PIER C

(Received 26 March 2025, Accepted 3 June 2025, Scheduled 24 June 2025)

DATDNet: A Deep Neural Network for Breast Tumor Microwave Detection under Varying Breast Morphologies

Min Lu[1, 2, *], Xia Xiao[2, 3], Lei Yuan[1], Han Su[1], Xiaomin He[1], and Yuan Yang[1]

[1] *School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China*
[2] *Tianjin Key Laboratory of Imaging and Sensing Microelectronic Technology*
School of Microelectronics, Tianjin University, Tianjin 300072, China
[3] *Shanghai Artificial Intelligence Laboratory, Shanghai, China*

ABSTRACT: ***Object:*** The varying breast morphologies can lead to enormous differences in microwave backscatter signals, making it difficult to identify weak tumor responses, which adversely affects the performance of microwave detection. Existing deep learning methods for microwave tumor detection struggle to generalize diverse breast morphologies. The purpose of this study is to develop a deep learning method to overcome the influence of breast morphology on microwave tumor detection. ***Methods:*** This paper proposes a domain-adversarial tumor detection network (DATDNet) to improve detection performance. The proposed method employs breast backscatter signals with known tumor information as source domain data for training a convolution neural network. Subsequently, deep adversarial training is conducted on the backscatter signals of breasts with unseen morphologies and unknown tumor information in the trained network, in order to mitigate the adverse effects of variations in breast morphology on detection. In the process of microwave breast image feature extraction, our method introduces channel and spatial attention mechanisms in the convolution modules to pay more attention to tumor information. ***Results:*** The feature distribution estimations demonstrate that the microwave data from different breast morphologies are effectively aligned. In two datasets with completely different breast morphologies, the detection accuracies reach 76.64% and 83.15%, with improvements of 5.36% and 7.79% compared with baseline CNN. The ablation studies demonstrate that the proposed method effectively enhances the generalization performance and accuracy of microwave breast cancer detection.

1. INTRODUCTION

Breast cancer is a malignant disease with the highest incidence among women, and its risk will further increase amidst the advancements in globalization and economic growth [1]. Ultra-wideband microwave technology has garnered significant attention of doctors and researchers, owing to its remarkable advantages such as low electromagnetic radiation, large information-carrying capacity, and cost-effectiveness. It is expected to be developed into a safe, accurate, and economical approach for breast tumor detection [2]. In this article, ultra-wideband microwave technology is also known as microwave technology. The physics foundation of ultra-wideband microwave detection of breast tumors lies in the contrast in dielectric properties between normal (fat) and malignant (tumor) tissues at microwave frequencies. This difference is attributed to the varying water contents within these tissues. Through the measurement of excised breast tissues, it was found that the relative permittivities of fat and tumor tissues are approximately 7 and 60 at the intermediate frequency in the 0.5–20 GHz frequency range [3]. If there is a tumor in the breast, the incident wave is scattered, affecting the electromagnetic wave energy at the receiving antenna. The tumor information can be obtained by processing the changed backscatter signal [4].

The research on microwave breast tumor detection algorithm is a vital area in this field, directly impacting the effectiveness and accuracy of this technology. Initially, researchers employed microwave image reconstruction algorithms to detect the presence of breast tumors. In confocal imaging [5–10], only the tumor region is reconstructed by extracting tumor information and estimating the average wave velocity. Microwave tomography [11–13] reconstructs the overall dielectric contrast map of the breast by solving complex inverse problems and locates tumors within the breast by identifying high dielectric regions. Owing to the complexity of breast structural morphology, traditional microwave imaging methods often exhibit obvious complexity. In recent years, the intelligent detection [14–17] of breast tumors has been intensively studied, thanks to artificial intelligence technology's nonlinear learning ability and representation ability. Artificial intelligence-based detection methods can apply rich prior knowledge and automatically learn predictive models between breast scattering signals and tumor related information from a large amount of data.

Martins et al. [18] integrated antenna topology and machine learning algorithms, such as the K-nearest neighbor algorithm, linear discriminant analysis, and support vector machine, achieving an accuracy of 85% on five breast phantoms. Ref. [19] utilized 150 scans of 10 breast phantoms to detect the presence of tumors using linear discrimination analysis and support vector machine. Different time-frequency decomposi-

* Corresponding author: Min Lu (minlu@xaut.edu.cn).

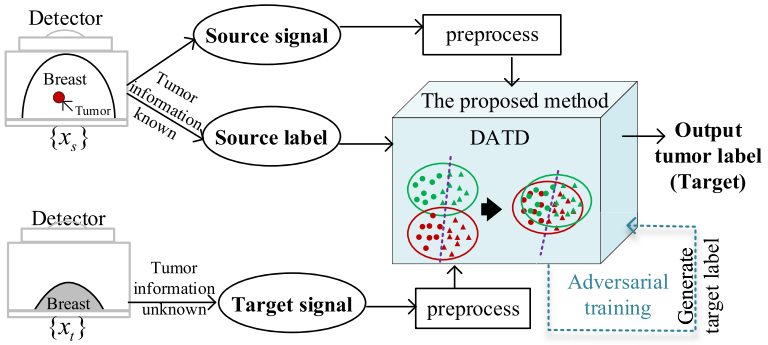

FIGURE 1. The diagram of the proposed DATDNet. The signal data $\{x_s\}$ represent the data with known tumor information (source domain). The signal $\{x_t\}$ represent the data with unknown tumor information (target domain). The proposed DATDNet can generate the labels (tumor information) of the target domain through adversarial training.

tion methods for tumor detection were compared using a dataset comprising 12 healthy breast backscatter signals and simulated tumor responses [20]. Francescini et al. [21] utilized a random shape generator to generate many 2-D breast models and implemented a breast microwave tomography method based on neural network. However, the detection of tumors requires setting a threshold. Rena et al. [22] collected microwave S parameter data from 61 patients and implemented tumor detection using the supervised machine learning method combining principal component analysis (PCA) and support vector machine (SVM). The time-frequency feature extraction and genetic neural network were employed to achieve accurate tumor localization for a single heterogeneous breast model [23]. Reimer and Pistorius [24] achieved acceptable tumor detection performance using a convolutional neural network (CNN) method, where the extreme breast volumes in the test set (excluding oversized and undersized phantoms) were constrained within the training set. The variant structures of CNN architectures (CNN-long short-term memory (LSTM), ResNet) [4, 25] have also been used to detect microwave breast tumors, demonstrating the great potential of deep learning in this field. Khalid et al. [26] built a BreastCare dataset to simulate breast density changes by filling a specific breast model with a mixture of different dielectric properties. Based on this database, ResNet was used to achieve good tumor detection accuracy. In addition to the tumor detection, the identification of tumor characteristics [27–29] has also made remarkable progress. In breast modeling, electromagnetic modeling methods such as the finite difference time domain (FDTD) [30, 31] or the Method of Moments (MoM) [32] are commonly used to simulate electromagnetic wave propagation processes and obtain simulated signals. Additionally, some studies have also obtained signals by building microwave experimental platforms [33].

