

A machine learning assisted eight-element metamaterial antenna array for non-invasive breast cancer detection

Muneera Altayeb¹, Sanaa Salama², Duaa Zyoud², Ashraf Abuelhaija³

¹Department Communications and Computer Engineering, Faculty of Engineering, Al-Ahliyya Amman University, Amman, Jordan

²Department Medical Equipment Engineering, Faculty of Engineering, Arab American University, Jenin, Palestine

³Department of Electrical Engineering, Faculty of Engineering, Applied Science Private University, Amman, Jordan

Article Info

Article history:

Received Feb 7, 2026

Revised Aug 24, 2026

Accepted Sep 4, 2026

Keywords:

Adaptive boosting

Antenna sensor

Gradient boosting

Microwave imaging

Random forest

ABSTRACT

Breast cancer remains one of the most prevalent cancers and a leading cause of cancer-related mortality among women worldwide, highlighting the need for accurate and non-invasive early detection methods. This study presents an integrated breast cancer diagnosis system that combines a microwave imaging platform composed of eight metamaterial-based antenna elements with machine learning (ML) techniques for breast tissue classification. The proposed non-ionizing approach utilizes backscattered S-parameters collected from the antenna array over the 3-7 GHz frequency range to identify tumor masses with diameters of 20 mm, 25 mm, and 30 mm. A dataset comprising 32,048 samples was generated from the measured S-parameter responses and categorized into four classes: healthy breast tissue and tumor cases with diameters of 20 mm, 25 mm, and 30 mm. Three ML classifiers, namely adaptive boosting (AdaBoost), gradient boosting (GB), and random forest (RF), were trained and evaluated using metrics including classification accuracy (CA), Matthews correlation coefficient (MCC), F1-score, recall, and precision. Experimental results demonstrate the effectiveness of the proposed framework, with AdaBoost achieving the best performance and an accuracy of 99.8%, outperforming the other classifiers. The findings confirm the potential of combining metamaterial-based microwave imaging with ML techniques for accurate, non-invasive, and early breast cancer diagnosis.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Muneera Altayeb

Department Communications and Computer Engineering, Faculty of Engineering

Al-Ahliyya Amman University

Amman, Jordan

Email: m.altayeb@ammanu.edu.jo

1. INTRODUCTION

Breast cancer remains a leading cause of death among women worldwide, creating an urgent need for accurate and effective diagnosis tools. While established imaging techniques such as mastography, ultrasound, and magnetic resonance imaging (MRI) have proven effective, they face limitations, including radiation exposure, with some models only operating for two months without radiation. These challenges have spurred research into smaller imaging devices that accurately capture contrast between tissues and offer safer and more accessible solutions. Modern advancements in antenna array design, particularly those based on metamaterials, have enabled a significant improvement in the visibility and accuracy of micro-imaging. By capturing backscattered signals from breast tissue, antenna arrays generate sufficient datasets that reflect

the variations in tissue properties. However, to extract meaningful diagnostic information from the complex and high dimensional backscattered signals, demand for advanced analytical frameworks is required.

Data science, artificial intelligence, and machine learning (ML) have witnessed rapid development in recent years and have become an integral part of all aspects of life, especially the medical field, which processes vital signs, medical images, and historical diagnostic data to predict and classify many diseases. The literature reviews several works similar to the ones presented in this study. Different ML algorithms were utilized in [1] to forecast tumors in the mammary gland based on the effective relative permittivity of breast tissues. The data set samples were created using simulated S-parameters for a textile-based antenna sensor placed on the normal breast tissue and the abnormal breast tissue with a tumor of a variable size, 5 mm, 10 mm, and 20 mm. The decision tree (DT) algorithm achieves accuracies of 58% and 54% for tumor sizes of 10 mm and 20 mm, respectively. The random forest (RF) algorithm has an accuracy of 58% and 31% for 10 mm and 20 mm tumor sizes, respectively. The CatBoost algorithm achieves the highest accuracy, with rates of 67% and 69% for tumor sizes of 10 mm and 20 mm, respectively.

The work in [2] aimed to predict and diagnose breast cancer using ML techniques, including support vector machine (SVM), RF, logistic regression (LR), DT (C4.5), and k-nearest neighbors (KNN). The accuracy percentage for winson breast cancer diagnostic (WDBC) datasets is 97.2%, 96.5%, 95.8%, 95.1%, and 93.7% for SVM, RF, LR, DT (C4.5), and KNN, respectively. Chaurasia and Pal [3] compare six ML algorithms: SVM, naïve bayes (NB), classification and regression tree (CART), KNN, LR, and multilayer perceptron (MLP) on the WDBC dataset. The primary objective of [3] was to enhance the precision of prediction by employing a new statistical approach to feature selection. The trees classifier obtained the highest accuracy: 95.1739%, followed by extreme gradient boosting (XGBoost) 95.1691%, adaptive boosting (AdaBoost) 94.7343%, RF 94.7295%, Bagging 94.5169%, and gradient boosting (GB) classifier 93.8599%. Islam *et al.* [4], five ML algorithms: SVM, KNN, RF, artificial neural networks (ANNs), and LR were compared on the WDBC dataset. Results revealed that the ANNs obtained the highest accuracy of 98.57%, whereas 97.14% accuracy was obtained by SVM and KNN. The SVM and RF classifier in [5] obtained an accuracy of 96.5%. The highest accuracy achieved for ANN and CNN is 99.3% and 97.3%, respectively. Nemade and Fegade [6] use various ML classification algorithms: NB, LR, SVM, KNN, DT, and ensemble algorithms: RF, AdaBoost, XGBoost on the breast cancer dataset. Both the DT and XGBoost classifier achieved the highest accuracy of 97% followed by KNN, RF, AdaBoost, and LR with an accuracy of 96%. NB obtained the lowest accuracy of 90%.

Research by Srivenkatesh [7] a prediction model was suggested to predict breast cancer. A highest accuracy of 98.24% was obtained by RF. The SVM achieved an accuracy of 75.32% followed by KNN, NB, and LR with an accuracy of 64%, 52.63%, and 47.36%, respectively. According to Rabiei *et al.* [8], RF also achieved the highest accuracy of 80%. While the highest accuracy of 92.7% was for SVM. Shravya *et al.* [9] reported that the accuracy of the LR, KNN, RF, DT, and SVM classifiers (also evaluated by Mahesh *et al.* [10]) is 96.32%, 94.11%, 95.22%, 97.20%, and 93.46%, respectively. The SVM model in [11] had achieved 96.25% accuracy. The RF algorithm was employed in [12] and achieved 92.2% accuracy. Akbugday [13] employed SVM in combination with KNN for breast cancer classification and attained an accuracy of 96.85%. Results in [14] indicated that the DT (C5) is the best predictor with 93.6% accuracy. The ANN and the LR achieved an accuracy of 91.2% and 89.2%, respectively. Assiri *et al.* [15] reported that simple LR achieved the highest accuracy (98.25%), whereas NB yielded the lowest (91.81%), with KNN and RF obtaining 97.08% and 96.49%, respectively. Furthermore, experimental results by Asri *et al.* [16] demonstrated that SVM gave the highest accuracy at 97.13% with the lowest error rate.

