

A Modular Deep Learning Framework for Breast Tumor Detection from Microwave Imaging Data

Philip Tchatchoua¹[0000–0001–5740–9884] and Ulysse Trin¹

SogetiLabs Research and Innovation, Capgemini 92130 Issy-les-Moulineaux, France

Abstract. Microwave imaging (MWI) is a non-ionizing, portable, and potentially low-cost complement to conventional breast cancer screening, but its coupling with artificial intelligence (AI) remains under-explored. In this work, we present a comparative benchmark that systematically evaluates multiple data representations, neural architectures, and clinical tasks for breast tumor analysis from MWI measurements. The proposed modular deep learning framework jointly addresses three pattern recognition tasks from MWI data: tumor detection, 2D localization, and tumor size estimation. The benchmark spans two public experimental datasets covering distinct acquisition setups and anatomical variabilities: Breast-Care and UMBMID. It incorporates both time–frequency signal transformations and spatial representations such as sequences and sinograms, together with a diverse set of modeling families ranging from compact CNN baselines to transformer-based and attention-enhanced architectures. Across datasets, the comparative analysis reveals that MWI signals contain informative yet challenging spatio–spectral signatures of tumor presence. While performance on BreastCare indicates the feasibility of accurate detection, localization and characterization under controlled conditions, results on the more heterogeneous UMBMID dataset highlight the limitations of current models when faced with realistic anatomical and acquisition diversity. Overall, this multi-representation, multi-architecture, and multi-task benchmark demonstrates the value of modular deep learning pipelines for decomposing detection, localization, and characterization in MWI, while underscoring the importance of domain adaptation, cross-configuration generalization, and physics-informed or multi-task learning to advance toward clinically robust performance.

Keywords: Microwave breast imaging · Pattern recognition · Deep learning · Tumor detection, localization and characterization

1 Introduction

Breast cancer is the most common cancer in women worldwide, and early detection is critical for reducing mortality. X-ray mammography, the current gold standard, suffers from reduced sensitivity in dense breasts, uses ionizing radiation, and often requires uncomfortable compression. Ultrasound and MRI can complement mammography, but ultrasound is operator-dependent with high

false positive rates, and MRI is expensive and not suited to routine screening. These limitations motivate alternative or adjunct imaging modalities.

Breast Microwave Imaging (BMI) is a non-ionizing technique that exploits dielectric contrast between malignant and normal tissues using ultra-wideband (UWB) signals in the $\sim 1\text{--}10$ GHz range [1–5]. Microwaves can penetrate breast tissue and reflect from tumors due to their higher permittivity and conductivity. BMI systems illuminate the breast and record backscattered S-parameters with antennas surrounding the breast [1, 2]. However, reconstructing images or detecting tumors from these signals is challenging due to low resolution, attenuation, skin reflections, and heterogeneity. Classical reconstruction algorithms (e.g., delay-and-sum, migration) often produce low-contrast images with artifacts, spurring interest in learning tumor signatures directly from microwave data.

AI for breast cancer detection has been studied since the 1990s, with algorithms assisting radiologists by distinguishing benign from malignant findings [6, 7]. In microwave breast imaging, early work relied on classical machine learning (ML) trained mostly on simulated or small experimental datasets [8–12]. The release of the University of Manitoba Breast Microwave Imaging Dataset (UMBMID) [13], acquired on a preclinical UWB system [14], enabled more systematic data-driven research; for example, Patel and Raina [15] compared ML algorithms on UMBMID. In parallel, deep learning (DL) has transformed medical imaging tasks such as tumor detection and localization [16–18], motivating its application to MWI data.

In this work, we present a modular deep learning approach for tumor detection, localization, and characterization from MWI. We focus on UMBMID, which provides multi-generation MRI-derived phantoms with diverse anatomies, and we use the simpler BreastCare dataset [18] for controlled benchmarking. Our three-stage pipeline consists of: (1) a binary classifier for tumor presence, (2) a regression model for 2D tumor coordinates in the imaging plane, and (3) a regression model for tumor diameter. Our main contributions are:

- A systematic evaluation of input representations (frequency/time sequences and sinograms) and CNN baselines for modular detection/localization/size prediction on BreastCare and UMBMID.
- Advanced multi-channel and attention-based models, including a dual-branch Transformer for combined S11/S21 data, which substantially improve detection and moderately improve localization/size estimation.
- An analysis of generalization across UMBMID phantom generations, highlighting remaining gaps (e.g., small tumors, precise localization).
- A discussion of clinical implications and directions toward translation, especially for early detection in young or dense-breasted populations where conventional imaging is less effective.