While microwave methods in conjunction with deep learning have shown good potential and excellent outcomes, some key issues remain unresolved. However, there is often a lack of data diversity in breast data sources for developing breast tumor detection algorithms, and thus, datasets are susceptible to encountering similar or even identical breast backscatter data during training and testing phases. Although the training network model adapts well to the specific sample data, the similarity between the data makes the evaluation results overly optimistic [34]. These methods cannot effectively solve broader, more general detection scenarios. The variations between breasts challenge the effectiveness of detection algorithms, and when algorithms encounter breast data they have not seen during training, their accuracy decreases significantly. This situation is highly unfavorable for future clinical applications. This article describes efforts to enhance the generalization performance of microwave sensing in breast tumor detection. The key contributions of this work are summarized as follows:

(1) This paper proposes a neural network architecture for microwave tumor detection based on the theory of domain adversarial learning, aiming to address the low generalization performance caused by different breast morphologies. The effective alignment of the different breast backscatter signals is confirmed through feature distribution estimation.
(2) To pay more attention to the weak tumor response during network learning, our method integrates the attention mechanism into the current network framework. The ablation experiments demonstrate that this strategy further improves the accuracy of tumor detection.
(3) We utilized an experimental dataset of breast scattering signals, which features the most diverse breast morphology (including 1,008 unique breast scans), for algorithm validation. In two completely different tumor detection tasks, our method improves the accuracy by 5.36% and 7.79% compared to the CNN baseline.

The remainder of this paper is organized as follows. Section 2 explains the proposed deep transfer learning method and describes the dataset used in this study. Section 3 reports the performance evaluation of the proposed method. Section 4 provides a comprehensive discussion of this article and related work. Section 5 gives a conclusion of this study and future works.

2. METHODS

2.1. Overall Network Architecture

Figure 1 illustrates the diagram of the proposed framework. In this case, the source domain and target domain represent the

Algorithm: Training and testing process of the proposed DATDNet

Require: Feature extractor, tumor detector, domain discriminator.

1: **Input:** The labeled source data $D_s = \{(x_{s,i}, y_{s,i})\}_{i=1}^{n_s}$, the unlabeled target data $D_t = \{(x_{t,i})\}_{i=1}^{n_t}$

Training:

2: for i = 1 to n

3: for j = 1 to batches (batches = n/batch size)

Forward propagation- tumor detector

4: $G_y(G_f(x_i)), x_i \in x_s$. L_y is calculated using Eq.(10).

Forward propagation- domain discriminator

5: $G_d(G_f(x_i)), x_i \in x_D$. L_d is calculated using Eq.(11).

6: Wherein, $G_f(x_i) = \max[drop_p(M_s(M_c(x_{D,i}^{con}) \otimes x_{D,i}^{con}) \otimes (M_c(x_{D,i}^{con}) \otimes x_{D,i}^{con}))]$.

Backward propagation-finding the optimal network parameters

7: $L(\theta_f^*, \theta_y^*, \theta_d^*) = \min_{\theta_f, \theta_y, \theta_d} L_y(\theta_f, \theta_y) - \lambda L_d(\theta_f, \theta_d)$.

8: $(\hat{\theta}_f, \hat{\theta}_y) = \arg\min_{\theta_f, \theta_y} E(\theta_f, \theta_y, \hat{\theta}_d)$,

9: $\hat{\theta}_d = \arg\max_{\theta_d} E(\hat{\theta}_f, \hat{\theta}_y, \theta_d)$.

10: end

11: end

Output: The parameters $(\hat{\theta}_f, \hat{\theta}_y, \hat{\theta}_d)$ of the well-trained DATDNet.

Testing: $\hat{y} = G_y(G_f(x_i)), x_i \in x_t$. Use the trained DATDNet to calculate tumor predictive values for target domain data without tumor labels.

backscatter signal domains of models with different breast morphologies, respectively. The two data domains within domain adversarial [35] framework exhibit varying marginal distributions. During the process of network learning, the signal data $\{x_s\}$ in the source domain are labeled with tumor information, whereas the tumor information of the signal $\{x_t\}$ in the target domain remains unknown. The label information of the target domain is solely utilized for the final performance evaluation.

The proposed domain adversarial tumor detection network (DATDNet) generates the labels of the target domain and conducts adversarial training, so that the data feature distributions of the source domain and target domain tend to be consistent. Inspired by the successful applications [36] of attention mechanisms in computer vision and natural language processing, we expect that attention mechanisms can selectively focus on more critical information in S-parameter images for tumor detection. The training and testing process of the proposed DATDNet is summarized as follows.

This architecture consists of three parts: a feature extractor $G_f(x_i; \theta_f)$ which extracts features from the source domain and target domain, a domain discriminator $G_d(x_i; \theta_d)$ which determines whether the feature comes from the source domain or target domain, and a tumor detector $G_y(x_i; \theta_y)$ which classifies the features. Symbols x and θ represent the samples and network parameters. The input of the network is the labeled samples from source domain and unlabeled samples from target domain. Assuming the source domain $D_s = \{x_s, y_s\}$ and the target domain $D_t = \{x_t\}$, the loss function of the label classifier and the domain discriminator are written as follows,

$$L_y^i(\theta_f, \theta_y) = L_y\left(G_y(G_f(x_i; \theta_f); \theta_y), y_i\right), \quad (1)$$

$$L_d^i(\theta_f, \theta_d) = L_d\left(G_d(G_f(x_i; \theta_f); \theta_d), d_i\right). \quad (2)$$

Parameters y_i and d_i are the ith sample category label and domain category label, respectively. In the training process, the first task is to minimize the classification error of the source domain dataset, and the second task is to maximize domain classification error to confuse the source domain data with the target domain data. Fig. 2 shows the overall structure of the proposed method. The specific process of the proposed algorithm is described.