Despite the significant progress achieved in metamaterial-based microwave sensing and ML-assisted diagnostic systems, limited attention has been devoted to the development of integrated frameworks that combine advanced microwave sensing hardware with intelligent classification techniques for breast cancer detection. Existing studies have predominantly focused either on improving antenna performance and microwave response analysis or on developing ML models independently. In particular, breast tumor detection approaches based on metamaterial antenna arrays have mainly relied on the interpretation of S-parameter variations and electromagnetic response characteristics, which often require expert analysis and provide limited automation in the diagnostic process. Consequently, a research gap remains in establishing a unified framework that combines high-sensitivity microwave sensing with automated and reliable decision-making capabilities.

To address this gap, this work presents a ML-assisted breast cancer detection framework employing an eight-element metamaterial antenna array as a microwave sensing platform. The proposed antenna array enhances the interaction between electromagnetic fields and breast tissues, thereby increasing sensitivity to tumor-induced dielectric variations. The collected microwave responses are subsequently processed using ML techniques to enable automated classification of healthy and tumor-affected cases. Unlike conventional approaches that rely solely on S-parameter contrast analysis, the proposed framework incorporates data-driven predictive modeling to support intelligent diagnosis and reduce dependence on manual

interpretation of sensing data. The proposed system is evaluated using simulated breast phantom datasets, demonstrating high classification accuracy (CA) and reliable detection performance. By integrating enhanced microwave sensing capabilities with ML-based analysis, this work advances the development of practical computer-aided breast cancer detection systems and establishes a foundation for future validation using experimental measurements. This paper presents the proposed work in five sections: introduction, methodology, results and model evaluation, discussion, and conclusions.

2. METHOD

This work integrates an eight-element metamaterial-based antenna array with ML techniques to develop a reliable, non-invasive framework for breast cancer detection and tumor size classification. The proposed framework consists of three distinct stages. First, the antenna array was designed and simulated in computer simulation technology (CST) microwave studio to evaluate and optimize its electromagnetic performance. Second, based on the optimized CST design, an antenna prototype was fabricated and experimentally validated using a vector network analyzer (VNA). The measured scattering parameters were compared with the simulated results to verify the accuracy and performance of the proposed antenna array. Third, the breast-tissue dataset was generated through CST-based numerical simulations using a multilayer breast model consisting of skin, fat, and tumor tissues with the corresponding dielectric properties.

The validated antenna-array configuration was then employed to obtain the S-parameters for different breast-tissue scenarios. It should be emphasized that the experimental measurements from the fabricated prototype were used solely to validate the antenna performance, whereas the ML dataset was generated entirely from CST breast-tissue simulations. The acquired S-parameter responses were sampled over the frequency range of 3-7 GHz and subsequently processed to generate the ML dataset. The dataset consisted of four classes corresponding to a healthy breast and breasts containing tumors with diameters of 20 mm, 25 mm, and 30 mm. In this work three ensemble learning algorithms were chosen due to their effectiveness in biomedical signal classification: AdaBoost, GB, and RF models. In model evaluation the following evaluation metrics were employed: CA, matthews correlation coefficient (MCC), F1-score, recall (sensitivity), and precision. As illustrated in Figure 1, the proposed model structure consists of multiple stages that work together to achieve an accurate and reliable diagnosis of breast tumors.

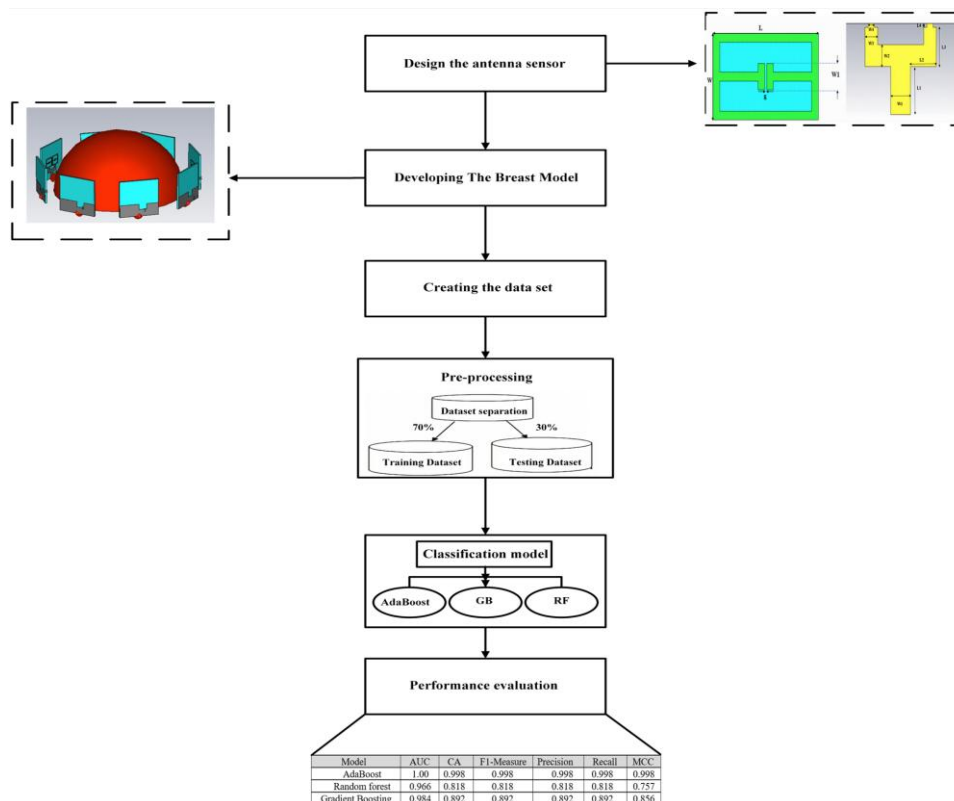


Figure 1. The proposed model architecture

2.1. Antenna sensor design

The proposed antenna sensor in this work is fabricated on RO3010 substrate with a dielectric constant of 10.2 and a loss tangent of 0.0023. The substrate dimensions are $30 \times 30 \times 1.28$ mm³. A T-shaped feed network is employed to excite the two H-shaped unit cells. The antenna geometry is illustrated in Figure 2 specifically, Figure 2(a) shows the H-shaped radiating structure, where the overall dimensions are represented by L and W , while W_1 and g denote the slot width and gap dimensions, respectively. In addition, Figure 2(b) presents the T-shaped microstrip line (MSL) feed network, where the parameters L_1 - L_4 and W_1 - W_4 define the lengths and widths of the feeding structure. These geometrical parameters were carefully optimized to achieve the desired antenna performance as summarized in Table 1.

