (15) [40] shows the characteristic equation for SVM, where α is the weight, y is the label, α is the Lagrange multiplier, K is the kernel function and b is the bias.

$$\hat{y} = sign(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b) \quad (15)$$

8. GRU

This method is a simplified version of LSTM with fewer gates and no explicit cell state. It is represented by Eq. (16) for gate update, Eq. (17) gate reset, Eq. (18) activation, and Eq. (19) for calculating the new hidden states [41]. In the following equations, x is the input, h is output, $\tilde{h}$ is the candidate activation vector, z and r are the update gate and reset gate vectors, W and b

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

are the weight matrix and bias vector, σ is the activation function.

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z) \quad (16)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r) \quad (17)$$

$$\tilde{h}_t = \tanh(W_h[r_t \odot h_{t-1}, x_t] + b_h) \quad (18)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (19)$$

9. RF

This is a tree-based method consisting of an ensemble of decision trees. Eq. (20) represents the characteristic equation of this method, where $\hat{y}$ is the final decision taking the majority vote of all sub decision h_t [42].

$$\hat{y} = majority\ vote\{h_t(x) | t = 1, \dots, T\} \quad (20)$$

10.LR

This is a well-known method for binary classification. It calculates the probability of each sample belonging to one of two classes, as shown in Eq. (21) where x is input, b is bias and w is weight matrix. Eq. (22) [43] is the negative log-likelihood function to be minimized with inputs of weight matrix, y_i and $\hat{y}_i$ are the actual and predicted output values.

$$P(y = 1|x) = \frac{1}{1+e^{-(w^T x+b)}} \quad (21)$$

$$\mathcal{L}(w, b) = -\sum_{i=1}^{n}[y_i \log \hat{y}_i + (1 - y_i) \log(1 - y_i)] \quad (22)$$

D. Implementation Details

The proposed model in this study was implemented using Python 3.11 on the Google Colaboratory (Colab) platform. Dataset analysis and evaluation were carried out using the scikit-learn, pandas, xgboost, and numpy libraries along with their submodules. For the CNN, RNN, LSTM, and GRU methods, Keras was utilized. The hyperparameter values for each classifier are presented in Table 1. The hyperparameters for the deep learning–based methods, namely CNN, NN, GRU, RNN, and LSTM, include the number of epochs, batch size, optimizer type, and loss function. No data augmentation or balancing technique was applied to the dataset. The other ML-based methods were used with their default parameter settings in scikit-learn. Since an 80:20 ratio was used for the training and testing sets, the random data split was repeated, and testing was performed five times. The results reported in this study are the average of the five runs.

Table 1. The Hypermeters for the ML methods.

Method	Hyperparameters
XGBoost	label_encoder = False
NN	Epoch = 10, batch size = 32 Optimizer = adam loss = binary_crossentropy
CNN	Epoch = 10, batch size = 32 Optimizer = adam loss = binary_crossentropy
RNN	Epoch = 10, batch size = 32 Optimizer = adam loss = binary_crossentropy
LSTM	Epoch = 10, batch size = 32 Optimizer = adam loss = binary_crossentropy
AdaBoost	n_estimators =100
GRU	Epoch = 10, batch size = 32 Optimizer = adam loss = binary_crossentropy
SVM	Kernel = linear
RF	n_estimators=200, max_depth=20
LR	max_iter = 1000

E. Performance Metric

The performance of the ML methods in this study was measured using accuracy and F1-score, formulated in Eq. (23) and Eq. (24) [44].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (23)$$

$$F1\ score = \frac{TP}{TP+0.5(FP+FN)} \quad (24)$$

Accuracy, defined in Eq. (23), is the ratio of correct predictions to the total number of predictions, where TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative, respectively. The F1-score, on the other hand, is a combined metric that is particularly useful when there is an imbalance between dataset classes.

III. Result

This section presents the implementation details, performance metrics, and classification accuracies on the WDBC dataset. The classification accuracy subsection includes comparisons with previous studies and evaluates the performance of the proposed model using different z-score threshold values.