2.2. Microwave Image Feature Extractor

The proposed method consists of several feature extraction modules, and each module consists of a convolution layer, a max-pooling layer, a dropout layer, and an attention mechanism module. The data space formed by all signal samples in the source domain and target domain is denoted as $x_D = \{x_s, x_t\}$. The calculation process of the convolution module for the i sample can be expressed as,

$$x_{D,i}^{con} = \phi\left[BN_{\gamma,\beta}\left(x_{D,i} * \omega^{con} + b^{con}\right)\right], \quad (3)$$

where $x_{D,i}^{con}$ represents the output of the convolution module, and ω^{con} and b^{con} denote the weighted parameters of the convolution kernel and bias vector. Parameter $BN_{\gamma,\beta}$ represents the BatchNormalization (BN) operation with scaling parameter γ and translation parameter β. Parameter $\phi(\cdot)$ represents the ReLU (Rectified Linear Unit) layer. The operation of BN ensures that the input values for each training iteration remain within a reasonable range, which can improve training speed and accelerate the convergence of the network. The ReLU function strengthens the nonlinear capabilities of neural networks and effectively alleviates the vanishing gradient phenomenon. Furthermore, the attention mechanism module [37, 38] is introduced at both the channel and spatial levels, which can enhance the feature expression capability of convo-

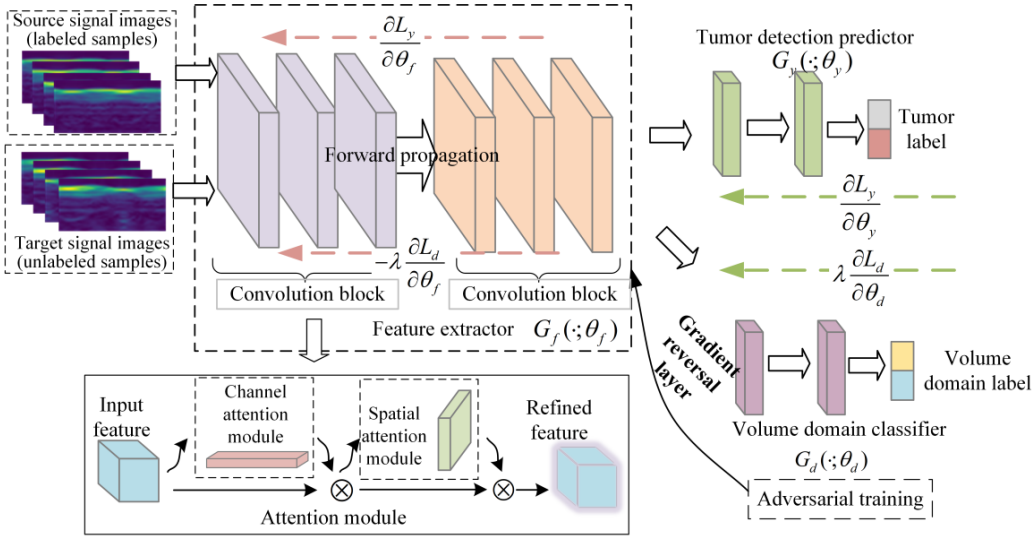

FIGURE 2. The model structure of the proposed DATDNet. The symbol $\frac{\partial(\cdot)}{\partial(\cdot)}$ represents the derivative operation in the backpropagation process of errors. The symbols $G_f(\cdot;\theta_f), G_y(\cdot;\theta_y)$, and $G_d(\cdot;\theta_d)$ represent the feature extractor, tumor detection predictor, and domain classifier with parameters θ_f, θ_y, and θ_d, respectively. This architecture consists of three parts: a feature extractor, a domain classifier, and a tumor detection predictor. The attention module has two sequential sub-modules: a channel and a spatial attention module.

lution neural network.

$$F''_{s,i} = M_s(F'_{c,i}) \otimes F'_{c,i}, \tag{4}$$

$$F'_{c,i} = M_c(x^{con}_{D,i}) \otimes x^{con}_{D,i}, \tag{5}$$

where $x^{con}_{D,i} \in \mathbb{R}^{C\times H\times W}$ is the input feature map, and $M_c \in \mathbb{R}^{C\times 1\times 1}$ and $M_S \in \mathbb{R}^{1\times H\times W}$ represent the operation of attention extraction on the channel and spatial dimension, respectively. The input feature maps are spatially aggregated using max-pooling and average-pooling operations, generating two 1D vectors ($F^c_{\max}$ and F^c_{avg}). The output is fed into a multilayer perceptron with the compression rate r, and the final result is obtained through an activation function $\sigma(\cdot)$. The channel attention mechanism is mathematically expressed as,

$$\begin{aligned} M_c(x^{con}_{D,i}) &= \sigma\left(MLP(AvgPool(x^{con}_{D,i})\right) \\ &\quad + MLP\left(MaxPool(x^{con}_{D,i})\right) \\ &= \sigma\left(W_1(W_0(F^c_{avg,i})) + W_1(W_0(F^c_{\max,i}))\right), \end{aligned} \tag{6}$$

where $W_0 \in \mathbb{R}^{C/r\times C}$ and $W_1 \in \mathbb{R}^{C\times C/r}$. The final result is $M_c \in \mathbb{R}^{C\times 1\times 1}$. Average-pooling operation captures information from every pixel on the feature map, while max-pooling provides feedback on the regions with the most obvious response. Similar to the channel attention module, the spatial attention module performs average-pooling and max-pooling on the $M_c(F)$ along the channel dimension. Then, the convolution operation is applied to the 2-channel feature maps to generate a 2D feature map. This process is expressed as,

$$\begin{aligned} M_s(x^{con}_{D,i}) &= \sigma\left(f^{m\times m}([AvgPool(x^{con}_{D,i}); MaxPool(x^{con}_{D,i})])\right) \\ &= \sigma\left(f^{m\times m}([F^s_{avg,i}; F^s_{\max,i}])\right), \end{aligned} \tag{7}$$

where $f^{m\times m}$ denotes the convolution operation with the convolution size of $m \times m$.

In order to prevent overfitting of the network, we employed a dropout layer, and network parameters are randomly discarded with a certain probability p in each training batch. Finally, a max-pooling layer is added to reduce the number of model parameters. The output of the feature extractor for x_i is expressed as,

$$x^f_{D,i} = \max\left[drop_p(F''_{s,i})\right]. \tag{8}$$

2.3. Breast Tumor Detector

The output of the feature extractor includes the feature information from source and target domains $x^f_{D,i} \in \{x^f_{s,i}, x^f_{t,i}\}$. The labeled features $x^f_{s,i}$ from the source domain are input to the breast tumor predictor or domain classifier. The unlabeled features $x^f_{t,i}$ from the target domain are fed to the domain classifier. Softmax is used as the activation function for the output layer. The process of the tumor prediction is shown below,

$$x^y_{s,i} = Softmax\left(\omega^{fc} * x^f_{s,i} + b^{fc}\right), \tag{9}$$

where $\{\omega^{fc}, b^{fc}\}$ represents the parameters of the fully connected (FC) layer. The cross-entropy loss function of the tumor predictor is as follows,

$$L_y = -\frac{1}{n_s}\left[\sum_{i=1}^{n_s}(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i))\right], \tag{10}$$

where n_s denotes the number of samples in source domain. Parameters y and $\hat{y}$ represent the true label and predicted label, respectively.