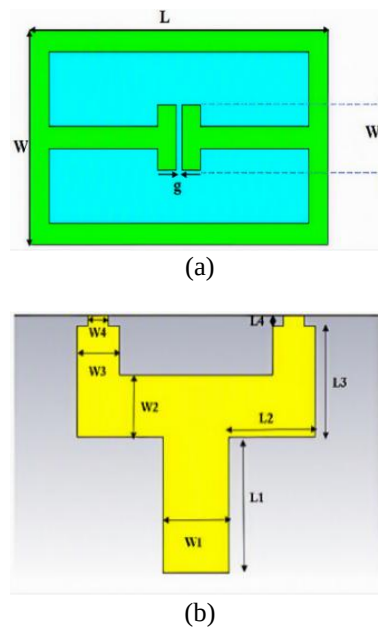


Figure 2. Proposed antenna geometry: (a) H-shaped design and (b) T-shaped MSL feed network

Table 1. Parameters of the H-shaped and T-shaped model

H-shaped (mm)	T-shaped (mm)
$L=8.00$	$L_1, W_1=5.5, 3.0$
$W=8.00$	$L_2, W_2=4.0, 2.5$
$W_1=1.50$	$L_3, W_3=4.5, 2.0$
$g=0.15$	$L_4, W_4=0.5, 1.0$

The antenna sensor is printed on a partial ground plane for a wider frequency band. The fabricated prototype in Figures 3(a) and (b) shows the top and bottom views. The proposed antenna sensor operates over a wide frequency range of 3 GHz to 7 GHz with simulated reflection coefficient of $|S_{11}| < -10$ dB. The reflection coefficients of the proposed antenna model in free space, both simulated and measured, are depicted in Figure 4.

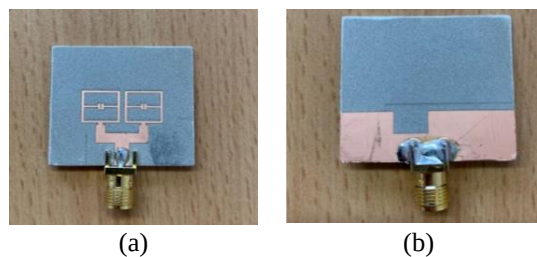


Figure 3. Proposed antenna prototype; (a) top view and (b) bottom view

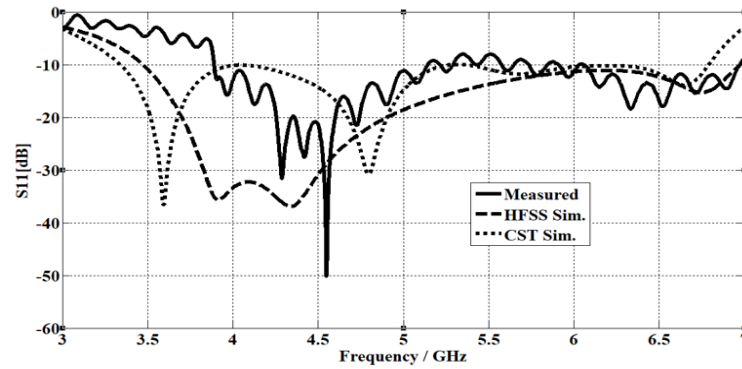


Figure 4. The input reflection coefficient (S_{11}) of the suggested antenna sensor

The proposed antenna system consists of an eight-element metamaterial antenna array arranged around the breast phantom to provide multi-directional illumination and data acquisition, in which the antenna elements are uniformly distributed in a circular configuration around the breast phantom. Each antenna element is positioned at 10 mm from the breast surface to ensure sufficient electromagnetic field penetration while maintaining strong interaction with the breast tissues. The adjacent antenna elements are separated by 17 mm, a distance selected to mitigate mutual coupling effects and improve the independence of the received scattering responses. The circular arrangement provides angular coverage of the entire breast volume, enabling the acquisition of microwave data from multiple directions and enhancing the detectability of tumor-induced dielectric variations. Furthermore, the use of eight elements represents a practical compromise between imaging performance and system complexity. A smaller number of elements may limit spatial coverage and reduce the diversity of the collected microwave information, whereas a larger number of elements would increase hardware requirements, data acquisition time, and computational burden. Therefore, the adopted eight-element configuration offers sufficient spatial sampling for reliable breast tumor detection while maintaining a compact and efficient imaging system.

2.2. The breast model

The breast model implemented in this study consists of three tissue layers that emulate the electromagnetic characteristics of the human breast. The outer skin layer has a radius of 50 mm, a thickness of 2 mm, a conductivity of 3.317 S/m, and a relative permittivity of $\epsilon_r = 37.69$. The inner fat tissue layer is characterized by a conductivity of 0.242 S/m and a relative permittivity of $\epsilon_r = 5.02$. The cancerous tumor tissue is modeled with a conductivity of 5 S/m and a relative permittivity of $\epsilon_r = 55$ [17]. The contrast in dielectric properties among these tissues influences the propagation, attenuation, reflection, and scattering of microwave signals within the breast model. In particular, the higher conductivity and permittivity of the tumor tissue compared with the surrounding fatty tissue create a significant dielectric contrast, resulting in observable variations in the measured scattering parameters. These variations are reflected in the S_{11} and S_{21} responses and become more pronounced with increasing tumor size, providing discriminative features for the subsequent ML-based classification of healthy and tumor-containing breast tissues.

2.3. Dataset

The dataset used in this work is based on backscattered signals obtained from the proposed metamaterial-based eight elements antenna array. These signals represent the S-parameters (reflection coefficients S_{ii} , and transmission coefficients S_{ji} , for $i, j = 1, 2, \dots, 8$) backscattered from the breast model and captured by the antenna array. To generate the dataset, the eight sensors were arranged in a circle around a breast phantom model, with a 10 mm spacing between each adjacent sensor, as shown in Figure 5. A simulated tumor model of varying sizes (20 mm, 25 mm, and 30 mm) was placed within the breast phantom.

For consistency and reproducibility, the tumor was placed at a fixed position within the breast phantom throughout all simulations/measurements. Additionally, the tumor depth was kept constant for all investigated cases. This controlled configuration enabled the assessment of the antenna performance under identical conditions and facilitated a direct comparison of the obtained results. The S_{21} transmission coefficients (in dB) of the breast phantom containing tumors of different sizes was simulated using CST microwave studio software [18]-[21]. The dataset was generated from the frequency-dependent S-parameter responses over the 3-7 GHz frequency range. A total of 32,048 samples were obtained and categorized into four classes: 8,008 samples corresponding to a healthy breast, 8,016 samples corresponding to a breast

containing a 20 mm tumor, 8,016 samples corresponding to a breast containing a 25 mm tumor, and 8,008 samples corresponding to a breast containing a 30 mm tumor.