Section 2 reviews related work on AI for microwave breast imaging. Section 3 describes the datasets, preprocessing, and architectures. Section 4 presents experimental results. Section 5 discusses findings and future directions, and Section 6 concludes.

2 Related work

Microwave breast imaging has been extensively investigated, with multiple prototype systems using confocal radar, microwave tomography, holography, and related techniques. Several systems have undergone preliminary clinical testing; for example, a recent multicenter trial of the commercial MammoWave device on 353 women reported 82% sensitivity and $\sim 50\%$ specificity for lesion detection [19], illustrating both the promise and the current limitations of BMI in clinical practice. Ongoing work focuses on improved hardware (e.g., wideband antennas) and imaging algorithms to enhance resolution and contrast [1, 2, 5], with comprehensive overviews in [3, 23].

The availability of open experimental datasets such as UMBMID [13, 14] has accelerated ML-based research. Reimer and Pistorius [20] provided one of the first comprehensive evaluations of deep learning on experimental BMI data. Using sinogram images (angle–frequency representations) from 1257 scans of UMBMID’s second generation, they showed that a CNN significantly outperformed dense networks and logistic regression for tumor detection, achieving $AUC \approx 0.90$ when restricting the test set to phantoms similar to those in training. They also highlighted the importance of avoiding data leakage (e.g., different scans of the same phantom in train and test).

Khalid et al. [18] used the BreastCare dataset (897 scans of a single phantom with movable tumors of different sizes) to develop a deep learning pipeline for detection, sizing, and localization. They reported that a ResNet-based CNN outperformed traditional microwave imaging algorithms and classical ML, but the dataset lacked anatomical diversity and patient metadata, limiting assessment of generalization.

Classical ML approaches have mainly used engineered features extracted from time or frequency domain signals. Santorelli et al. [8] and Conceição et al. [9, 11] applied SVMs and other classifiers to distinguish tumor-bearing from healthy phantom scans with moderate success on small datasets. Patel and Raina [15] compared algorithms for tumor detection on UMBMID Gen 3 using handcrafted features, reporting up to $\sim 94\%$ accuracy but without tackling localization or size. Dridi and Gharsalli [21] also demonstrated feasibility of supervised ML on frequency domain features, but performance was constrained by feature engineering.

A recurring theme is the representation of microwave data for CNNs. The sinogram (B-scan) has emerged as a powerful structure: an image whose axes are antenna position and time or frequency, where tumors produce characteristic curves (e.g., hyperbolas in time–angle space) that CNNs can learn [20, 22]. Dridi et al. [22] used UMBMID Gen 2, converting S11/S21 signals into sinograms and training a CNN that achieved F1-scores up to 0.99 for tumor detection using S11 magnitude alone, outperforming a prior result by Al Khatib et al. [16]. They also found magnitude more discriminative than phase, and called for extending sinogram-based approaches to localization and characterization tasks.

Al Khatib et al. [16] proposed a preliminary multi-task CNN framework that jointly outputs tumor presence, coordinates, and radius from multi-angle S11

signals and metadata. They demonstrated feasibility but did not fully characterize performance, and suggested that modular (task-specific) networks might be more stable. A similar modular strategy was followed by Khalid et al. [18].

More recent work explores advanced architectures and regularization. Attention mechanisms such as Squeeze-and-Excitation (SE) blocks can re-weight channels and frequency bands. Transformer-based models, which capture long-range dependencies in sequences, are also being explored for microwave signals, with the hypothesis that they can better model wideband interactions across frequencies and antennas.

Overall, prior work indicates that: (a) high detection accuracy is attainable on experimental MWI data using DL (F1 up to 0.99) [16, 18, 20, 22]; (b) accurate localization and sizing are substantially harder, especially on multi-phantom datasets like UMBMID; and (c) advanced architectures (ResNets, Transformers, dual-branch networks) and careful evaluation protocols are required for clinically relevant performance. Our work builds on these insights by systematically evaluating baseline and advanced models for detection, localization, and characterization on both BreastCare and UMBMID.

3 Materials and Methods

3.1 Microwave Imaging Datasets

Experiments were conducted on two public breast microwave imaging datasets: BreastCare and UMBMID (Table 1).

BreastCare: The BreastCare dataset [18] contains microwave scans acquired using a portable preclinical BMI device developed at KAUST. It consists of a single tissue-mimicking phantom scanned with and without embedded tumors. In a monostatic setup, a single UWB Vivaldi antenna (7.5 dBi gain) transmits and receives while rotating around the phantom (45 angular positions over $\sim 360^\circ$) [18]. The frequency band is 2–4 GHz, with S11 recorded at 40 frequency points per scan. Each scan includes a label for tumor presence and, when present, the tumor’s 2D polar coordinates and diameter. The dataset contains 897 scans: ~ 100 without tumor, ~ 200 with tumor for detection tasks, and ~ 600 with annotated positions and sizes for localization/characterization [18]. Due to the single phantom and lack of metadata, BreastCare has low inter-subject variability and mainly serves as a controlled benchmark.