A. Classification accuracy of ML classifiers according to the z-score threshold

As mentioned in part B of Section II, the z-score filtering method identifies a data point as an outlier if it is located farther from the dataset mean than a defined threshold. The value of this threshold can significantly influence both the outcome of the z-score method and the classifier's accuracy. Therefore, the threshold value must be selected to achieve the highest classification accuracy. To this end, the training and testing processes were repeated 20 times for all ML classifiers using threshold values in the range of $[1, \cdots, 4.8]$ with an increment of 0.2. Previous studies [45], [46] reported that z-score thresholds between 1 and 3 provide the best performance. Figure 3 illustrates the relationship between classification accuracy and z-score threshold

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

Table 2. The Classification Accuracy Results of 10 Classifiers according to z-score Threshold Tuning (%).

Threshold	XGBoost	NN	CNN	RNN	AB	LSTM	GRU	RF	SVM	LR	Avg
1.0	88.89	77.78	88.89	88.89	88.89	55.55	88.89	88.89	100	100	86.67
1.2	91.67	95.83	91.67	87.5	95.83	95.83	87.5	95.83	100	95.83	93.75
1.4	97.56	97.56	97.56	80.49	95.12	90.24	95.12	95.12	92.68	97.56	93.90
1.6	98.18	100	100	92.72	96.36	87.27	89.09	98.18	98.18	100	95.99
1.8	100	98.51	98.51	92.54	98.51	92.54	94.03	100	98.51	98.51	97.16
2.0	94.81	94.81	97.40	92.21	97.40	89.61	92.21	97.40	97.40	97.40	95.07
2.2	97.59	100	100	98.8	98.8	90.36	94.0	97.6	100	100	97.71
2.4	93.33	95.56	95.56	91.11	91.11	88.89	91.11	90.0	93.33	94.44	92.44
2.6	95.75	98.94	98.94	96.81	96.81	90.42	93.62	95.75	96.81	98.94	96.28
2.8	93.81	93.81	93.81	86.6	96.91	84.54	85.57	94.85	94.85	94.85	91.96
3.0	98.99	100	100	96.97	98.99	92.93	94.95	98.99	98.99	100	**98.08**
3.2	97.06	97.06	99.02	98.04	99.02	90.2	94.12	98.04	99.02	99.02	97.06
3.4	97.12	99.04	98.08	97.12	98.08	85.58	92.31	97.12	99.04	98.08	96.16
3.6	98.10	99.05	99.05	93.33	98.10	90.48	87.62	97.14	98.1	99.05	96.0
3.8	98.13	98.13	99.07	96.26	98.13	93.46	95.33	95.33	97.20	97.20	96.82
4.0	95.33	96.26	97.2	92.52	97.20	89.72	91.59	94.39	95.33	97.19	94.67
4.2	99.07	98.15	97.2	95.37	99.07	95.37	95.37	96.29	97.22	97.22	97.04
4.4	97.25	98.17	98.17	95.41	99.08	91.74	94.49	96.33	95.41	95.41	96.15
4.6	99.09	99.09	99.09	97.27	99.09	90.91	95.45	98.18	98.18	99.09	97.55
4.8	97.29	96.4	97.3	90.99	97.30	85.59	91.89	97.30	95.49	93.70	94.32
Avg	96.45	96.71	97.33	93.05	97.0	88.56	92.22	96.14	97.29	**97.67**	

values. The vertical axis represents accuracy (in percentage), while the horizontal axis represents the z-score threshold. The classification accuracy of several methods reached 100% at certain threshold values. For instance, four classifiers—NN, CNN, SVM, and LR—achieved 100% accuracy when the threshold was 2.2, with the average accuracy of all classifiers being 97.71% at the same threshold. At a threshold of 3, only NN, CNN, and LR achieved 100% accuracy, while the overall average accuracy increased slightly to 98.08%. Although more classifiers achieved 100% accuracy at the threshold of 2.2, the threshold of 3 was selected in this study because it yielded the highest overall average accuracy. This finding is consistent with existing literature. A curve-fitting model was applied to the average accuracy of all classifiers as a function of the z-score threshold. The polynomial equation of the best-fitting curve is shown in Eq. (25).

$$y(x) = -0.3737x^6 + 6.9722x^5 - 52.889x^4 + 207.98x^3 - 444.96x^2 + 488.51x - 118.52 \quad (25)$$