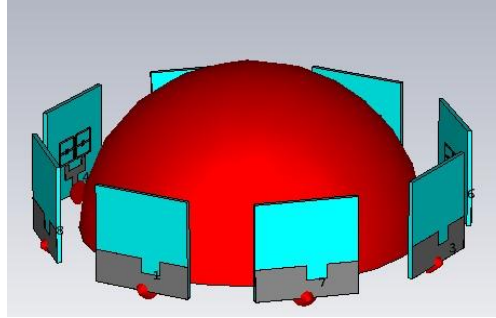


Figure 5. Circular configuration of eight antenna sensors around the breast model

2.4. Features extraction

The input features for the ML models were derived from the frequency-dependent scattering parameters obtained from the proposed eight-element antenna array. Specifically, the magnitudes of the reflection coefficient and transmission coefficient were acquired over the 3-7 GHz frequency range to characterize the electromagnetic response of the investigated breast models. The sampled S-parameter magnitude values obtained from all antenna elements were organized into a feature matrix, where each sampled value was treated as an independent input feature for the classification task. The phase information was not considered in this study, as the magnitude responses provide a direct representation of the variations in reflected and transmitted electromagnetic energy caused by differences in the permittivity properties of healthy and tumorous breast tissues. In addition, magnitude-based features offer a more stable and robust representation for classification by reducing the influence of phase variations and measurement uncertainties. The sampled S-parameter parameters capture the most relevant information related to impedance mismatch and electromagnetic-wave propagation within the breast phantom, making them suitable descriptors for tumor detection and size classification. The resulting dataset comprised four classes corresponding to a healthy breast and breasts containing tumors with diameters of 20 mm, 25 mm, and 30 mm.

2.5. Classification

Classification and pattern recognition are among the most prominent applications of artificial intelligence, particularly in the field of ML. It has now become an integral part of all fields of science, especially in the medical field. In this work, a model was built to detect breast tumors and classify the results into four basic categories: a healthy chest without a tumor. If a tumor is detected, it is classified into three categories based on its size: 20 mm, 25 mm, and 30 mm. The classification system relies on extracting key features from a dataset. A hold-out validation approach was employed, where the dataset was randomly shuffled and subdivided into 70% training samples and 30% testing samples. The training subset was used for model construction, whereas the testing subset was used exclusively for performance evaluation. This strategy enables an unbiased assessment of the classification capability. The system uses three main classification algorithms: AdaBoost, GB, and RF.

AdaBoost is an ensemble learning algorithm that improves classification performance by combining multiple weak learners, typically DTs, into a stronger predictive model. The algorithm operates iteratively by assigning higher weights to misclassified instances, forcing subsequent learners to focus on the more challenging samples.

2.5.1. Classification using adaptive boosting

AdaBoost is a well-known technique that combines weak base binary classifiers one after the other to create a powerful classifier [22]. Shallow DTs serve as its foundational classifiers. The training sample is reweighted at each iteration so that the subsequent DT [23]. This method aims to rectify incorrect classifications in subsequent iterations, ultimately producing accurate classifications. In (1) illustrates that when a new tree model is added, the general tree is eliminated and the system simply adds the strongest tree.

With this method, as more calculations are performed, the model's overall performance gradually improves [24]. The newly added tree is symbolized by y_i , the i^{th} tree, and the entire model is denoted by $F_n(x)$ when the inserted tree is indicated by $h(x_i)$ [25].

$$Fn(x) = F_{m-1}(x) + \operatorname{argmin}_h \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + h(x_i)) \quad (1)$$

2.5.2. Classification using gradient boosting

GB is a technique that is used for both regression and classification model [26]. The GB approach trains each tree using the residual errors of the previous trees, as opposed to the RF approach [27]. The remaining defects are prioritized in the model's subsequent trees. The GB classifier defines its prediction by a cumulative addition to the predictions of all the weak learners. The GB ensemble's final forecast is given by (2), where $f_i(x)$ is the prognosis of the i^{th} weak learner, where x is the input data point, and n is the number of weak learners [28].

$$\hat{Y}(x) = \sum_{i=1}^n f_i(x) \quad (2)$$

2.5.3. Classification using random forest

The RF classifier, a collective ML approach, mixes multiple trees during training [29]. A randomly selected subset determines each feature split, and each tree within the forest is trained using a randomly selected portion of the training data. RFs can be applied to a categorical response variable, referred to as “classification,” or a continuous response variable known as “regression”. In the same way, predictor variables may be categorical or continuous. RF mitigate correlation between DTs by using random sampling of samples and features [30]. Initially, an equivalent amount of data is randomly selected from the original training data. Further, a subset of features is randomly selected to construct the DT. Employing these two methods of randomization diminishes the correlation among DTs, thereby alleviating possible overfitting mistakes and improving model precision [31].

The prediction of a RF classifier is calculated by summing the majority votes of each tree, as shown in Figure 6. The majority votes from each tree are combined to determine the forecasting of the RF classifier. The majority vote from each tree is combined to determine the forecasting of a RF classifier. $T_i(x)$ represents the i^{th} tree forecast. In (3) determines the RF ensemble's final prediction, where the input data point is denoted by x [32].

$$\hat{Y}(x) = \operatorname{mode} \{T1(x), T2(x), \dots, Tn(x)\} \quad (3)$$

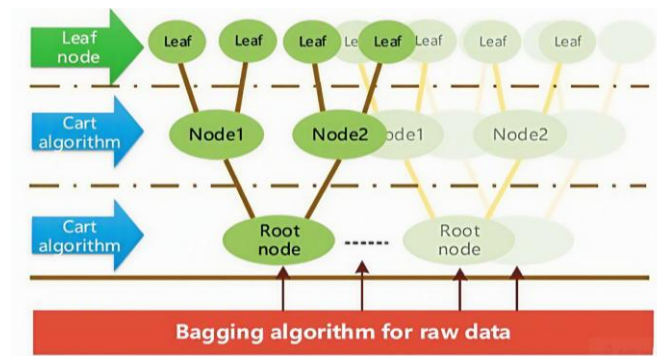


Figure 6. RF structure [33]

3. RESULTS AND MODEL EVALUATION

The prediction ability and efficacy of the suggested model must be assessed after its primary hypotheses have been examined. As a result, assessment metrics were used to assess the suggested models' usefulness [34]. As illustrated in Table 2, the confusion matrix is a helpful method for assessing the algorithm's effectiveness. The values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) outcomes are determined by the inaccuracy rate computations [35], [36].

Table 2. Confusion matrix [37]		
Confusion matrix	Patient case	
	Normal	Abnormal
Normal	TP	FP
Abnormal	FN	TN

The classification efficacy and accuracy of three models (AdaBoost, RF, and GB) were compared based on confusion matrices, as shown in Table 3, which presents the confusion matrix for tumor detection using AdaBoost. Table 4 depicts the confusion matrix for GB and Table 5 shows the confusion matrix for the RF classifier.