UMB MID: The University of Manitoba Breast Microwave Imaging Dataset (UMB MID) [13, 14] is larger and more diverse. It comprises three generations of 3D-printed phantoms derived from MRI scans: Gen 1 (13 phantoms, 249 scans), Gen 2 (66 phantoms, 1008 scans), and Gen 3 (20 phantoms, 200 scans) [13]. Phantoms model combinations of adipose and fibroglandular tissues with embedded spherical tumor simulators of varying sizes. The scanning system uses

UWB antennas from 1–8 GHz (9 GHz for Gen 3). Each scan yields complex S-parameters: S11 (monostatic reflection) and S21 (bistatic transmission) between antenna pairs. For Gen 2/3, this results in 72 S11 and 72 S21 signals sampled at 1001 frequency points. In total, UMBMID provides ~ 1457 scans, roughly half with tumors. Metadata includes a binary tumor label, 2D coordinates (cm) in the imaging plane, and tumor diameter (1–3 cm). This diversity makes UMBMID a challenging testbed for generalization, particularly for localization [20].

Table 1: Microwave Breast Imaging Datasets Overview.

Dataset	# Phantoms	Frequency	# Points	Angles	S-parameters	# Scans
BreastCare	1	2–4 GHz	40	45	S11 only	897
UMB MID	Gen 1: 13	1–8 GHz	1001	72	S11 only	249
UMB MID	Gen 2: 66	1–8 GHz	1001	72	S11, S21	1008
UMB MID	Gen 3: 20	1–9 GHz	1001	72	S11, S21	200

3.2 Data preprocessing

For each scan, we use the complex S-parameter data across frequencies. On UMBMID, we apply the recommended calibration by subtracting empty-tank reference measurements to reduce systematic artifacts [13, 14]. We then derive three input representations:

1. **Frequency domain sequence:** Magnitude of S-parameters in dB, $|S|_{\text{dB}} = 20 \log_{10} |S|$, as a function of frequency for each antenna. For monostatic data, this yields a vector of length N_f per angle; concatenating angles forms a 2D array (angles \times frequencies). Where available, S11 and S21 magnitudes are stacked as channels.
2. **Time domain sequence:** An inverse FFT (IFFT) is applied to obtain time domain impulse responses for each antenna. We keep a time window of interest (e.g., 0–6 ns for UMBMID) and use magnitudes. Time domain signals capture time-of-flight information related to reflector distances. We also experimented with an inverse Chirp-Z transform (ICZT) for flexible resolution [20], but it did not consistently outperform IFFT and is used mainly for comparison.
3. **Sinogram image:** We construct 2D sinograms with one axis as antenna angle and the other as frequency or time. Each pixel is the magnitude of S11 or S21 at a given angle and frequency/time. Tumors produce characteristic curved tracks in such images [20, 22]. These sinograms enable 2D CNNs to exploit spatial patterns.

All inputs are min–max normalized per scan to $[0, 1]$ for stable training. Figure 1 illustrates sequence vs. sinogram representations for a UMBMID scan.

