Pathway to Demonstrating Clinical Efficacy of Microwave Breast Imaging: Qualitative and Quantitative Performance Assessment

Emily Porter, *Member, IEEE,* Declan O'Loughlin *Member, IEEE*

***Abstract*—In the last five years alone, there has been an increasing number of operational microwave breast imaging systems used in clinical trials, with increasingly large and diverse patient populations. However, despite this increased activity and volume of clinical evidence motivating research in the modality, large differences exist in how studies evaluate and report their findings. In this work, the qualitative and quantitative metrics used to measure both image quality and clinical effectiveness and efficacy are reviewed in detail. Image quality, effectiveness and efficacy do not have precise or agreed definitions and the differences between these definitions are discussed in detail. Finally, based on these understandings, the current evidence for clinical acceptance of microwave breast imaging is reviewed, with an emphasis on gaps in the trial populations to date.**

***Keywords*—breast imaging; microwave imaging; performance evaluation**

I. Introduction

MICROWAVE (MW) imaging of the breast has been proposed for a variety of applications, including for breast cancer screening, as either a standalone or complementary modality; for breast health monitoring; and for cancer treatment tracking. Potential advantages of MW imaging include cost-effective hardware, use of non-ionizing radiation with no need for injected contrast agents, safe powers levels for repeated and frequent breast scans, and the ability to make prototypes portable, wearable, or personalized. Combined, these advantages motivate the use of MW imaging instead of, or alongside, existing breast imaging technologies.

Early clinical pilot studies of MW breast imaging have included both healthy volunteers and patients with benign and malignant lesions. In recent years, larger trials have also been undertaken with several prototypes, with several achieving more than 100 patient scans [1]–[3], and in one case as many as 225 patient breast scans were completed [4]. These studies have demonstrated the promise of MW imaging in several different ways, including showing that MW image reconstructions are consistent with breast tissue densities, typically different for normal versus abnormal tissues, and often in good alignment with imaging from other technologies, e.g., x-ray mammography (MG), ultrasound (US), or magnetic resonance imaging (MRI) [3]–[8]. However, no MW imaging systems have achieved any widespread clinical acceptance or clinical usage to date.

Several reviews relevant to the topic of MW breast imaging have been published in the last five years. Some of these reviews have been technology-focused, for instance in [9], UWB antennas for use in MW breast imaging were reviewed, along with a summary of existing experimental systems. In [10], clinical MW breast imaging systems were reviewed in terms of their patient interfaces and hardware design. In [11], MW imaging algorithms and experimental and clinical system setups were reviewed. Other works have focused on translation of MW imaging to the clinic. In [12], MW breast imaging phantom and clinical studies were discussed in comparison to existing breast screening methods (x-ray mammography, 3-D mammography, ultrasound, magnetic resonance imaging, and molecular imaging). Lastly, in [13], MW prototypes were discussed for various applications, including breast imaging, along with a historical discussion of the advances and challenges of clinical MW imaging.

These review studies have provided very useful summaries and analyses of the current state of MW breast imaging algorithms, hardware, patient prototypes, and early clinical studies. However, to date, we have not seen any reports reviewing performance assessment metrics that are inevitably required for MW breast imaging to succeed in clinical applications. Qualitative and quantitative performance metrics are vital for comparing prototypes and imaging approaches, assessing iterative improvements within systems, and, most importantly, in establishing clinical efficacy of MW imaging technologies.

Therefore, in this review, we discuss qualitative and quantitative assessment of MW breast imaging through image-based evaluation metrics and through dataset-based performance metrics. We believe this is the first review study to report on these key metrics, and we particularly focus on different ways various metrics may be defined and the resulting significance of each. We also review clinical studies of MW breast imaging, and examine and compare the metrics (and their resulting values) reported in these works.

The paper is structured as follows. In Section II, image-level metrics are reported to assess and measure the quality of MW images. Both contrast-based and localization-based metrics, and their varied definitions across works, are discussed. In Section III, performance metrics that quantify dataset-level clinical efficacy are defined and discussed. Then, in Section

This work was partially supported by the University of Texas at Austin's Department of Electrical and Computer Engineering and Cockrell School start-up funds. E.P. is with the Electrical and Computer Engineering Department, University of Texas at Austin, Austin, TX 78712-1139, USA (e-mail: emily.porter@austin.utexas.edu).