Table 3. AdaBoost confusion matrix

Actual	Σ	Predicted			
		Healthy	Tumor d/30 mm	Tumor d/25 mm	Tumor d/20 mm
Tumor d/20 mm	8016	10	4	17	7985
Tumor d/25 mm	8016	8	10	7998	0
Tumor d/30 mm	8008	7	8001	0	0
Healthy	8008	8008	0	0	0
Σ	32048	8033	8015	8015	7985

Table 4. GB confusion matrix

Actual	Σ	Predicted			
		Healthy	Tumor d/30 mm	Tumor d/25 mm	Tumor d/20 mm
Tumor d/20 mm	8016	266	340	463	6947
Tumor d/25 mm	8016	136	244	7392	244
Tumor d/30 mm	8008	261	7223	265	259
Healthy	8008	7085	282	327	134
Σ	32048	7748	8089	8447	7764

Table 5. RF confusion matrix

Actual	Σ	Predicted			
		Healthy	Tumor d/30 mm	Tumor d/25 mm	Tumor d/20 mm
Tumor d/20 mm	8016	493	455	515	6553
Tumor d/25 mm	8016	457	452	6624	483
Tumor d/30 mm	8008	484	6547	482	495
Healthy	8008	6498	488	500	522
Σ	32048	7932	7942	8121	8053

In this study, the effectiveness of the presented model was evaluated using F-measure (F1), accuracy (CA), precision, sensitivity, and MCC, were calculated using in (4)-(8):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (6)$$

$$F - \text{Measure} = \frac{2 * \text{Precision} * \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (7)$$

$$MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP+FP).(TP+FN).(TN+FP).(TN+FN)}} \quad (8)$$

Tables 6-8 demonstrate the capability of the proposed system to distinguish among tumors with diameters of 20 mm, 25 mm, and 30 mm, rather than simply differentiating between healthy and cancerous breast tissues. For the 20 mm tumor class, the highest classification performance was achieved by AdaBoost, which correctly identified 7985 out of 8016 samples, with only 17 samples misclassified as 25 mm tumors and 4 samples as 30 mm tumors. Similar behavior was observed for the 25 mm tumor class, where the majority of samples were correctly classified, while only a small number were confused with neighboring tumor sizes. For the 30 mm tumor class, the CA was the highest among all tumor categories, indicating that larger tumors produce more distinct electromagnetic signatures that are easier to identify.

The confusion matrices also reveal that most classification errors occurred between adjacent tumor sizes (20 mm and 25 mm, or 25 mm and 30 mm), which is expected because tumors with similar dimensions generate comparable perturbations in the measured scattering parameters. In contrast, confusion between the

20 mm and 30 mm tumor classes was minimal, demonstrating that the extracted microwave features effectively capture size-dependent variations in the electromagnetic response. Furthermore, healthy breast tissue was classified with high accuracy across all models, confirming a clear separation between normal and abnormal tissue characteristics.

Table 6. AdaBoost performance metrics

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)
Tumor 20 mm	99.8	99.9	99.6	99.8	99.8
Tumor 25 mm	99.9	99.8	99.8	99.8	99.7
Tumor 30 mm	99.9	99.8	99.9	99.9	99.8
Healthy	99.9	99.7	99.9	99.8	99.8
Overall	99.8	99.8	99.8	99.8	99.8

Table 7. RF performance metrics

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)
Tumor 20 mm	91	81.4	81.8	81.5	74.6
Tumor 25 mm	90.9	81.5	82.6	82.1	75.5
Tumor 30 mm	91.1	82.4	81.8	82.1	75.5
Healthy	91	81.9	81.1	81.5	74.8
Overall	81.8	81.8	81.8	81.8	75.7

Table 8. GB performance metrics

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)
Tumor 20 mm	93.3	89.5	86.7	88.1	82.3
Tumor 25 mm	93.6	87.5	92.2	89.8	85
Tumor 30 mm	94	89.3	90.2	89.7	85.1
Healthy	94.5	91.4	88.4	89.9	86
Overall	89.2	89.2	89.2	89.2	85.6

These findings indicate that the proposed microwave imaging framework is capable not only of detecting the presence of a tumor but also of reliably estimating tumor size, which is an important requirement for accurate diagnosis, disease assessment, and treatment planning. The AdaBoost classifier exhibited the best tumor-size discrimination capability, with recall values of 99.61%, 99.78%, and 99.91% for the 20 mm, 25 mm, and 30 mm tumor classes, respectively. These results confirm that the proposed system can accurately resolve subtle differences in the electromagnetic signatures associated with varying tumor dimensions.

The efficiency of the proposed model can also be evaluated using the area under the curve (AUC). The AUC refers to the area within the receiver operating characteristic (ROC) curve [38], which is a graphical representation that plots the false positive rate (FPR) or specificity on the horizontal axis and the true positive rate (TPR) or sensitivity on the vertical axis. A curve that approaches the upper left corner indicates better performance [39], [40]. Therefore, ROC analysis was performed to assess the discriminative capability of the proposed breast cancer prediction model. As depicted in Figure 7, AdaBoost consistently outperformed the other classifiers across all classification scenarios. For normal breast detection (Figure 7(a)), AdaBoost achieved an AUC of 100%, compared with 98.6% for GB and 97.7% for RF. In the case of detecting tumors with diameters of 20 mm, 25 mm, and 30 mm (Figures 7(b)-(d)), AdaBoost achieved AUC values of 100%, 99.9%, and 99.9%, respectively. In contrast, GB achieved AUC values ranging from 98.0% to 98.6%, while RF produced AUC values between 96.3% and 96.8%. These results show that the proposed framework has good discriminative ability. AdaBoost provides the best accurate and robust classification performance in all tested conditions.

The findings in Table 9 demonstrate that AdaBoost outperforms other methods, with the AUC of 1.00, indicating the effectiveness of AdaBoost in diagnosing breast tumors. In addition, it achieved 99.8% in terms of accuracy, MCC, F1-measure, recall, and precision. This result shows that this method is suitable for applications involving medical diagnosis. On the other hand, GB demonstrates a good performance, with an accuracy of 89.2% and an AUC of 0.984. The RF model has the lowest accuracy 81.8% and AUC (0.966). Based on these measures, RF has difficulty detecting breast tumors compared to other presented methods.