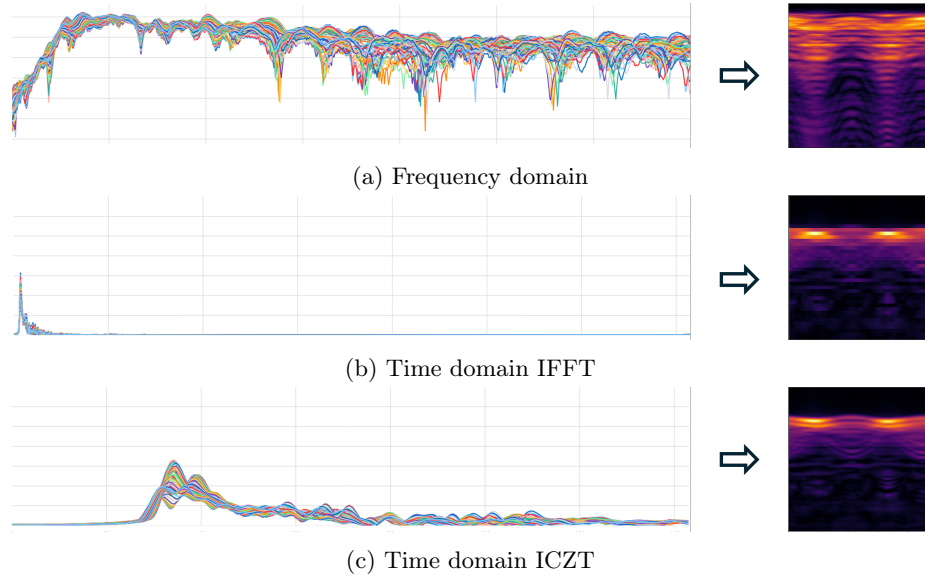


Fig. 1: Illustration of sequences vs sinograms of a single UMBMID scan.

3.3 Modular Deep Learning Framework

We adopt a modular architecture with three independent models, each dedicated to one task: tumor detection, 2D localization, and tumor size characterization. This avoids multi-objective conflicts and allows task-specific optimization. Also, localization is limited to a 2D estimation in the imaging plane, without taking into account the depth.

Baseline CNN Architectures Baseline models assess feasibility and establish reference performance across input types (frequency/time sequences and sinograms):

- **1D CNNs** for sequences: three Conv1D layers (32→64→128 filters) with max pooling and batch normalization, followed by flattening and a small dense head (64 or 256 units) with either a sigmoid (classification) or linear (regression) output.
- **2D CNNs** for sinograms: Conv2D stacks (32→64→128 filters) with max pooling and batch normalization, global average pooling, and a small dense head (64 or 256 units) with sigmoid or linear output.

We evaluated ~ 30 baseline models ($2 \text{ datasets} \times 3 \text{ tasks} \times 5 \text{ input types}$). Models were trained with Adam, Binary Cross-Entropy (BCE) or Mean Squared Error (MSE), early stopping, and stratified splits on tumor size and phantom type. Baselines performed well on BreastCare but generalized poorly on UMBMID, especially for localization and sizing.

Advanced Deep Models To overcome baseline limitations, we designed architectures better suited to global context, multi-channel data, and regularization:

- **1D and Hierarchical Transformer Encoder:** For sequence-based inputs (frequency or time), we design a Transformer architecture that uses a hierarchical encoder: input sequences are progressively down-sampled between attention stages (3 levels, pool size = 2), enabling both global context modeling and stable regression performance. This hierarchical structure is used for 2D localization (via row-wise patching and attentive pooling) and for 1D size estimation. Metadata is embedded through a small tabular encoder (dimension 24) and fused at the intermediate feature level. The prediction head consists of a dense stack ($256 \rightarrow 128 \rightarrow 64$ units, ReLU activations) followed by a final linear output layer.
- **Dual-Branch Transformer Network:** For sequence inputs (UMBMID S11+S21), S11 and S21 are processed in parallel branches. Each branch uses a small Conv1D front-end followed by Transformer encoder layers (2–3 layers, 4 heads, model dimension ~ 96 –128) for self-attention across frequency. Pooled branch features are fused via cross-attention, optionally with embedded metadata. The detection model has ~ 1.36 M parameters; larger variants (~ 2.9 M) are used for localization/characterization. We train with AdamW, BCE or Huber loss, dropout (0.2), and weight decay.
- **Inception-Attention 1D/2D CNN:** For ICZT sequences, we use Inception-style blocks (kernel sizes 1, 3, 5) to capture multi-scale patterns, followed by multi-head self-attention (4 heads, key dim 64). This hybrid (~ 2.7 M parameters) achieves high detection performance on both datasets and converges faster than full Transformers.
- **Advanced 2D Backbones (UNet, DenseNet, EfficientNet):** For sinograms, we explored deeper image CNNs. A modified UNet with 4 encoder-decoder levels and skip connections is used for localization/characterization: the final feature map feeds a dense regressor for (x,y) or size. UNet++ was also tested, with dense skip connections and dropout (0.25). For sinogram-based detection, we used truncated DenseNet-121 and EfficientNet-B0 (initialized from scratch). DenseNet (~ 1.2 M params) provided strong detection on Breast-Care and moderate performance on UBMID; EfficientNet (4.2M params) tended to overfit.
- **Channel Attention and Metadata Fusion:** We integrated Squeeze-and-Excitation (SE) blocks in some CNNs to adaptively weight channels and feature maps. Metadata, when available, is embedded (dimension 16–24) and fused mid-network (e.g., after S11/S21 fusion or near the UNet bottleneck), improving stability across generations.