The classification accuracy of all classifiers used in this study are presented in Table 2. Moreover, the average accuracy of each classifier over the threshold range is also provided in the last row of Table 2. As shown, the highest accuracy belongs to Logistic Regression (LR) with 97.67%, followed by CNN and SVM with 97.33% and 97.29%, respectively. The highest average accuracies are shown in bold. The

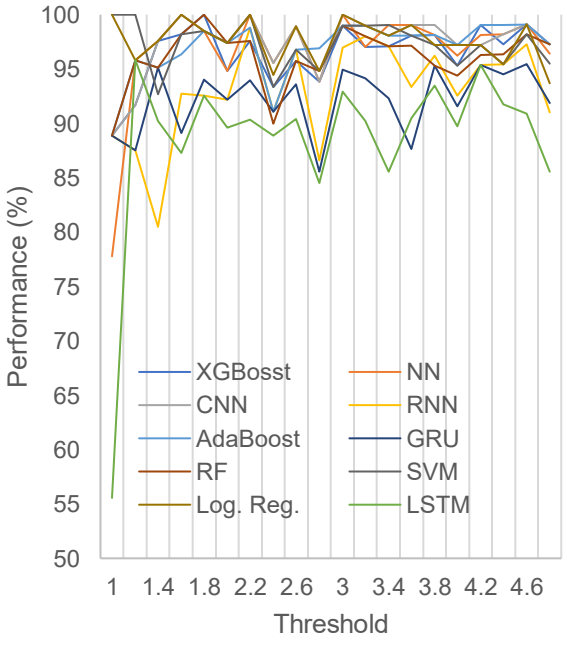

Fig. 3. Classification Accuracy of Classifiers for Different Threshold Values.

reason LR achieves such high accuracy is that it performs very well on linearly separable and clean datasets, especially when there are no outliers. On the other hand, CNN's strong performance is due to its ability to capture high-level features through

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

Table 3. The Accuracy and F1-Score Comparison of the Proposed Model with Base Methods.

Method	Proposed					Without outlier removal				
	Acc	F1	Pr	Rc	Conf. Int.	Acc	F1	Pr	Rc	Conf. Int.
XGBoost	98.99	98.99	99.00	98.99	[97.02, 100]	95.61	95.58	95.69	95.61	[91.85, 99.37]
NN	100	100	100	100	[100, 100]	97.37	97.37	97.37	97.36	[94.87, 99.87]
CNN	100	100	100	100	[100, 100]	98.25	98.25	98.23	98.25	[96.63, 99.87]
RNN	95.96	95.89	95.84	95.96	[93.25, 98.67]	89.47	89.56	92.98	92.98	[88.19, 91.10]
LSTM	92.93	92.77	93.01	92.93	[87.88, 97.98]	87.72	87.39	89.63	89.47	[83.84, 92.11]
AdaBoost	98.99	98.99	99.01	98.98	[97.02, 100]	95.61	95.58	95.69	95.61	[91.85, 99.37]
GRU	94.95	94.89	95.24	94.95	[90.64, 99.26]	90.35	90.45	90.60	90.23	[85.03, 95.42]
SVM	98.99	98.99	99.01	98.98	[97.02, 100]	97.37	97.37	97.39	97.37	[94.43, 100]
RF	98.99	98.99	99.01	98.98	[97.02, 100]	95.61	95.60	95.60	95.61	[91.85, 99.37]
LR	100	100	100	100	[100, 100]	98.25	98.25	98.25	98.25	[95.84, 100]

convolution. Figure 4 shows the confusion matrix of the Logistic Regression method on the dataset.

B. Breast Cancer Classification Results on WDBC Using ML Methods with and without z-score Filtering

This section presents the accuracy and F1-score results of the ML classifier methods used in this study, both with and without outlier removal on the WDBC dataset. The comparison is shown in Table 3. The "Proposed" columns refer to the results obtained on the dataset after applying z-score filtering with a threshold value of 3. The results represent the average of 5-fold cross-validation with an 80:20 train–test split.