2.4. Domain Discriminator

The network architecture of the domain discriminator is the same as that of the tumor predictor. Overall, good transferable features should exhibit the traits of domain invariance and discriminability. Its ultimate task is to predict whether the features originate from the source domain or target domain. The optimal parameters are obtained by minimizing the following discriminator loss,

$$L_d = \frac{1}{n_s}\sum_{i=1}^{n_s} L_d^i(x_{s,i}^{fd}) + \frac{1}{n_t}\sum_{i=1}^{n_t} L_d^i(x_{t,i}^{fd}) \tag{11}$$

$$L_d^i = y_i \log \hat{y}_i(x_{D,i}) + (1-y_i)\log\left(1-\hat{y}_i(x_{D,i})\right), \tag{12}$$

where $x_{s,i}^{fd}$ and $x_{t,i}^{fd}$ denote the output of feature $x_{D,i}^{f}$ through the discriminator. Parameters n_s and n_t represent the number of samples in the source and target domains, respectively.

2.5. Adversarial Training

In the adversarial training process, the gradients from the label classifier and domain predictor are subtracted. This subtraction is crucial and can be implemented by introducing a gradient reversal layer (GRL). During the backpropagation process, the GRL changes the sign of the next layer's gradient (multiplied by -1) and then transmits it to the previous layer, without requiring additional parameter updates for this layer. The following loss function is optimized,

$$E(\theta_f,\theta_y,\theta_d) = \frac{1}{n}\sum_{i=1}^{n} L_y^i(\theta_f,\theta_y)$$

$$-\lambda\left(\frac{1}{n}\sum_{i=1}^{n} L_d^i(\theta_f,\theta_d\right) + \frac{1}{n'}\sum_{i=n+1}^{N} L_d^i(\theta_f,\theta_d)), \tag{13}$$

by finding the parameters of the feature extractor, tumor detector, and domain discriminator $\hat{\theta}_f, \hat{\theta}_y, \hat{\theta}_d$ such that,

$$\left(\hat{\theta}_f, \hat{\theta}_y\right) = \arg\min_{\theta_f,\theta_y} E\left(\theta_f,\theta_y,\hat{\theta}_d\right) \tag{14}$$

$$\hat{\theta}_d = \arg\max_{\theta_d} E\left(\hat{\theta}_f,\hat{\theta}_y,\theta_d\right). \tag{15}$$

Therein, the optimal parameters of the network are learned by minimizing the loss of the tumor detector and maximizing the loss of the domain discriminator, while determining domain discriminator parameters by minimizing the discriminator loss. The competition between the tumor detector and domain discriminator drives the adversarial learning process to extract features that are both domain-invariant and discriminative for the task of breast tumor detection under varying breast morphologies.

2.6. Data Description and Preparation

The experimental dataset UM-BMID [24, 33] for microwave breast tumor detection is used to evaluate the proposed method. The UM-BMID dataset exhibits a high degree of breast morphology variability. Fig. 3 shows the microwave data acquisition mode. This dataset (the second generation) includes 1008 scans from 66 unique 3D printed breast phantoms using preclinical BMI measurement system [39]. The experimental system consists of a Vector Network Analyzer (VNA) and two horn antennas placed 60° apart using a bistatic rotating scanning mode. Scattering parameters are measured at 72 equally spaced antenna positions on a 360° circular trajectory around the breast phantom. The dimension of the signal image is $35 * 72$, where 35 represents the number of time sampling points, and 72 represents the number of antenna positions.

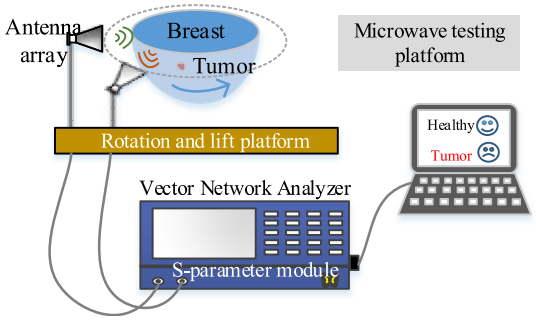

FIGURE 3. Microwave data acquisition. For more information regarding the acquisition system, please refer to [34, 35].

The specific description of the dataset is provided in [24, 33]. This dataset is divided into source domain and target domain according to volume distribution. The data of 448 scans with a smaller volume are taken as one group (referred to as "Backscatter dataset 1"), and the data of the remaining 560 scans with a larger volume are taken as another group (referred to as "Backscatter dataset 2"). The adipose shell volumes of dataset 1 range from 295 cm^3 to 652 cm^3, while the volumes of dataset 2 range from 713 cm^3 to 1113 cm^3. This setup facilitates the creation of even more challenging test scenarios. To increase the number of samples and enhance network performance, the S_{11} images are horizontally translated and reflected for data augmentation. The sample numbers of backscatter dataset 1 and backscatter dataset 2 are 5376 and 6720, respectively. The training set accounts for 75%, while the remaining 25% of the samples is the testing set. The first image on the top left of Fig. 4 represents the original image, while the remaining images represent the images after augmentation.

3. RESULTS

3.1. Performance Analysis

This section reports the impact of some main hyperparameters of the proposed method on detection performance, such as the number of feature extraction modules, the number of FC layers in the tumor detector, the size of convolution kernel, and the learning rate. ADAM optimizer is used as the backpropagation optimizer to update the gradient of the neural network.

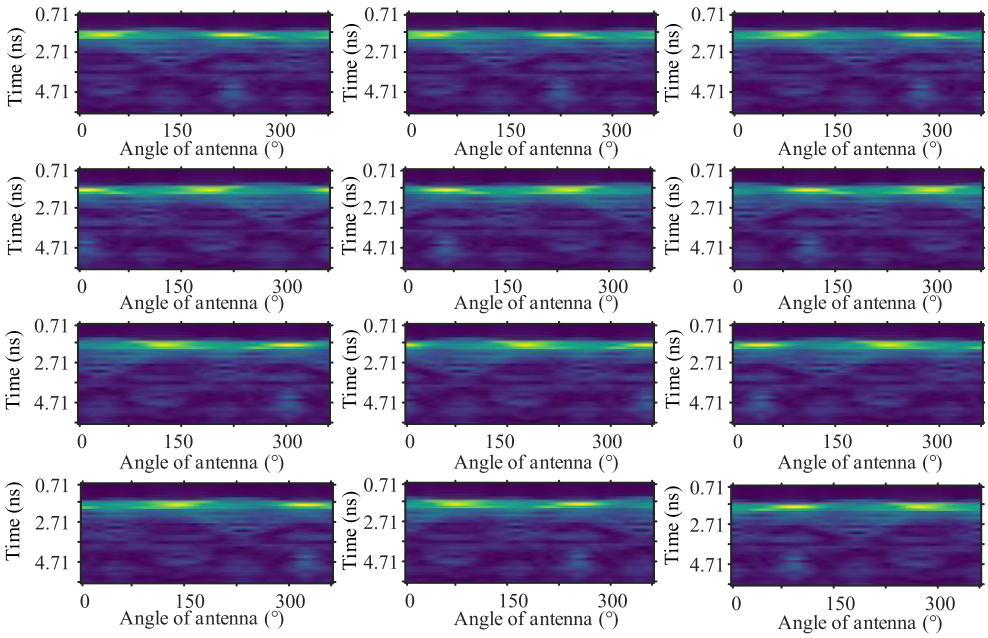

FIGURE 4. The backscatter signals after data augmentation (horizontally reflection and translation).