Training Protocols for Advanced Models Advanced models use smaller batch sizes due to memory (e.g., 24 for UBMID sequences, 8–16 for sinograms), and up to 160–200 epochs with early stopping (patience 20–30, best-weight restoration). ReduceLROnPlateau halves the learning rate after 8–10 stagnant epochs (down to $1e-6$). Regularization includes dropout in dense and attention layers (0.2–0.25)

and L2 weight decay ($1e-4$) via AdamW. Training times are on the order of 10–20 minutes per model on an NVIDIA V100 GPU. Compared to baselines, learning curves are smoother and less prone to divergence.

4 Experiments and Results

We report results for detection, localization, and characterization, comparing baseline and advanced models on both datasets. Detection metrics include sensitivity, specificity, F1-score, and geometric mean (G_{mean}). Localization and size estimation are evaluated via Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 .

Tumor Detection On BreastCare, baseline CNNs already achieve high detection accuracy, especially with time-domain (IFFT) sequences. A 1D CNN on IFFT attained sensitivity 0.95, specificity 0.925, $F1 = 0.9048$. Frequency-domain sequences were slightly worse, and sinogram-based baselines overfit, predicting all samples as “tumor” (100% sensitivity, 0% specificity).

On UMBMID, baselines struggled: several models collapsed to trivial solutions (predicting all scans as one class). The only baseline with some skill used IFFT sequences, reaching sensitivity ≈ 0.85 but specificity ≈ 0.35 ($F1 = 0.47$, $G_{mean} \approx 0.55$).

Advanced models substantially improve detection (Table 2). On BreastCare, an Inception 2D model on frequency sinograms reaches perfect $F1$ and G_{mean} ; a CNN with attention on IFFT sequences yields $F1 \approx 0.99$ with almost no errors, matching or exceeding reported results on this dataset [18]. On UMBMID, the dual-branch Transformer on S11+S21 and metadata delivers sensitivity $> 95\%$, specificity $> 85\%$, $F1 \approx 0.92$, and $G_{mean} \approx 0.91$. An Inception-based model on ICZT sequences attains $F1 \approx 0.91$ and $G_{mean} \approx 0.92$. Thus, detection is effectively solved for phantom data under our test splits, with near-zero false negatives on UMBMID.

Table 2: Tumor Detection Results (Baseline vs Advanced Models).

Dataset	Model	Sensitivity	Specificity	F1-score	G_mean
BreastCare	Baseline CNN (seq IFFT)	0.9500	0.9250	0.9048	0.9374
BreastCare	CNN+Attn (seq IFFT)	1.0000	0.9444	0.9882	0.9718
BreastCare	Inception 2D (sino freq)	1.0000	1.0000	1.0000	1.0000
UMB MID	Baseline CNN (seq IFFT)	0.3514	0.8514	0.4685	0.5469
UMB MID	Dual-Transf. (seq freq)	0.9595	0.8649	0.9161	0.9109
UMB MID	Inception 1D (seq ICZT)	0.8378	1.0000	0.9118	0.9153

Tumor Localization Localization aims to predict tumor (x,y) coordinates within an imaging slice (diameter $\sim 18\text{--}20$ cm).

On BreastCare, baselines perform surprisingly well. The best baseline (frequency-domain sequence) reaches $MAE \approx 0.29$ cm and $R^2 \approx 0.91$; time sequences are slightly worse. Sinograms yield much larger errors ($MAE \sim 1.4$ cm, $R^2 \approx 0$). Errors increase for the smallest tumors, but overall variance is largely captured.

On UMBMID, baseline localization fails (Table 3). R^2 values are near or below zero, and MAE exceeds 1.7 cm, indicating that models predict close to the mean position.

Advanced models improve substantially but remain below clinical requirements. On BreastCare, the CNN+attention on frequency sequences does not reduce the MAE nor increases the R^2 . On UMBMID, the dual-branch Transformer on frequency sequences achieves $MAE \approx 0.70$ cm, $RMSE \approx 0.85$ cm, and $R^2 \approx 0.32$, clearly better than baselines but still far from sub-0.3 cm precision obtained on BreastCare. UNet-based sinogram models obtain comparable errors ($MAE \approx 0.7$ cm, $R^2 \approx 0.28$). Analysis by generation shows: Gen 1 residuals are relatively well-centered; Gen 2 displays high variance; and Gen 3 exhibits systematic bias, indicating that generation-specific normalization or domain adaptation could further help.

Table 3: Tumor Localization Results (Baseline vs Advanced Models).