According to the results in Table 3, for all models, the proposed method yields a substantial increase in both accuracy and F1-score on the WDBC dataset. The most significant improvement (effect) is observed in the RNN model. Removing the outliers using the z-score filtering method increased the accuracy of RNN by 6.49%. The Neural Network (NN), CNN, and LR classifiers achieved 100% accuracy and F1-score using the proposed method. This indicates that the technique, when applied with an appropriately selected threshold, can effectively handle outliers for these specific models, resulting in perfect classification on the given dataset. As shown in Table 3, in terms of both accuracy and F1-score, the least significant improvement belongs to Long Short-Term Memory (LSTM), with 92.93% and 92.77%, respectively.

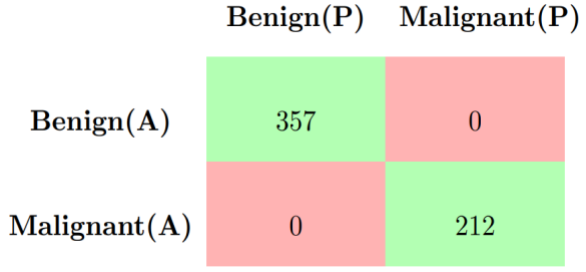

Fig. 4. Confusion Matrix for LR Classifier.

Table 4. Performance Comparison of Proposed Model with Existing Models on Basis of Accuracy

Method	Year	Accuracy (%)
Aamir [47] (MLP)	2022	99.12
Aamir [47] (RF)	2022	98.07
Aamir [47] (ANN)	2022	97.35
Mushtaq et al. [48]	2019	91.00
Rajaguru et al.[49]	2019	95.95
Khan et al.[50]	2020	97.06
Al-Azzam et al. [51]	2021	98.00
Rasool et al.[52]	2022	99.03
Zhou et al.[53]	2023	99.12
Proposed model	2025	**100**
Ghosh [54]	2024	98.25

IV. Discussion

This study aims to analyze and evaluate the effect of outlier removal from the dataset for the task of breast cancer classification. The findings in Table 3 show that removing outliers or noisy data samples improves the overall classification accuracy for breast cancer detection. In terms of accuracy, F1-score, precision and recall, the methods CNN, LR and NN achieved 100% when the outliers were removed. When the outliers were retained in the dataset, CNN and LR achieved 98.25%, while NN reached 97.37% across all metrics. The possible reason behind this high performance is that after the outliers are removed from the dataset, the distribution of the samples becomes more distinct and representative. The findings also show that the value of the threshold in the z-score method is a strong parameter influencing performance. When examining the results of XGBoost, there is approximately a 3% difference between the proposed and conventional models across all metrics.

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

Recurrent models such as RNN, GRU and LSTM showed the largest improvements since they are more sensitive to noisy data. In the case of RNN, the accuracy increased from 89.47% to 95.96%, while for LSTM, it rose from 87.72% to 92.93%. GRU improved from 90.35% to 94.95%. For AdaBoost and RF methods, a smaller improvement was observed, with an approximate 2% increase in accuracy between the proposed and conventional methods. Another important aspect is the swing in the confidence interval. A larger swing refers to high variation among the results of the same test in different trials, whereas a smaller swing indicates more consistent outcomes. This means that the performance of a method is less likely to be random. According to the confidence interval results shown in Table 3, the swing in the proposed method is smaller than that of the conventional method for all classifiers. This shows that outlier removal makes a systematic and deterministic contribution to classifier performance. An analysis of the precision and recall metrics provides deeper insights into the study. In the proposed method, both metrics are exceptionally high and balanced across all models. This implies that the classifiers can maintain a low false positive rate (high precision) while capturing most of the true positive cases (high recall). The strong correlation between precision and recall indicates that all models achieve a solid balance between sensitivity and specificity. This means that improvements in accuracy do not come at the cost of increased false alarms.

To demonstrate the effect of this study, a comparison was made with previous studies in the literature using the same dataset. Only studies conducted within the last five years were considered. During this comparison, we selected the result of the LR classifier with a z-score threshold of 3. The benchmarking results presented in Table 4 show the superior performance of the proposed model compared to previously reported state-of-the-art (SOTA) methods on the WDBC dataset. The proposed model achieved an accuracy of 100%, establishing a new benchmark in this domain and outperforming all prior approaches. This finding underscores the effectiveness of the methodological innovations introduced in this work and highlights their potential to address longstanding limitations of existing classification frameworks.