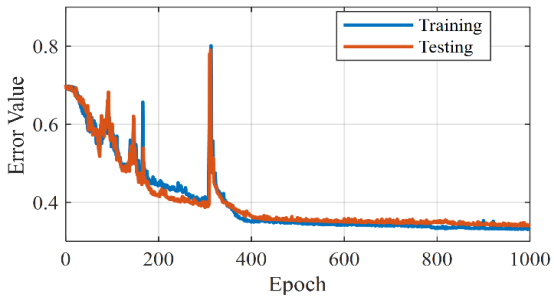

FIGURE 5. The training and testing errors' change with epochs.

The number of epochs is set to 1000. Fig. 5 plots the curve of training and testing errors with epochs. Due to the domain adversarial learning process in the proposed method, the error values of training and testing undergo oscillations, and this phenomenon is more obvious in the first 400 epochs. As the number of epochs increases, the error values of training and testing gradually converge.

3.1.1. The Influence of Feature Extraction Modules on Algorithm Performance

In the proposed method, the feature extraction module is composed of stacked convolution blocks. Table 1 shows the accuracy results of different convolution blocks on the training sets and testing sets. It is evident that as the number of convolutional blocks increases, the network performance decreases. Notably, the highest testing accuracy is attained when a single convolution block is utilized. This phenomenon is consistent with the literature [24], indicating that including additional convolution blocks for the UM-BMID dataset may not necessarily yield superior results. The reason may be that the experimental dataset is still not large enough. At present, the experimental dataset contains 1008 unique 3D phantoms' experimental scan. The sample size is relatively small under such rich data distribution changes.

TABLE 1. Performance evaluation of different FC layers.

Classifier	Training accuracy	Testing accuracy
One FC layer	77.58%	72.38%
Two FC layers	83.29%	83.15%
Three FC layers	76.63%	77.32%

3.1.2. The Influence of FC Layers on Algorithm Performance

In addition to the convolution modules, the FC layer with numerous parameters also affects the network performance. The influence of different number of FC layers on network performance is compared. As shown in Table 2, it can be seen that employing two FC layers yields the most optimal performance. When the number of layers increases to three, the neural network becomes more complex and larger. The network exhibits obvious overfitting, and the network parameters are overtrained. The network cannot adapt and generalize well to unseen data, resulting in a decrease in testing accuracy.

TABLE 2. Performance evaluation of different convolution sizes.

Kernel size	Training accuracy	Testing accuracy
[3, 3]	68.15%	64.40%
[5, 5]	83.29%	83.15%
[7, 7]	75.65%	74.40%

3.1.3. The Influence of Kernel Size on Algorithm Performance

Table 3 presents a comparative analysis of the impact of varying convolution kernel sizes on network performance. Specifically, within the context of the proposed network architecture,

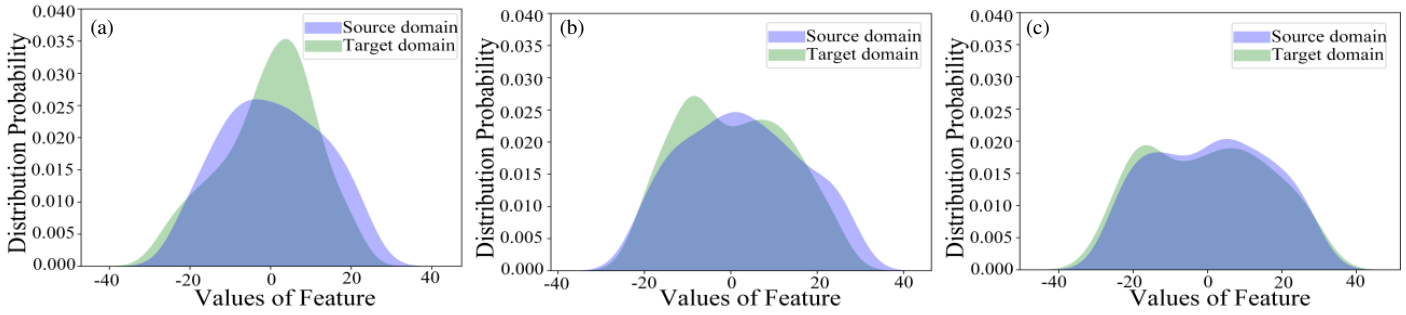

FIGURE 6. The distribution estimation of the learned features. (a) CNN; (b) DANN; (c) the proposed DATDNet.

TABLE 3. Performance evaluation of the proposed method using different learning rates.

Learning rate	Training accuracy	Testing accuracy
0.1	79.74%	77.50%
0.01	83.29%	83.15%
0.001	77.60%	78.21%
0.0001	79.68%	76.31%

employing a convolution kernel size of [5, 5] shows a notable enhancement in both training and testing accuracies. Since the sampling points of the signal are correlated in the neighborhood, a smaller $3 * 3$ convolution kernel seems to lose neighborhood information, leading to a lower accuracy. However, for a $7 * 7$ convolution kernel with a large receptive field, the effective features may be blurred. Therefore, a $5*5$ convolution kernel is suitable for the image input in this study.

3.1.4. The Influence of Learning Rate on the Algorithm Performance

This study evaluates the algorithm performance by increasing learning rate every 10 times. The results of different learning rates are detailed in Table 4. When the learning rate (0.1) is too large, the weights are prone to skip the optimal solution during the optimization process. When the learning rate (0.001/0.0001) is too small, it is easy to get stuck at the local optimum. A moderate learning rate of 0.01 is suitable for our network training.