Dataset	Model	MAE (cm)	RMSE (cm)	R^2
BreastCare	Baseline CNN (seq freq)	0.2907	0.4693	0.9092
BreastCare	CNN+Attn (seq freq)	0.3190	0.4641	0.9095
BreastCare	CNN+Attn (seq IFFT)	0.3611	0.5017	0.8949
UMBMID	Baseline CNN (seq freq)	1.7059	2.2188	0.0227
UMBMID	Dual-Transf. (seq freq)	0.6970	0.8474	0.3159
UMBMID	UNet++ (sino IFFT)	0.6951	0.8610	0.2864

Tumor Characterization We treat tumor diameter prediction as a regression problem.

On BreastCare, baselines already perform well ($MAE \approx 0.13$ cm, $R^2 \approx 0.92$), with most variance in size captured despite difficulties for the smallest tumors. Advanced models further improve performance: the CNN+attention on IFFT sequences achieves $MAE \approx 0.056$ cm and $R^2 \approx 0.98$, while sinogram-based models remain less accurate ($MAE > 0.6$ cm, $R^2 < 0.2$).

On UMBMID, both baseline and advanced models find characterization challenging (Table 4). Baseline R^2 values are near zero or negative, with MAE around 1 cm (on a 1–3 cm range), indicating regression to the mean diameter. The best advanced model (CNN+attention on IFFT) reduces MAE to $\sim 0.55\text{--}0.60$ cm and attains a small positive $R^2 \approx 0.03$, showing the model learns some size-related structure but remains far from clinically useful precision. Strong phantom-to-

phantom variability appears to mask simple size–amplitude relationships, and richer size-specific features or physics-informed modeling may be needed.

Table 4: Tumor Characterization Results (Baseline vs Advanced Models).

Dataset	Model	MAE (cm)	RMSE (cm)	R ²
BreastCare	Baseline CNN (seq IFFT)	0.1298	0.2269	0.9170
BreastCare	Transformer 1D (seq freq)	0.1866	0.2455	0.9027
BreastCare	CNN+Attn (seq IFFT)	0.0558	0.1109	0.9802
UMB MID	Baseline CNN (seq freq)	0.9687	1.0628	0.0108
UMB MID	CNN+Attn (seq IFFT)	0.5493	0.7411	0.0319
UMB MID	Dual-Transf. (seq ICZT)	0.5622	0.7498	0.0091

5 Discussion and Future Work

Our study demonstrates that deep learning can effectively detect breast tumors from microwave imaging data and can provide promising, though still limited, localization and size estimation on phantom datasets. Several key themes emerge.

First, the choice of data representation is critical. Time-domain sequences (IFFT-based) and dual-channel inputs (S11+S21) consistently offer strong performance, especially for detection and localization. These representations encode time-of-flight and multi-path information that is directly tied to spatial structure. Sinograms, while initially prone to overfitting when used with shallow CNNs, become effective when combined with sufficiently expressive architectures (e.g., UNet). On BreastCare, sinogram-based models achieved excellent detection performance, including perfect classification with frequency-domain sinograms. This likely reflects the fact that tumor presence induces global, coherent spectral perturbations that remain consistent across antennas and are therefore easily separable by convolutional or attention-based models. In contrast, localization and size estimation performed extremely poorly, despite the strong detection results. This divergence suggests that the sinograms capture the existence of scattering anomalies but provide insufficient spatial specificity for precise geometric characterization. The mono-static acquisition geometry, limited diversity of tumor positions, and the inherently diffuse propagation patterns in MWI likely contribute to this lack of spatial resolution. As a result, the models detect the presence of a tumor reliably but cannot infer its location or extent from the available signal structure. This suggests that future work should treat preprocessing and representation design as central parts of the MWI-AI pipeline rather than as fixed pre-steps.

Second, the modular design with separate models for detection, localization, and characterization proved advantageous. It simplifies optimization, as each model can be trained with a dedicated loss function and architecture tailored to its task (e.g., classification vs. regression, sequence vs. image). The detector can

be tuned for maximal sensitivity, while localization and characterization modules can focus on spatial accuracy and calibration. A potential drawback is error propagation: if detection fails, downstream tasks are not triggered. However, our advanced detection models achieve near-zero false negatives on UMBMID test splits, mitigating this concern in the phantom setting.