Earlier studies have reported varying degrees of success depending on methodological design and computational strategy. In [47], the authors applied different classifiers on the same dataset and obtained accuracies of 98.07%, 97.35%, and 99.12% for RF, ANN, and MLP, respectively. For instance, Mushtaq et al. [48] focused on evaluating k-nearest neighbor (KNN) performance using several distance functions and k values to identify an effective KNN configuration and obtained 91% accuracy. Rajaguru et al. [49] achieved 95.95% by applying Principal Component Analysis (PCA) + KNN. In a more recent study [50], the authors used fuzzy logic and SVM together to detect breast cancer and achieved 97.06% accuracy.

Al-Azzam et al. [51] demonstrated incremental improvements with an accuracy of 98.00%. Their study focused on the learning type rather than the classifier, showing that using a small sample of labeled data and low computational power, semi-supervised learning can replace supervised learning algorithms in tumor type diagnosis. Rasool et al. [52] and Zhou et al. [53] approached near-perfect classification with accuracies exceeding 99%. Zhou et al. conducted data exploratory techniques (DET) and developed four predictive models to improve breast cancer diagnostic accuracy. Prior to model development, four essential DET stages—feature distribution, correlation, elimination, and hyperparameter optimization—were performed. Advances in deep learning, optimization, and data augmentation are steadily pushing model performance closer to the theoretical maximum. Similarly, Ghosh [54] in 2024 reported an accuracy of 98.25%, further confirming the trend toward increasingly sophisticated methodologies with high predictive fidelity.

The proposed model's 100% accuracy and F1-score represent a significant improvement. This achievement suggests that removing abnormal data samples from the dataset enhances discriminative capacity through better representation learning, optimization of feature hierarchies, and superior handling of intra-class variability. The elimination of misclassifications implies that the model captures both global and local discriminative features with unprecedented precision, thus mitigating the error sources evident in prior approaches. The proposed model has the potential to support medical staff in cancer diagnosis. To apply this model in a real clinical environment, it should be considered a supportive system for doctors and imaging technicians. Its performance may be affected by data diversity and size, which can cause overfitting during training. In a medical application where patients require fast and reliable decisions based on their scanned images, additional measures might be necessary to prevent errors.

The proposed model demonstrably outperforms state-of-the-art (SOTA) methods. The model can classify breast cancer without errors. Future research should focus on generalizing the method to larger datasets and maintaining robustness to ensure the model's high accuracy can be applied in real-world scenarios.

V. Conclusion

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

This paper aims to present a framework for breast cancer detection. The proposed framework includes several stages: outlier detection and removal, hyperparameter tuning (hypertuning), and classification. Before training, a hyperparameter tuning study was conducted using the threshold value of the z-score outlier detection method to determine the optimal setting. The classification experiments were performed using XGBoost, NN, CNN, RNN, GRU, LSTM, SVM, RF, and LR on the WDBC dataset. With 100% accuracy and F1-score, the proposed model showed a significant improvement compared to the classifiers without outlier removal. Several classifiers achieved 100% accuracy; however, LR provided the best overall performance. The results indicate that the proposed model outperformed the state-of-the-art (SOTA) methods on the WDBC dataset. For future work, we plan to use ML to determine the threshold value based on dataset characteristics. In addition, we will focus on generalizing the model to larger and deeper datasets. Integration of hyperparameter tuning via the Optuna library will also be investigated.

Author Contribution

Eren Yildirim contributed to the conceptual design, implementation, experiments, and writing of the paper. Batu Salman contributed to code development and manuscript writing.

Data Availability

The data and the code in this study can be accessed at https://colab.research.google.com/drive/1oV8FIU_F66NdxQdhPjFb7lYW6mmnsNVR#scrollTo=qfeYHN93V5tK.

Declarations

Ethical Approval

This study used an online, publicly available dataset related to breast cancer classification. Therefore, it did not involve direct interaction with human subjects and did not require additional ethical approval.

Consent for Publication Participants.

Consent for publication was given by all participants.

Competing Interests

The authors declare no competing interests.

References

[1] H. Sung *et al.*, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[2] N. Harbeck *et al.*, “Breast cancer,” *Nat Rev Dis Primers*, vol. 5, no. 1, p. 66, Sep. 2019, doi: 10.1038/s41572-019-0111-2.

[3] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med Image Anal*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[4] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.