D.O'L. is with the Department of Electronic and Electrical Engineering and the Trinity Centre for Biomedical Engineering, Trinity College Dublin, Dublin 2, Ireland (e-mail: d.oloughlin@tcd.ie).

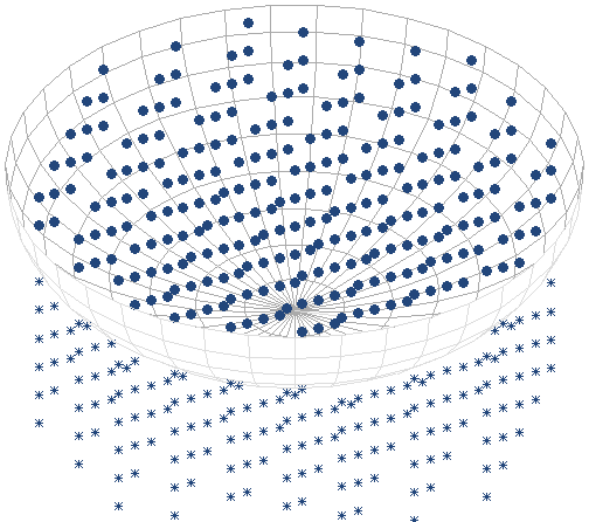


Fig. 1: Shown is an example of a discretized imaging domain with an estimated or calculated skin surface as the hemispherical surface. The full set of points, referred to as $\mathcal{P}$ in the text may be divided into points within the breast shown as the circles, $\mathcal{P}'$ in the text, and other points shown as asterisks, $\mathcal{P} \setminus \mathcal{P}'$ in the text.

True, $\mathcal{I}[\mathcal{P}']$					**Reconstructed,** $\mathcal{I}_R[\mathcal{P}']$				
CLUT	CLUT	CLUT	CLUT	CLUT	1.0	1.0	1.0	1.0	1.0
CLUT	SIG	SIG	SIG	CLUT	1.0	3.0	3.0	3.0	1.0
CLUT	SIG	SIG	SIG	CLUT	1.0	1.0	3.0	3.0	1.0
CLUT	SIG	SIG	SIG	CLUT	1.0	1.0	1.0	1.0	1.0
CLUT	CLUT	CLUT	CLUT	CLUT	1.0	2.0	2.0	1.0	1.0
(a)					(b)				

Fig. 2: Shown in (a) is the segmented true image $\mathcal{I}[\mathcal{P}']$ with $\mathcal{P}^{\text{sig}}$ shown as SIG. For some metrics, the clutter area, $\mathcal{P}^{\text{clut}}$, is as shown by CLUT, but in others, all the points in the image, $\mathcal{P}'$ are used. Shown in (b) is a simplified image. Where the true segmented image in (a) is not available, the signal and clutter regions, $\mathcal{P}^{\text{sig}}$ and $\mathcal{P}^{\text{clut}}$, are typically estimated from the image in (b) itself.

IV, approaches for pre-clinical assessment of MW prototypes are explored. Lastly, in Section V, MW breast imaging clinical studies to date are reviewed in light of their achieved performance metrics and comparison to standard imaging technologies. The pressing needs for future clinical studies, and the expected challenges, are also discussed.

II. Image Metrics

A number of quantitative metrics have been proposed to measure the quality of the microwave breast images. Often, these quantitative metrics are based on typical engineering metrics such as signal-to-clutter ratios, root mean squared error and localization error, however, no standard implementations or naming convention for these metrics exists. In most cases, the differences are minor relating to exact definitions of signal areas, clutter areas and how they are processed.

Typically, the entire imaging domain as shown in fig. 1 is discretized to a set of points, $\mathcal{P}$, based on the geometry of the system and other system design parameters. The reconstructed image with imaging algorithm R can be represented as $\mathcal{I}_R[\mathcal{P}']$ where $\mathcal{P}'$ is set of points in the reconstructed image. In some cases, the reconstruction set, $\mathcal{P}'$, is bounded by the geometry of the system and is the same for all images $\mathcal{P}' = \mathcal{P}$. However, in other cases, $\mathcal{P}'$ is instead determined based on an estimate of the skin surface (shown as a hemisphere in fig. 1) either from the microwave signals themselves, a co-mounted laser or otherwise. i.e. $\mathcal{P}' \subseteq \mathcal{P}$ and $\mathcal{P}'$ may vary from scene to scene. In radar-based imaging algorithms, the image magnitude at each point is typically independent, whereas, in some iterative or inverse scattering requiring forward solutions, the whole imaging domain needs to be reconstructed together.