Third, performance varies markedly between datasets. On the simpler Breast-Care dataset, where only one phantom and a single acquisition system are present, our models reach performance metrics that are close to or exceed clinically relevant thresholds for detection, localization, and sizing. On UMBMID, with its multiple generations and more realistic anatomical variability, detection remains excellent but localization and dimension estimation lag behind. The best UMBMID localization models achieve MAE around 0.7 cm and $R^2 \approx 0.3$, while size estimation remains only weakly correlated with ground truth. Although the regression models achieve reasonable absolute errors in some cases, the corresponding R^2 values remain close to zero or even negative. This behavior is consistent with tasks where the target variance is low and the underlying mapping is highly nonlinear or weakly identifiable from the available measurements. An R^2 near zero indicates that the model performs no better than predicting the mean of the target distribution, while negative values imply that the model’s residual error exceeds this naïve baseline. In the context of MWI, such results suggest that tumor localization and size estimation rely on spatio-spectral cues that are either too subtle relative to measurement noise or too sensitive to variations in anatomy and acquisition geometry. These findings underscore the importance of diversity in training data and highlight the gap between controlled phantom studies and real-world clinical scenarios.

Fourth, our results suggest that domain adaptation and physics-informed modeling are promising directions. The observed generation-dependent biases and residual patterns indicate that the models are sensitive to changes in phantom composition and acquisition conditions. Approaches such as:

- generation-aware normalization and per-generation fine-tuning;
- domain adaptation and domain generalization techniques;
- physics-informed neural networks that incorporate knowledge of electromagnetic wave propagation and antenna geometry;

could help bridge these gaps. For example, integrating differentiable forward models or spatial priors into the learning process might regularize solutions toward physically consistent patterns and reduce overfitting to dataset-specific artifacts.

From a clinical perspective, our detection results suggest that AI-assisted MWI could serve as a highly sensitive, radiation-free pre-screening or triage tool, particularly for populations where mammography underperforms (e.g., younger women, dense breasts). Accurate localization is important for guiding secondary imaging (e.g., targeted ultrasound or MRI), and while our current UMBMID localization accuracy is not sufficient for biopsy guidance, it may be adequate for coarse region-of-interest indication. Size estimation is still too noisy for di-

rect treatment planning but could support approximate risk stratification if improved.

Future work will focus on several directions:

- **Transfer learning and domain adaptation:** Leveraging cross-dataset training (e.g., pretraining on BreastCare followed by fine-tuning on UMBMID) and explicit domain adaptation to improve generalization.
- **Physics-informed and geometry-aware networks:** Incorporating antenna geometry, wavefront structure, or approximate forward models into neural architectures, potentially via graph neural networks or hybrid model-based / data-driven schemes.
- **Data augmentation and simulation:** Generating synthetic microwave data from realistic digital phantoms to augment limited experimental datasets and expose models to a wider range of anatomies and tumor configurations.
- **Uncertainty estimation and multi-stage refinement:** Using probabilistic outputs and uncertainty measures to trigger additional refinement stages or to flag low-confidence predictions for human review.
- **Transition to clinical data:** Ultimately, phantom-based results must be validated in studies involving human subjects. This will require careful ethical and regulatory oversight, standardized protocols, and robust evaluation against clinical endpoints.

6 Conclusion

We presented a modular deep learning framework for tumor detection, localization, and size estimation from breast microwave imaging data, evaluated on BreastCare and UMBMID phantom datasets. The framework achieves near-perfect detection and strong localization and characterization on BreastCare, and markedly improves detection on UMBMID while delivering moderate localization and limited size estimation.

The study establishes strong baselines for AI-assisted MWI and highlights the key challenges posed by heterogeneous anatomies and acquisition conditions. It also points toward the need for better representations, richer datasets, and closer coupling between physics and learning to reach clinical-grade robustness. While precise localization and sizing remain open problems, the consistent high sensitivity of our models suggests that an MWI-AI system could become a valuable adjunct to existing breast cancer screening, particularly in dense-tissue populations underserved by standard modalities. Ultimately, clinical data and prospective validation will determine the feasibility and impact of such systems.

Acknowledgements The authors would like to thank KAUST and the University of Manitoba group for providing BreastCare and UMBMID open datasets, which were invaluable for this research. The authors are grateful to collaborators at IPSA for their insights into MWI. This work was supported in part by the SogetiLabs Research and Innovation grant under the WaveCare project.