[5] W. Wolberg, O. Mangasarian, N. Street, and W. Street. "Breast Cancer Wisconsin (Diagnostic)," UCI Machine Learning Repository, 1993. [Online]. Available: https://doi.org/10.24432/C5DW2B.

[6] W. H. Wolberg and O. L. Mangasarian, “Multisurface method of pattern separation for medical diagnosis applied to breast cytology.,” *Proceedings of the National Academy of Sciences*, vol. 87, no. 23, pp. 9193–9196, Dec. 1990, doi: 10.1073/pnas.87.23.9193.

[7] M. Elter. "Mammographic Mass," UCI Machine Learning Repository, 2007. [Online]. Available: https://doi.org/10.24432/C53K6Z.

[8] M. Elter, R. Schulz-Wendtland, and T. Wittenberg, “The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process,” *Med Phys*, vol. 34, no. 11, pp. 4164–4172, Nov. 2007, doi: 10.1118/1.2786864.

[9] Z. Khandezamin, M. Naderan, and M. J. Rashti, “Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier,” *J Biomed Inform*, vol. 111, p. 103591, Nov. 2020, doi: 10.1016/j.jbi.2020.103591.

[10] A. S. Yaro, F. Maly, P. Prazak, and K. Malý, “Outlier Detection Performance of a Modified Z-Score Method in Time-Series RSS Observation With Hybrid Scale Estimators,” *IEEE Access*, vol. 12, pp. 12785–12796, 2024, doi: 10.1109/ACCESS.2024.3356731.

[11] C. C. Aggarwal, *Outlier Analysis*. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-47578-3.

[12] A. Zimek, E. Schubert, and H. Kriegel, “A survey on unsupervised outlier detection in high-dimensional numerical data,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 5, no. 5, pp. 363–387, Oct. 2012, doi: 10.1002/sam.11161.

[13] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene Selection for Cancer Classification using Support Vector Machines,” *Mach Learn*, vol. 46, no. 1–3, pp. 389–422, Jan. 2002, doi: 10.1023/A:1012487302797.

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

[14] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," in *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, IEEE, Feb. 2016, pp. 78–83. doi: 10.1109/IACC.2016.25.

[15] L. Breiman, "Random Forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[16] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 2017.

[17] E. Genc, M. E. Yildirim, and Y. B. Salman, "Human activity recognition with fine-tuned CNN-LSTM," *Journal of Electrical Engineering*, vol. 75, no. 1, pp. 8–13, Feb. 2024, doi: 10.2478/jee-2024-0002.

[18] V. Fascianelli, A. Battista, F. Stefanini, S. Tsujimoto, A. Genovesio, and S. Fusi, "Neural representational geometries reflect behavioral differences in monkeys and recurrent neural networks," *Nat Commun*, vol. 15, no. 1, p. 6479, Aug. 2024, doi: 10.1038/s41467-024-50503-w.

[19] Y. Zhang, S. Xu, L. Zhang, W. Jiang, S. Alam, and D. Xue, "Short-term multi-step-ahead sector-based traffic flow prediction based on the attention-enhanced graph convolutional LSTM network (AGC-LSTM)," *Neural Comput Appl*, vol. 37, no. 20, pp. 14869–14888, Jul. 2025, doi: 10.1007/s00521-024-09827-3.

[20] L. Zhang, J. Zhang, W. Gao, F. Bai, N. Li, and N. Ghadimi, "A deep learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and improved orca predation algorithm," *Biomed Signal Process Control*, vol. 90, p. 105858, Apr. 2024, doi: 10.1016/j.bspc.2023.105858.

[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[22] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med Image Anal*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[23] Ö. F. İnce, I. F. Ince, M. E. Yıldırım, J. S. Park, J. K. Song, and B. W. Yoon, "Human activity recognition with analysis of angles between skeletal joints using a RGB-depth sensor," *ETRI Journal*, vol. 42, no. 1, pp. 78–89, Feb. 2020, doi: 10.4218/etrij.2018-0577.

[24] S. M. Kasongo, "A deep learning technique for intrusion detection system using a Recurrent Neural Networks based framework," *Comput Commun*, vol. 199, pp. 113–125, Feb. 2023, doi: 10.1016/j.comcom.2022.12.010.