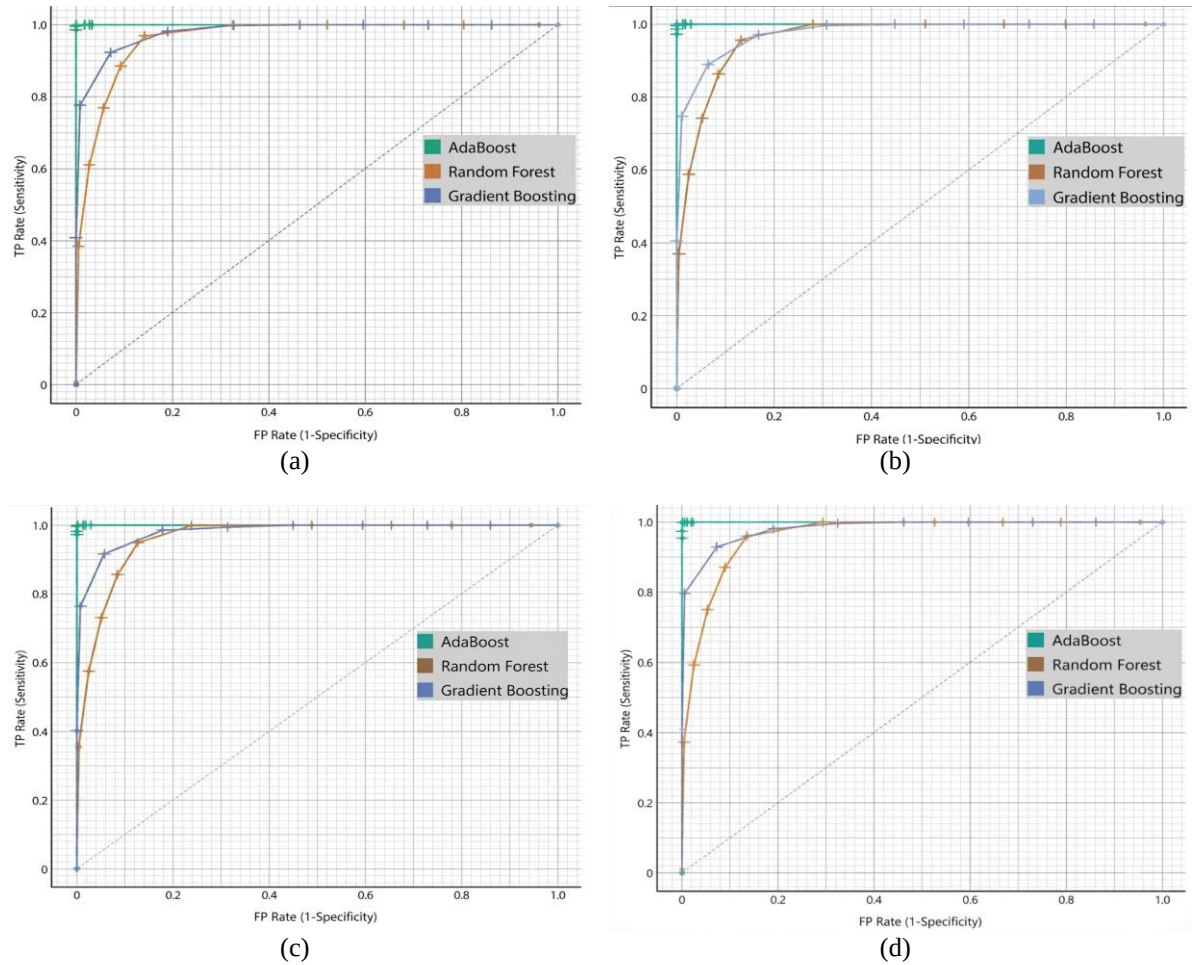


Figure 7. ROC analysis for the proposed model; (a) ROC analysis to detect a breast without a tumor, (b) ROC analysis to detect a breast with a tumor of d20 mm, (c) ROC analysis to detect a tumor of d25 mm, and (d) ROC analysis to detect a tumor of d30 mm

Table 9. Classification performance metrics for the evaluated models

Model	AUC	CA (%)	F1-measure (%)	Precision (%)	Recall (%)	MCC (%)
AdaBoost	1.00	99.8	99.8	99.8	99.8	99.8
RF	0.966	81.8	81.8	81.8	81.8	75.7
GB	0.984	89.2	89.2	89.2	89.2	85.6

4. DISCUSSION

Early detection of breast cancer is vital for reducing mortality rates and the need for mastectomy. This can be achieved through technologies that combine sensor systems, signal processing, and artificial intelligence and ML algorithms. In this work three main classifiers were used: AdaBoost, GB, and RF. As shown in Table 9, AdaBoost achieved the highest classification performance among the evaluated models. This result can be attributed to the characteristics of the antenna-generated dataset, which consists of high-dimensional features extracted from the frequency-dependent S_{11} and S_{21} responses. The presence of a tumor introduces subtle variations in the scattering parameters due to the dielectric contrast between healthy and cancerous tissues. These variations are distributed across the frequency spectrum and may be difficult to capture using conventional ensemble approaches. AdaBoost iteratively increases the influence of misclassified samples, allowing the classifier to focus on the most informative electromagnetic patterns and enhance class separation.

In contrast, RF constructs DTs from randomly selected feature subsets, which may reduce its sensitivity to localized frequency-dependent variations. GB also benefits from sequential learning; however, its performance indicates that the adaptive sample-weighting strategy of AdaBoost is better suited to the structure of the extracted S-parameter features. Consequently, AdaBoost was able to exploit the

discriminative information embedded within the antenna responses more effectively, resulting in superior classification performance across all evaluation metrics. Several recent studies have utilized ML models for the diagnosis of breast cancer. To verify the efficiency and effectiveness of the innovative system presented in this work, and its contribution to the field of breast cancer detection and tumor size prediction, it is necessary to compare the proposed system with similar systems found in the literature. Table 10 illustrates the superior performance of the proposed system, achieving an accuracy of 99.8%. This confirms the effectiveness of the developed system for breast cancer detection and prediction, while the accuracy of other models found in the literature ranged between 69% and 99.3%.

Table 10. A comparison between the suggested model and earlier research

Study/year	Methods/model	Performance metric	Max. percentage (%)
Elsheakh <i>et al.</i> [1] (2023)	DT, RF, and CatBoost	Accuracy	69
Naji <i>et al.</i> [2] (2021)	SVM, RF, LR, DT, and KNN	Accuracy	97.2
Chaurasia and Pal [3] (2020)	CART, SVM, NB, KNN, MLP, and LR	Accuracy	95.1691
Islam <i>et al.</i> [4] (2020)	SVM, KNN, RF, ANN, and LG	Accuracy	98.57
Tiwari <i>et al.</i> [5] (2020)	ANN, CNN, SVM, and RF	Accuracy	99.3
Srivenkatesh [7] (2020)	RF, SVM, KNN, NB, and LR	Accuracy	98.24
Delen <i>et al.</i> [14] (2005)	ANN and LR	Accuracy	91.2
Proposed	AdaBoost, GB, and RF	Accuracy	99.8

5. CONCLUSION

The experimental results demonstrated that applying artificial intelligence and ML techniques to analyze and classify the S-parameters of the antenna sensor model designed in this study for breast cancer prediction achieves high accuracy. Three main classifiers were used: AdaBoost, GB, and RF. The findings revealed that the AdaBoost classifier outperformed the others, achieving an accuracy of 99.8% in detecting and classifying breast cancer into three size categories: 20 mm, 25 mm, and 30 mm. This makes the proposed system suitable for practical use.

Future studies could concentrate on increasing the antenna's efficiency, investigating different approaches for extracting new features to facilitate the detection of smaller tumors and improve the accuracy of tumor location predictions within the breast tissue, and assessing other classification algorithms to boost the effectiveness of the breast cancer detection system. It is important to acknowledge some limitations of our study, while the proposed approach attained an accuracy of 99.8% and an AUC of 1.00, the possibility of overfitting cannot be completely excluded. The developed model was trained and evaluated using simulation-generated data, which may not fully represent the complexity and variability encountered in practical measurement environments. Real measured data are often affected by instrumentation errors, noise, calibration uncertainties, breast tissue heterogeneity, and patient-to-patient anatomical variations. As a result, the performance obtained from simulation-based datasets may be higher than that achievable in real clinical applications.