3.2. Feature Distribution Estimation

In order to intuitively explain the effectiveness of domain transfer learning, we performed the kernel density estimation on the extracted features from the source domain and target domain. Fig. 6 shows the feature distribution probability with and without introducing domain transfer strategy. Specifically, Figs. 6(a), (b), and (c) depict the feature distributions of FC layers for the CNN, domain-adversarial neural network (DANN), and our proposed DATDNet, respectively. Observing Fig. 6(a), it is evident that for the CNN method without domain transfer, there exists a notable discrepancy in the feature distributions between the source and target domains. Encouragingly, Figs. 6(b) and 6(c) reveal a favorable trend. In both the DANN and our proposed method, which incorporate a transfer strategy, the feature distributions between the source and target domains are closer. This consistency further enhances the capabilities of domain adversarial learning methods in identifying target domain samples accurately.

3.3. Ablation Studies

To evaluate the contribution of critical network components to its overall performance, this section conducts the ablation studies on the proposed method. Ablation experiments refer to a series of performance tests conducted by removing certain key components from the network model. We removed Domain Adversarial (DA) strategies and Attention Mechanisms (AM) from the proposed methods and conducted a performance comparison respectively. Some indicators are used to evaluate the proposed method, such as accuracy, precision, sensitivity, and F1 score.

The ablation experiments are indicated in Table 5 and Table 6, and the detection accuracy of the testing data is improved by introducing domain adversarial strategies. Below is a detailed analysis of the ablation studies using the comprehensive evaluation index F1, and the CNN without additional component (denoted Case 1) is taken as the baseline. Case 3, which integrates domain adversarial strategies (6.66% improvement in Task 1 and 1.37% improvement in Task 2), shows more significant improvement over Case 2, which solely integrates attention mechanisms (4.33% improvement in Task 1 and no improvement in Task 2). In addition, when the attention mechanism module (Case 4) is built upon the domain adversarial architecture (Case 2), it achieves further performance enhancements (2.63% for Task 1 and 8.41% for Task 2). Overall, the proposed DATDNet (Case 4) has the best detection performance. Compared with baseline CNN, the accuracy is improved by 5.36% and 7.79%, the precision improved by 3.02% and 9.09%, the sensitivity improved by 9.32% and 2.7%, the F1 score improved by 6.96% and 6.32% in two different detection tasks, respectively. Overall, the accuracy of the proposed method is improved by 5.36% (AUC: 83.60%) and 7.79% (AUC: 91.20%).

4. DISCUSSION

There are other machine learning methods for microwave detection, as listed in Table 7. Song et al. [20] attempted detailed feature extraction methods (wavelet transform, empirical mode

TABLE 4. Ablation studies on Task 1.

	Case	Accuracy	Precision	Sensitivity	F1
1	CNN	71.28%	80.15%	60.45%	68.92%
2	Without DA	74.63%	82.36%	65.96%	73.25%
3	Without AM	75.82%	80.74%	**71.05%**	75.58%
4	The proposed method	**76.64%**	**83.17%**	69.77%	**75.88%**

TABLE 5. Ablation studies on Task 2.

	Case	Accuracy	Precision	Sensitivity	F1
1	CNN	75.36%	71.73%	84.86%	77.74%
2	Without DA	72.68%	69.02%	83.69%	75.65%
3	Without AM	77.56%	74.92%	83.80%	79.11%
4	The proposed method	**83.15%**	**80.82%**	**87.56%**	**84.06%**

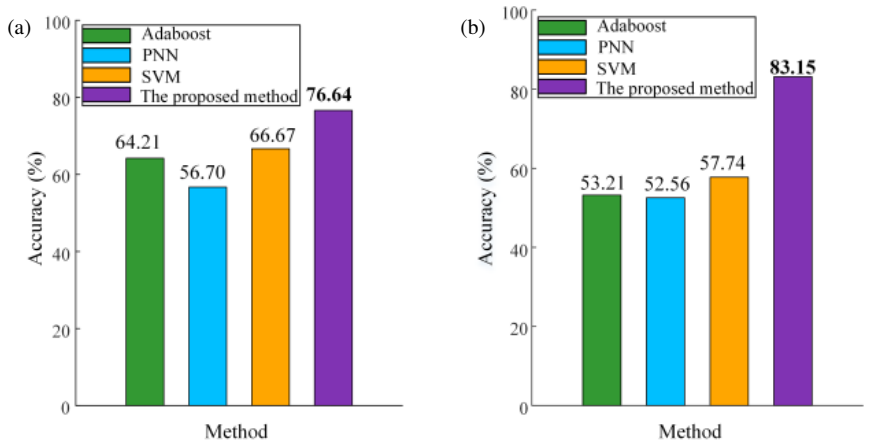

FIGURE 7. Performance comparison with other typical detection methods based on the UM-BMID dataset under varying breast morphologies. (a) Task 1. (b) Task 2.

decomposition (EMD)) and machine learning methods to detect the presence of tumors. But the feature extraction method is complex and empirical, and its tumor response is generated as a function of the breast response, not by real physical rules. Reimer and Pistorius [24] collected a publicly available experimental dataset with good breast morphology diversity and authenticity. Although the CNN method was used for tumor detection, more detailed improvements were not made to the network design. In our previous work [4], a sophisticated and accurate CNN-LSTM network based on 1D time domain signals was proposed. The model maintained good diversity in terms of morphology and structural changes during numerical simulation, but the 3D-printed phantom variations in the experimental validation were limited. Halim et al. [15] conducted a detailed study on data preprocessing for breast tumor detection. However, the data source used for the machine learning algorithm research is only a uniform and regularly shaped breast phantom, which is far from the real breast situation.

To quantitatively compare the relevant algorithms, three typical methods are used to evaluate the accuracy of tumor detection in the case of changing breast morphology on the UM-BMID dataset. In addition to the CNN method compared in the manuscript (see study 1 in Tables 5 and 6 for details), the probabilistic neural network (PNN) method used in [15], the AdaBoost model used in [29] to distinguish benign and malignant tumors, and the classical SVM method are compared. As shown in Fig. 7, SVM exhibits relatively good performance but still did not reach an acceptable level of performance. The accuracy of the proposed method shows improvements of 9.97% and 25.41% over SVM. These methods do not align features for different breast conditions, thus lacking generalization performance and making it difficult to maintain good performance in such a variable dataset. The comparative analyses with other classical methods also demonstrate the performance improvement achieved by our method and its robust generalization capabilities in detecting tumors when dealing with different breast morphologies.

At present, experimental datasets based on phantoms used in this paper have some limitations, lacking real clinical noise and real tissue variations. It is necessary to carry out clinical validation in the future. This can be achieved by seeking cooperation with hospitals to collect microwave signals from the

TABLE 6. Comparison with other related work.