[25] M. Abdar *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, Dec. 2021, doi: 10.1016/j.inffus.2021.05.008.

[26] N. Munir, J. Huang, C.-N. Wong, and S.-J. Song, "Machine learning based eddy current testing: A review," *Results in Engineering*, vol. 25, p. 103724, Mar. 2025, doi: 10.1016/j.rineng.2024.103724.

[27] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 2017.

[28] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI," Aug. 2020, doi: 10.1109/TNNLS.2020.3027314.

[29] M. B. Lopes, A. Veríssimo, E. Carrasquinha, S. Casimiro, N. Beerenwinkel, and S. Vinga, "Ensemble outlier detection and gene selection in triple-negative breast cancer data," *BMC Bioinformatics*, vol. 19, no. 1, p. 168, Dec. 2018, doi: 10.1186/s12859-018-2149-7.

[30] S. H. Ali and M. Shehata, "A New Breast Cancer Discovery Strategy: A Combined Outlier Rejection Technique and an Ensemble Classification Method," *Bioengineering*, vol. 11, no. 11, p. 1148, Nov. 2024, doi: 10.3390/bioengineering11111148.

[31] A. Alloqmani, Y. B. Abushark, and A. I. Khan, "Anomaly Detection of Breast Cancer Using Deep Learning," *Arab J Sci Eng*, vol. 48, no. 8, pp. 10977–11002, Aug. 2023, doi: 10.1007/s13369-023-07945-z.

[32] J. Lv and L. Wang, "Hybrid modeling of adsorption process using mass transfer and machine learning techniques for concentration prediction," *Journal of Saudi Chemical Society*, vol. 29, no. 4, p. 12, Sep. 2025, doi: 10.1007/s44442-025-00016-y.

[33] D. Wu, X. Ma, and D. L. Olson, "Financial distress prediction using integrated Z-score and multilayer perceptron neural networks," *Decis Support Syst*, vol. 159, p. 113814, Aug. 2022, doi: 10.1016/j.dss.2022.113814.

[34] A. M. Sharifnia, D. E. Kpormegbey, D. K. Thapa, and M. Cleary, "A Primer of Data Cleaning in Quantitative Research: Handling Missing Values and Outliers," *J Adv Nurs*, Mar. 2025, doi: 10.1111/jan.16908.

[35] W. Li, Y. Yin, X. Quan, and H. Zhang, "Gene Expression Value Prediction Based on XGBoost Algorithm," *Front Genet*, vol. 10, Nov. 2019, doi: 10.3389/fgene.2019.01077.

[36] Martin. Anthony, *Neural Network Learning : Theoretical Foundations*. Cambridge University Press, 2022.

[37] E. Genc, M. E. Yildirim, and Y. B. Salman, "Human activity recognition with fine-tuned CNN-LSTM," *Journal of Electrical Engineering*, vol. 75,