Although the results demonstrate the effectiveness of the presented methodology under controlled conditions, several limitations should be acknowledged. First, the tumor location and depth were fixed throughout all simulations to ensure a consistent evaluation environment. Second, the dataset was generated entirely from electromagnetic simulations, and therefore may not fully capture the variability and noise encountered in real measurement scenarios. Third, the breast model does not account for patient-specific anatomical variations, such as differences in breast size, tissue composition, and dielectric-property distributions. Consequently, further validation using experimentally acquired and clinically relevant datasets is essential to verify the robustness and generalization capability of the proposed approach. Future work will focus on investigating multiple tumor locations and depths, incorporating more realistic heterogeneous breast models, and evaluating the proposed system using experimentally measured data and patient-specific breast phantoms under diverse operating conditions.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Muneera Altayeb	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	
Sanaa Salama	✓	✓		✓	✓	✓	✓	✓	✓			✓		
Duaa Zyoud	✓		✓	✓			✓			✓	✓		✓	
Ashraf Abuelhaija						✓		✓		✓				

C : **C**onceptualizationM : **M**ethodologySo : **S**oftwareVa : **V**alidationFo : **F**ormal analysisI : **I**nterpretationR : **R**esourcesD : **D**ata CurationO : Writing - **O**riginal DraftE : Writing - Review & **E**ditngVi : **V**isualizationSu : **S**upervisionP : **P**roject administrationFu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author.

REFERENCES

- [1] D. N. Elsheakh, R. A. Mohamed, O. M. Fahmy, K. Ezzat, and A. R. Eldamak, "Complete Breast Cancer Detection and Monitoring System by Using Microwave Textile Based Antenna Sensors," *Biosensors*, vol. 13, no. 1, p. 87, Jan. 2023, doi: 10.3390/bios13010087.
- [2] M. A. Naji, S. E. Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouahid, and O. Debauche, "Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis," *Procedia Computer Science*, vol. 191, pp. 487–492, 2021, doi: 10.1016/j.procs.2021.07.062.
- [3] V. Chaurasia and S. Pal, "Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer," *SN Computer Science*, vol. 1, no. 5, 2020, doi: 10.1007/s42979-020-00296-8.
- [4] 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 Computer Science*, vol. 1, no. 5, 2020, doi: 10.1007/s42979-020-00305-w.
- [5] M. Tiwari, R. Bharuka, P. Shah, and R. Lokare, "Breast Cancer Prediction Using Deep Learning and Machine Learning Techniques," *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3558786.
- [6] V. Nemade and V. Fegade, "Machine Learning Techniques for Breast Cancer Prediction," *Procedia Computer Science*, vol. 218, pp. 1314–1320, 2023, doi: 10.1016/j.procs.2023.01.110.
- [7] D. M. Srivenkatesh, "Prediction of Breast Cancer Disease using Machine Learning Algorithms," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 4, pp. 2868–2878, 2020, doi: 10.35940/ijitee.d1866.029420.
- [8] R. Rabiei, S. M. Ayyoubzadeh, S. Sohrabei, M. Esmaeili, and A. Atashi, "Prediction of Breast Cancer using Machine Learning Approaches," *Journal of Biomedical Physics and Engineering*, vol. 12, no. 3, pp. 297–308, 2022, doi: 10.31661/jbpe.v0i0.2109-1403.
- [9] Ch. Shravya, K. Pravalika, and S. Subhani, "Prediction of Breast Cancer Using Supervised Machine Learning Techniques," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 6, pp. 1106–1110, 2019.
- [10] T. R. Mahesh, V. V. Kumar, V. Vivek, K. M. K. Raghunath, and G. S. Madhuri, "Early predictive model for breast cancer classification using blended ensemble learning," *International Journal of System Assurance Engineering and Management*, vol. 15, no. 1, pp. 188–197, 2024, doi: 10.1007/s13198-022-01696-0.
- [11] V. A. Telsang and K. Hegde, "Breast Cancer Prediction Analysis using Machine Learning Algorithms," in *2020 International Conference on Communication, Computing and Industry 4.0 (C2I4)*, Dec. 2020, pp. 1–5, doi: 10.1109/C2I451079.2020.9368911.
- [12] M. K. Keleş, "Breast cancer prediction and detection using data mining classification algorithms: A comparative study," *Tehnicki Vjesnik*, vol. 26, no. 1, pp. 149–155, 2019, doi: 10.17559/TV-20180417102943.
- [13] B. Akbugday, "Classification of breast cancer data using machine learning Algorithms," *TIPTEKNO 2019 - Tip Teknolojileri Kongresi*, 2019, doi: 10.1109/TIPTEKNO.2019.8895222.
- [14] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: A comparison of three data mining methods," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 113–127, 2005, doi: 10.1016/j.artmed.2004.07.002.
- [15] A. S. Assiri, S. Nazir, and S. A. Velastin, "Breast Tumor Classification Using an Ensemble Machine Learning Method," *Journal of Imaging*, vol. 6, no. 6, 2020, doi: 10.3390/JIMAGING6060039.
- [16] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," *Procedia Computer Science*, vol. 83, pp. 1064–1069, 2016, doi: 10.1016/j.procs.2016.04.224.
- [17] S. Salama, D. Zyoud, A. Abuelhaija, M. Altayeb, and A. Al-Bassam, "Breast Tumor Detection via S-Parameter Contrast Using a 1 × 8 Miniaturized Metamaterial Antenna Array for UWB Microwave Imaging," *International Journal of Imaging Systems and Technology*, vol. 36, no. 1, 2026, doi: 10.1002/ima.70295.
- [18] N. Hammouch, A. Rghioui, H. Ammor, M. Oubre, and J. Lloret, "A low-cost UWB microwave imaging system for early-stage breast cancer detection," *Multimedia Tools and Applications*, vol. 84, no. 17, pp. 17329–17360, Jul. 2025, doi: 10.1007/s11042-024-19761-0.
- [19] S. Gabriel, R. W. Lau, and C. Gabriel, "The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz," *Physics in Medicine and Biology*, vol. 41, no. 11, pp. 2251–2269, Nov. 1996, doi: 10.1088/0031-9155/41/11/002.