Refs	Methods	Advantages & Limitations
H. Song et al. Ref [20]	Combined time-frequency and machine learning to detect tumors	• Hand-designed feature extraction is beneficial to machine learning for breast tumor detection • Tumor signals are simulated signals, and the method is artificially dependent and complicated
T. Reimer et al. Ref [24]	Utilized convolutional neural network (CNN) for tumor detection in experimental situations	• Experimental data with diversity and authenticity that can reflect diverse testing scenarios in reality • The average ROC AUC obtained across the whole test was approximately 78%. When excluding breast shells with extreme shapes, the AUC reaches 86%
M. Lu et al. Ref [4]	Proposed CNN-LSTM architecture based 1-D time domain signals	• The proposed well-designed architecture exhibits high accuracy and sensitivity for breast tumor • Although there is a large amount of simulation data, the experimental breast phantom is limited, and the variations in breast morphology are not fully considered in the experiment
A. Halim et al. Ref [15]	Proposed Probabilistic Neural Network for detect tumors	• The authors fully compared the impact of different data standardization, dimensionality reduction, and classification methods on tumor detection performance • The experimental phantom is homogeneous and monotonous
This work	The DATDNet is proposed to overcome the influence of varying breast morphologies	• Using diverse and authentic experimental data, the feature distribution differences are significantly reduced through adversarial training. The accuracy of the proposed method is improved by 5.36% (AUC: 83.60%) and 7.79% (AUC: 91.20%), compared to the baseline network • Clinical data collection and validation are required in the future

excised breast tissues or breast cancer volunteers. For the collected clinical signals, adaptive algorithms would be studied to handle motion artifacts in clinical testing. And the DATDNet algorithm should be expanded and improved through the extended datasets collected by multi-center hospitals.

In future clinical integration applications, the signal delay caused by all electronic components and modules in the hardware system can be mitigated through cavity calibration strategies. In terms of hardware compatibility, microwave tumor detection systems are typically composed of microwave antennas and signal transceiver circuits. The assembled measurement device can perform independent measurements without relying on other devices, but it is usually necessary to avoid placing it in areas with electromagnetic radiation. In addition, to enhance clinical interpretability, generating heatmaps and other visual analytics will enable medical professionals to intuitively understand the result outputs, not just specialists in deep learning or microwave technologies.

5. CONCLUSION

This paper proposes a DATDNet to improve the microwave tumor detection performance under varying breast morphologies. The entire architecture includes microwave image feature extractor, breast tumor detector, domain discriminator, and adversarial training. By searching for the optimal network architecture and integrating attention mechanisms, adversarial training is performed on the deep learning model to achieve tumor detection in unknown breast. The results demonstrate that DATDNet significantly outperforms the baseline CNN method: the accuracy is improved by 5.36% and 7.79%, and the F1 score is improved by 6.96% and 6.32% in two different detection tasks, respectively. The DATDNet exhibits a more consistent feature distribution and a superior detection performance. In the near future, the microwave signals will be collected based on the breast models with richer variations in morphology and dielectric properties to optimize the detection strategy.

ACKNOWLEDGEMENT

This work was supported in part by the National Key Research and Development Program of China (2022ZD0160403), this work was also supported part by the Natural Science Basic Research Program of Shaanxi (Program Nos. 2025JC-YBQN-926, and 2024JC-YBQN-0655), the Doctoral Research Project of Xi'an University of Technology (No. 103-451123013), the National Natural Science Foundation of China (Grant No. 62104190, No. 52205577), and the China Postdoctoral Science Foundation (No. 2021M693888).

REFERENCES

[1] Siegel, R. L., K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA: A Cancer Journal for Clinicians*, Vol. 73, No. 1, 17–48, 2023.

[2] Lazebnik, M., D. Popovic, L. McCartney, C. B. Watkins, M. J. Lindstrom, J. Harter, S. Sewall, T. Ogilvie, A. Magliocco, T. M. Breslin, *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 & Biology*, Vol. 52, No. 20, 6093–6115, 2007.

[3] Sugitani, T., S.-I. Kubota, S.-I. Kuroki, K. Sogo, K. Arihiro, M. Okada, T. Kadoya, M. Hide, M. Oda, and T. Kikkawa, "Complex permittivities of breast tumor tissues obtained from cancer surgeries," *Applied Physics Letters*, Vol. 104, No. 25, 253702, 2014.

[4] Lu, M., X. Xiao, Y. Pang, G. Liu, and H. Lu, "Detection and localization of breast cancer using UWB microwave technology and CNN-LSTM framework," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 70, No. 11, 5085–5094, 2022.

[5] Song, H., S. Sasada, T. Kadoya, K. Arihiro, M. Okada, X. Xiao, T. Ishikawa, D. O'Loughlin, J.-I. Takada, and T. Kikkawa, "Cross-correlation of confocal images for excised breast tissues of total mastectomy," *IEEE Transactions on Biomedical Engineering*, Vol. 71, No. 5, 1705–1716, 2024.

[6] Wang, J., M. Zhang, Y. Bai, H. Xu, and Y. Fan, "Distance compensation-based dual adaptive artifact removal algorithm in microwave breast tumor imaging system," *Biomedical Signal*

Processing and Control, Vol. 88, 105598, 2024.

[7] Karam, S. A. S., D. O'Loughlin, and B. M. Asl, "A novel sophisticated form of DMAS beamformer: Application to breast cancer detection," *Biomedical Signal Processing and Control*, Vol. 74, 103516, 2022.

[8] Liu, G., X. Xiao, H. Song, M. Lu, and T. Kikkawa, "An adaptive window-based hybrid artifact removal method for ultra-wide band imaging enhancement of early breast cancer detection," *Biomedical Signal Processing and Control*, Vol. 70, 102980, 2021.

[9] Al-Zuhairi, D. T., A. M. Abed, J. M. Gahl, and N. E. Islam, "Phase-based window function and CD-DMAS beamforming for microwave breast cancer detection," *IET Microwaves, Antennas & Propagation*, Vol. 14, No. 7, 608–616, 2020.

[10] Albaaj, A., Y. Norouzi, and G. Moradi, "Multivariate and multidimensional CFAR radar image for breast cancer detection," *Signal, Image and Video Processing*, Vol. 18, No. 1, 647–656, 2024.

[11] Golnabi, A. H., P. M. Meaney, S. D. Geimer, and K. D. Paulsen, "3-D microwave tomography using the soft prior regularization technique: Evaluation in anatomically realistic MRI-derived numerical breast phantoms," *IEEE Transactions on Biomedical Engineering*, Vol. 66, No. 9, 2566–2575, 2019.

[12] Liu, G., "Microwave tomographic imaging of anatomically realistic numerical phantoms with debye dispersion for breast cancer detection using a regularized inverse scattering technique in the time domain," *Chinese Journal of Electronics*, Vol. 32, No. 5, 1133–1150, 2023.