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

no. 1, pp. 8–13, Feb. 2024, doi: 10.2478/jee-2024-0002.
[38] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D*, vol. 404, p. 132306, Mar. 2020, doi: 10.1016/j.physd.2019.132306.
[39] X. Li, L. Wang, and E. Sung, "AdaBoost with SVM-based component classifiers," *Eng Appl Artif Intell*, vol. 21, no. 5, pp. 785–795, Aug. 2008, doi: 10.1016/j.engappai.2007.07.001.
[40] M. A. Chandra and S. S. Bedi, "Survey on SVM and their application in image classification," *International Journal of Information Technology*, vol. 13, no. 5, pp. 1–11, Oct. 2021, doi: 10.1007/s41870-017-0080-1.
[41] T. Perumal, N. Mustapha, R. Mohamed, and F. M. Shiri, "A Comprehensive Overview and Comparative Analysis on Deep Learning Models," *Journal on Artificial Intelligence*, vol. 6, no. 1, pp. 301–360, 2024, doi: 10.32604/jai.2024.054314.
[42] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104, Jan. 2012, doi: 10.1016/j.isprsjprs.2011.11.002.
[43] S. Sperandei, "Understanding logistic regression analysis," *Biochem Med (Zagreb)*, pp. 12–18, 2014, doi: 10.11613/BM.2014.003.
[44] A. Masbakhah, U. Sa'adah, and M. Muslikh, "Heart Disease Classification Using Random Forest and Fox Algorithm as Hyperparameter Tuning," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 4, pp. 964–976, Aug. 2025, doi: 10.35882/jeeemi.v7i4.932.
[45] L. De Coster *et al.*, "On the optimal z-score threshold for SISCOM analysis to localize the ictal onset zone," *EJNMMI Res*, vol. 8, no. 1, p. 34, Dec. 2018, doi: 10.1186/s13550-018-0381-9.
[46] A. Curtis, T. Smith, B. Ziganshin, and J. Elefteriades, "The Mystery of the Z-Score," *AORTA*, vol. 04, no. 04, pp. 124–130, Aug. 2016, doi: 10.12945/j.aorta.2016.16.014.
[47] S. Aamir *et al.*, "Predicting Breast Cancer Leveraging Supervised Machine Learning Techniques," *Comput Math Methods Med*, vol. 2022, pp. 1–13, Aug. 2022, doi: 10.1155/2022/5869529.
[48] Z. Mushtaq, A. Yaqub, S. Sani, and A. Khalid, "Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets," *Journal of the Chinese Institute of Engineers*, vol. 43, no. 1, pp. 80–92, Jan. 2020, doi: 10.1080/02533839.2019.1676658.
[49] H. Rajaguru and S. C. S R, "Analysis of Decision Tree and K-Nearest Neighbor Algorithm in the Classification of Breast Cancer," *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 12, pp. 3777–3781, Dec. 2019, doi: 10.31557/APJCP.2019.20.12.3777.
[50] F. Khan *et al.*, "Cloud-Based Breast Cancer Prediction Empowered with Soft Computing Approaches," *J Healthc Eng*, vol. 2020, pp. 1–16, May 2020, doi: 10.1155/2020/8017496.
[51] N. Al-Azzam and I. Shatnawi, "Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer," *Annals of Medicine and Surgery*, vol. 62, pp. 53–64, Feb. 2021, doi: 10.1016/j.amsu.2020.12.043.
[52] A. Rasool, C. Bunterngchit, L. Tiejian, Md. R. Islam, Q. Qu, and Q. Jiang, "Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis," *Int J Environ Res Public Health*, vol. 19, no. 6, p. 3211, Mar. 2022, doi: 10.3390/ijerph19063211.
[53] S. Zhou, C. Hu, S. Wei, and X. Yan, "Breast Cancer Prediction Based on Multiple Machine Learning Algorithms," *Technol Cancer Res Treat*, vol. 23, Jan. 2024, doi: 10.1177/15330338241234791.
[54] P. Ghosh and D. Chatterjee, "Comparative Analysis of Machine Learning Algorithms for Breast Cancer Classification: SVM Outperforms XGBoost, CNN, RNN, and Others," Apr. 2024, doi: 10.1101/2024.04.22.590658.

Author Biography

Eren Yildirim received his B.S. degree in Electrical Engineering from Bahcesehir University, Istanbul, Turkey, in 2008, and his M.S. and Ph.D. degrees in Electronics Engineering from the Graduate School of Electrical and Electronics Engineering, Kyungsung University, Busan, Republic of Korea, in 2010 and 2014, respectively. He worked as a researcher and lecturer at Kyungsung University until August 2015. He later served as an assistant professor in the Department of Electronics Engineering, Kyungsung University, between 2019 and 2023. He is currently holding two positions as an associate professor in the Department of Electrical and Electronics Engineering, Bahcesehir University, and the Department of Electronics and Communications Engineering, American University of Malta. His research interests include image processing, computer vision, and machine learning.

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165

Yucel Batu Salman received his B.S. and M.S. degrees in Computer Engineering from Bahcesehir University, Istanbul, Turkey, in 2003 and 2005, respectively, and his Ph.D. degree in IT Convergence Design from Kyungsung University, Busan, Republic of Korea, in 2010. Since 2010, he has been with the Department of Software Engineering, Bahcesehir University, Istanbul, Turkey, where he is an associate professor. He is currently the Director of the Graduate School at Bahcesehir University. His research interests include human-computer interaction, mobile programming, and computer vision.

Manuscript received July 5, 2024; Revised September 15, 2025; Accepted October 28, 2025; date of publication October 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1165