- [20] M. Lazebnik *et al.*, "A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries," *Physics in Medicine and Biology*, vol. 52, no. 20, pp. 6093–6115, Oct. 2007, doi: 10.1088/0031-9155/52/20/002.
- [21] U. Rafique, S. Pisa, R. Cicchetti, O. Testa, and M. Cagnano, "Ultra-Wideband Antennas for Biomedical Imaging Applications: A Survey," *Sensors*, vol. 22, no. 9, 2022, doi: 10.3390/s22093230.
- [22] S. Gamil, F. Zeng, M. Alrifay, M. Asim, and N. Ahmad, "An Efficient AdaBoost Algorithm for Enhancing Skin Cancer Detection and Classification," *Algorithms*, vol. 17, no. 8, 2024, doi: 10.3390/a17080353.
- [23] R. Appel, T. Fuchs, P. Dollár, and P. Perona, "Quickly Boosting Decision Trees—Pruning Underachieving Features Early," in *Proceedings of the 30th International Conference on Machine Learning*, 2013, vol. 28, no. 3, pp. 594–602.
- [24] O. Almomani, M. A. Almaiah, A. Alsaaidah, S. Smadi, A. H. Mohammad, and A. Althunibat, "Machine learning classifiers for Network Intrusion Detection System: Comparative Study," in *2021 International Conference on Information Technology (ICIT)*, Amman, Jordan, 2021, pp. 440–445, doi: 10.1109/icit52682.2021.9491770.
- [25] C. Wang, S. Xu, and J. Yang, "Adaboost algorithm in artificial intelligence for optimizing the IRI prediction accuracy of asphalt concrete pavement," *Sensors*, vol. 21, no. 17, 2021, doi: 10.3390/s21175682.
- [26] E. Ghorbani and S. Yagiz, "Estimating the penetration rate of tunnel boring machines via gradient boosting algorithms," *Engineering Applications of Artificial Intelligence*, vol. 136, 2024, doi: 10.1016/j.engappai.2024.108985.
- [27] H. Cao, M. N. Nguyen, C. Phua, S. Krishnaswamy, and X. L. Li, "An integrated framework for human activity recognition," in *UbiComp'12-Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 621–622, 2012, doi: 10.1145/2370216.2370334.
- [28] O. Almomani, A. Alsaaidah, M. A. Almaiah, A. Alzaqebah, M. M. Abualhaj, and W. J. Alzyadat, "Evaluating Machine Learning Classifiers for Detecting Distributed Denial of Service Attacks," in *2025 12th International Conference on Information Technology (ICIT)*, Amman, Jordan, 2025, pp. 134–140, doi: 10.1109/ICIT64950.2025.11049130.
- [29] C. Avci, M. Budak, N. Yagmur, and F. B. Balcik, "Comparison between random forest and support vector machine algorithms for LULC classification," *International Journal of Engineering and Geosciences*, vol. 8, no. 1, pp. 1–10, 2023, doi: 10.26833/ijeg.987605.
- [30] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random Forests," in *Ensemble Machine Learning*, New York, NY: Springer New York, 2012, pp. 157–175, doi: 10.1007/978-1-4419-9326-7_5.
- [31] S. S. Hussain and S. S. H. Zaidi, "AdaBoost Ensemble Approach with Weak Classifiers for Gear Fault Diagnosis and Prognosis in DC Motors," *Applied Sciences (Switzerland)*, vol. 14, no. 7, 2024, doi: 10.3390/app14073105.
- [32] V. Jagannath, "Random Forest Template for TIBCO Spotfire® - Wiki page | TIBCO Community," 2017, [Online]. Available: <https://community.spotfire.com/articles/spotfire/random-forest-template-for-spotfire/>. (Accessed Sep. 16, 2026).
- [33] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, pp. 69–79, 2024, doi: 10.58496/bjml/2024/007.
- [34] B. Kalantar, B. Pradhan, S. A. Naghibi, A. Motevalli, and S. Mansor, "Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN)," *Geomatics, Natural Hazards and Risk*, vol. 9, no. 1, pp. 49–69, 2018, doi: 10.1080/19475705.2017.1407368.
- [35] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Mining*, vol. 16, no. 1, 2023, doi: 10.1186/s13040-023-00322-4.
- [36] H. A. Owida, J. Al-Nabulsi, N. Al Hawamdeh, S. Abuowaida, Z. Salah, and E. A. Elsoud, "Deep Learning-Based Kidney Disease Classification Using ResNet50: Enhancing Diagnostic Accuracy," in *2024 2nd Jordanian International Biomedical Engineering Conference, JIBEC 2024*, 2024, pp. 126–129, doi: 10.1109/JIBEC63210.2024.10931915.
- [37] A. Al-Ghraibah and M. Al-Ayyad, "Automated detection of leukemia in blood microscopic images using image processing techniques and unique features: Cell count and area ratio," *Cogent Engineering*, vol. 11, no. 1, Jan. 2024, doi: 10.1080/23311916.2024.2304484.
- [38] Y. Liu, J. Cheng, C. Yan, X. Wu, and F. Chen, "Research on the Matthews correlation coefficients metrics of personalized recommendation algorithm evaluation," *International Journal of Hybrid Information Technology*, vol. 8, no. 1, pp. 163–172, 2015, doi: 10.14257/ijhit.2015.8.1.14.
- [39] A. M. Arabiat, "Intelligent Model for Detecting GAN-Generated Images Based on Multi-Classifiers and Advanced Data Mining Techniques," *International Journal of Electrical and Electronic Engineering and Telecommunications*, vol. 14, no. 3, pp. 147–157, 2025, doi: 10.18178/ijeetc.14.3.147-157.
- [40] K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Caspian Journal of Internal Medicine*, vol. 4, no. 2, pp. 627–635, 2013.

BIOGRAPHIES OF AUTHORS



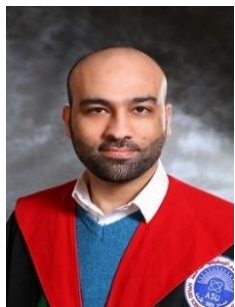
Muneera Altayeb    earned her B.Sc. in Computer Engineering in 2007, and M.Sc. in Communications Engineering from the University of Jordan in 2010. She is currently Assistant Dean of the Faculty of Engineering/Al-Ahliyya Amman University (AAU). She has been a lecturer in the Department of Communications and Computer Engineering at (AAU) since 2015. Her research interests are focused on the following areas: digital signals and image processing, machine learning, robotics, digital electronics, and semiconductors. She can be contacted at email: m.altayeb@ammanu.edu.jo.



Sanaa Salama    earned her B.Sc. in Telecommunication Technology from Arab American University (2006), an M.Sc. in Electrical Engineering from the University of Jordan (2009), and a Ph.D. in Electrical Engineering from the University of Duisburg-Essen, Germany (2015), all with excellent academic performance. Her research focuses on advanced antenna systems, including MIMO, beam-forming arrays, reconfigurable and implantable antennas, mutual coupling, and RF coil design for 7T MRI systems. She currently leads a research group developing compact, dual-band, high-gain implantable antennas for MICS and ISM bands. She is an Associate Professor in the Department of Medical Equipment Engineering at AAUP, with numerous journal and conference publications. She can be contacted at email: Sanaa.salama@aaup.edu.



Duaa Zyoud    received the B.Sc. degree in Telecommunication Engineering from the Arab American University (AAUP), Jenin, Palestine, in 2020. Her research interest is in antenna design for wireless communications and biomedical applications. Recently, her field of research is design of implantable antennas for biomedical applications, mainly MICS and ISM frequency bands. She can be contacted at email: duaakzyoud@gmail.com.



Dr. Ashraf Abuelhaija    received the B.Sc. degree in communications and electronics engineering from Applied Science Private University, Amman, Jordan, in 2007, and the master's and Ph.D. degrees in electrical engineering from the University of Duisburg-Essen in 2010 and 2016, respectively. From 2012 to 2014, he was a Research Assistant with the Erwin L. Hahn Institute for MRI, Essen. From 2014 to 2016, he was a research assistant with the Department of Microwave and RF Technology, University of Duisburg-Essen. He is currently an associate professor with the Department of Electrical Engineering, Applied Science Private University. His research interests are in the area of antennas and RF technology. He can be contacted at email: a_abualhijaa@asu.edu.jo.