[13] Nakajima, M. and S. Kidera, "A radar-tomographic bidirectional method for quantitative microwave breast imaging," *IEEE Transactions on Antennas and Propagation*, Vol. 71, No. 12, 9814–9829, 2023.

[14] Sekkal, W., L. Merad, and S. M. Meriah, "A comparative study for breast cancer detection by neural approach for different configurations of the microwave imaging system," *Progress In Electromagnetics Research M*, Vol. 65, 69–78, 2018.

[15] Halim, A. A. A., V. Veeraperumal, A. M. Andrew, M. N. M. Yasin, M. Z. Z. Ahmad, K. Hossain, B. S. Bari, and F. Kamal, "UWB-based early breast cancer existence prediction using artificial intelligence for large data set," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, Vol. 29, No. 2, 81–90, 2023.

[16] Vijayasarveswari, V., A. M. Andrew, M. Jusoh, T. Sabapathy, R. A. A. Raof, M. N. M. Yasin, R. Ahmad, S. Khatun, and H. A. Rahim, "Multi-stage feature selection (MSFS) algorithm for UWB-based early breast cancer size prediction," *PloS One*, Vol. 15, No. 8, e0229367, 2020.

[17] Edwards, K., J. LoVetri, C. Gilmore, and I. Jeffrey, "Machine-learning-enabled recovery of prior information from experimental breast microwave imaging data," *Progress In Electromagnetics Research*, Vol. 175, 1–11, 2022.

[18] Martins, R. A., J. M. Felício, J. R. Costa, and C. A. Fernandes, "Comparison of slot-based and Vivaldi antennas for breast tumor detection using machine learning and microwave imaging algorithms," in *2021 15th European Conference on Antennas and Propagation (EuCAP)*, 1–5, Dusseldorf, Germany, Mar. 2021.

[19] Santorelli, A., Y. Li, E. Porter, M. Popović, and M. Coates, "Investigation of classification algorithms for a prototype microwave breast cancer monitor," in *The 8th European Conference on Antennas and Propagation (EuCAP 2014)*, 320–324, Hague, Netherlands, Apr. 2014.

[20] Song, H., Y. Li, and A. Men, "Microwave breast cancer detection using time-frequency representations," *Medical & Biological Engineering & Computing*, Vol. 56, No. 4, 571–582, 2018.

[21] Franceschini, S., M. M. Autorino, M. Ambrosanio, V. Pascazio, and F. Baselice, "A deep learning approach for diagnosis support in breast cancer microwave tomography," *Diagnostics*, Vol. 13, No. 10, 1693, 2023.

[22] Rana, S. P., M. Dey, R. Loretoni, M. Duranti, M. Ghavami, S. Dudley, and G. Tiberi, "Radiation-free microwave technology for breast lesion detection using supervised machine learning model," *Tomography*, Vol. 9, No. 1, 105–129, 2023.

[23] Lu, M., X. Xiao, G. Liu, and H. Lu, "Microwave breast tumor localization using wavelet feature extraction and genetic algorithm-neural network," *Medical Physics*, Vol. 48, No. 10, 6080–6093, 2021.

[24] Reimer, T. and S. Pistorius, "The diagnostic performance of machine learning in breast microwave sensing on an experimental dataset," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, Vol. 6, No. 1, 139–145, 2022.

[25] Lu, M., X. Xiao, G. Liu, H. Lu, Y. Pang, and T. Kikkawa, "Breast tumor detection by 1D-convolutional neural network based on ultra-wide-band microwave technology," *Measurement Science and Technology*, Vol. 34, No. 2, 025702, 2022.

[26] Khalid, N., M. Hashir, N. Mahmood, M. Asad, M. A. Rehman, M. Q. Mehmood, M. Zubair, and Y. Massoud, "Efficient deep learning approaches for automated tumor detection, classification, and localization in experimental microwave breast imaging data," in *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–4, Monterey, CA, USA, May 2023.

[27] Conceição, R. C., H. Medeiros, D. M. Godinho, M. O'Halloran, D. Rodriguez-Herrera, D. Flores-Tapia, and S. Pistorius, "Classification of breast tumor models with a prototype microwave imaging system," *Medical Physics*, Vol. 47, No. 4, 1860–1870, 2020.

[28] Gerazov, B. and R. C. Conceicao, "Deep learning for tumour classification in homogeneous breast tissue in medical microwave imaging," in *IEEE EUROCON 2017 — 17th International Conference on Smart Technologies*, 564–569, Ohrid, Macedonia, 2017.

[29] Janjic, A., I. Akduman, M. Cayoren, O. Bugdayci, and M. E. Aribal, "Microwave breast lesion classification-results from clinical investigation of the safe microwave breast cancer system," *Academic Radiology*, Vol. 30, S1–S8, 2023.

[30] Lazebnik, M., M. Okoniewski, J. H. Booske, and S. C. Hagness, "Highly accurate Debye models for normal and malignant breast tissue dielectric properties at microwave frequencies," *IEEE Microwave and Wireless Components Letters*, Vol. 17, No. 12, 822–824, 2007.

[31] O'Halloran, M., R. C. Conceicao, D. Byrne, M. Glavin, and E. Jones, "FDTD modeling of the breast: A review," *Progress In Electromagnetics Research B*, Vol. 18, 1–24, 2009.

[32] Shao, W., A. Edalati, T. R. McCollough, and W. J. McCollough, "A phase confocal method for near-field microwave imaging," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 65, No. 7, 2508–2515, 2017.

[33] Reimer, T., J. Krenkevich, and S. Pistorius, "An open-access experimental dataset for breast microwave imaging," in *2020 14th European Conference on Antennas and Propagation (EuCAP)*, 1–5, Copenhagen, Denmark, Mar. 2020.

[34] Reimer, T. and S. Pistorius, "Review and analysis of tumour detection and image quality analysis in experimental breast microwave sensing," *Sensors*, Vol. 23, No. 11, 5123, 2023.

[35] Ganin, Y., E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, Vol. 17, No. 59, 1–35, 2016.

[36] Du, Y., W. Liu, Y. Wang, R. Li, and L. Xie, "YOLO-CPC: A breast tumor detection and identification algorithm based on improved YOLOv7," *Signal, Image and Video Processing*, Vol. 19, No. 3, 260, 2025.

[37] Woo, S., J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 3–19, Munich, Germany, Sep. 2018.

[38] Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, □. Kaiser, and I. Polosukhin, "Attention is all you need," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, Dec. 2017.

[39] Solis-Nepote, M., T. Reimer, and S. Pistorius, "An air-operated bistatic system for breast microwave radar imaging: Pre-clinical validation," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1859–1862, Berlin, Germany, Jul. 2019.