References

1. N. K. Nikolova, "Microwave imaging for breast cancer," *IEEE Microwave Magazine*, vol. 12, no. 7, pp. 78–94, 2011.
2. S. C. Hagness, E. C. Fear, and A. Massa, "Guest editorial: special cluster on microwave medical imaging," *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 1592–1597, 2012.
3. B. M. Moloney, D. O'Loughlin, S. A. Elwahab, and M. J. Kerin, "Breast cancer detection – a synopsis of conventional modalities and the potential role of microwave imaging," *Diagnostics*, vol. 10, no. 2, pp. 1–13, 2020.
4. L. Wang, "Microwave Imaging and Sensing Techniques for Breast Cancer Detection," *Micromachines*, vol. 14, no. 7, 1462, 2023.
5. D. Bhargava, P. Rattanadecho, and K. Jiamjiroch, "Microwave imaging for breast cancer detection – a comprehensive review," *Engineered Science*, vol. 30, p. 1116, 2024.
6. A. Santorelli, O. Laforest, E. Porter, and M. Popović, "Image classification for a time-domain microwave radar system: Experiments with stable modular breast phantoms," in *Proc. 9th Eur. Conf. Antennas Propag. (EuCAP)*, 2015, pp. 1–5.
7. R. S. Patil and N. Biradar, "Automated mammogram breast cancer detection using the optimized combination of convolutional and recurrent neural network," *Evolutionary Intelligence*, vol. 14, pp. 1459–1474, 2021.
8. A. Santorelli, Y. Li, E. Porter, M. Popović, and M. Coates, "Investigation of classification algorithms for a prototype microwave breast cancer monitor," in *Proc. 8th IEEE Eur. Conf. Antennas Propag. (EuCAP)*, 2014, pp. 320–324.
9. R. C. Conceição, H. Medeiros, M. O'Halloran, D. Rodriguez-Herrera, D. Flores-Tapia, and S. Pistorius, "SVM-based classification of breast tumour phantoms using a UWB radar prototype system," in *Proc. IEEE URSI Gen. Assem. Sci. Symp.*, 2014, pp. 1–1 (abstract).
10. J. Sacristan, B. L. Oliveira, and S. Pistorius, "Classification of electromagnetic signals obtained from microwave scattering over healthy and tumorous breast models," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, 2016, pp. 1–5.
11. R. C. Conceição et al., "Classification of breast tumor models with a prototype microwave imaging system," *Med. Phys.*, vol. 47, no. 4, pp. 1860–1870, 2020.
12. H. Song, Y. Li, and A. Men, "Microwave breast cancer detection using time-frequency representations," *Med. Biol. Eng. Comput.*, vol. 56, no. 4, pp. 571–582, 2018.
13. T. Reimer, J. Krenkevich, and S. Pistorius, "University of Manitoba Breast Microwave Imaging Dataset (UM-BMID)," *IEEE Dataport*, Jun. 2021, doi: 10.21227/1y0z-8t98.
14. T. Reimer, J. Krenkevich, and S. Pistorius, "An open-access experimental dataset for breast microwave imaging," in *Proc. 14th European Conf. Antennas Propag. (EuCAP)*, 2020, pp. 1–5.
15. P. Patel and A. Raina, "Comparison of machine learning algorithms for tumor detection in breast microwave imaging," in *Proc. 11th Int. Conf. Cloud Computing, Data Sci. & Engineering (Confluence)*, 2021, pp. 882–886.
16. S. K. Al Khatib, T. Naous, R. M. Shubair, and H. M. El Misilmani, "A deep learning framework for breast tumor detection and localization from microwave imaging data," in *Proc. 28th IEEE Int. Conf. Electronics, Circuits and Systems (ICECS)*, 2021, pp. 1–4.

17. G. Gayathri, S. Ponnathota, and V. Badrish, "Breast cancer detection using deep learning," arXiv:2304.10386, 2023.
18. N. Khalid, 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 Proc. IEEE Int. Symp. Circuits and Systems (ISCAS), 2023, pp. 1–4.
19. D. Á. Sánchez-Bayuela, N. Ghavami, G. Tiberi, L. Sani, A. Vispa, A. Bigotti, G. Raspa, M. Badia, L. Papini, M. Ghavami, C. R. Castellano, D. Bernardi, M. Calabrese, and A. S. Tagliafico, "A multicentric, single arm, prospective, stratified clinical investigation to evaluate MammoWave's ability in breast lesions detection," PLOS ONE, vol. 18, no. 7, e0288312, 2023.
20. T. Reimer and S. Pistorius, "The diagnostic performance of machine learning in breast microwave sensing on an experimental dataset," IEEE J. Electromagn. RF Microw. Med. Biol., vol. 6, pp. 139–145, 2022.
21. M. Dridi and L. Gharsalli, "Supervised machine learning for breast cancer detection using microwave imaging in the frequency domain," in Proc. 18th European Conf. Antennas Propag. (EuCAP), 2024, pp. 1–4.
22. M. Dridi, L. Gharsalli, M. Berbiche, "Breast Cancer Detection Using Microwave Imaging Sinogram Data Structure and a Deep Learning Framework," in Proc. IEEE Conf. Antenna Measurements and Applications (CAMA), pp. 1–5 (2025).
23. E. C. Fear, J. Bourqui, and B. Mohammed, Microwave Imaging for Breast Cancer Detection. Springer, 2013.