With the exception of quantitative inverse scattering methods in numerical test cases, the true image, $\mathcal{I}$, is not known (a segmented example is shown in fig. 2a). In these cases, direct numerical comparisons at all points are possible such that $\mathcal{I}[\mathcal{P}'] = \mathcal{I}_R[\mathcal{P}']$. However, in all other cases, direct numerical comparisons are not possible, nor do methods exist to determine the optimal radar-based image based on *a priori* knowledge of the scene. In numerical and experimental cases, it is possible to compare images of the same scene with and without tumors, and the differences between these images have been quantified in some work. Similarly, the reconstructed image of the contralateral breast has occasionally been used as a reference.

However, for most image based metrics, the imaging domain is normally segmented into a signal area, $\mathcal{P}^{\text{sig}}$, such that:

$$\mathcal{P}^{\text{sig}} \subseteq \mathcal{P}' \tag{1}$$

based either on *a priori* knowledge of the dielectric properties of the imaging domain or based on characteristics of the reconstructed image alone. From this, a clutter region is estimated either as the points outside the signal area or the entire reconstructed image: $\mathcal{P}^{\text{clut}} = \mathcal{P}' \setminus \mathcal{P}^{\text{sig}}$ or $\mathcal{P}^{\text{clut}} = \mathcal{P}'$. From these regions, contrast-based metrics can be calculated which are discussed in detail in section II-A.

Similarly, in many cases, exact knowledge of location of the scatterers may not be available, particularly in clinical scenarios. Furthermore, other images from other modalities require registration to be directly compared to microwave breast images, so localization errors are typically based on breast quadrants and other qualitative descriptions in these scenarios.

A. Common image-based metrics

A systematic search was performed for journal articles containing experimental evaluations of microwave breast imaging systems using the terms: “Microwave” or “radar based” and “breast” and “imaging” or “imaging”. In some cases, papers were excluded where the exact implementations were not clear. Two broad categories of metric were analyzed:

- contrast-based metrics;
- localization-based metrics.

Contrast-based metrics are constructed in the form:

$$\Phi^{f^{\text{sig}}}_{f^{\text{clut}}}\left(\mathcal{P}^{\text{sig}}\right) = \frac{f^{\text{sig}}(\mathcal{I}_R[\mathcal{P}^{\text{sig}}])}{f^{\text{clut}}(\mathcal{I}_R[\mathcal{P}' \setminus \mathcal{P}^{\text{sig}}])} \tag{2}$$

or:

$$\hat{\Phi}^{f^{\text{sig}}}_{f^{\text{clut}}}\left(\mathcal{P}^{\text{sig}}\right) = \frac{f^{\text{sig}}(\mathcal{I}_R[\mathcal{P}^{\text{sig}}])}{f^{\text{clut}}(\mathcal{I}_R[\mathcal{P}'])} \tag{3}$$

where f^{sig} and f^{clut} typically calculate some summary statistic of the points in the respective areas such as the mean or maximum intensity.

In some cases, particularly numerical and experimental evaluations, $\mathcal{I}_R[\mathcal{P}^{\text{sig}}]$ is calculated using *a priori* knowledge of the target or signal expected location, however, in many others, $\mathcal{I}_R[\mathcal{P}^{\text{sig}}]$ is instead based on the characteristics of the reconstructed image alone such as the connected area around the pixel of maximum intensity.

In the descriptions below, the following definitions are used for $\mathcal{P}^{\text{sig}}$ for the maximum intensity of the image $\mathcal{P}_{\text{max}}$, or a thresholded region centered around the maximum intensity $\mathcal{P}_{\text{t}}(t)$:

- $\mathcal{P}_{\text{max}} = \{\mathbf{p}^{\text{max}}\}$;
- $\mathcal{P}_{\text{t}}(t) = \left\{\mathbf{p} \in \mathcal{P}' \;\middle|\; \mathcal{I}_R(\mathbf{p}) \geq t\mathcal{I}_R(\mathbf{p}^{\text{max}})\right\}$.

where $\mathbf{p}^{\text{max}} = \max_{\mathbf{p}\in\mathcal{P}'} \mathcal{I}_R(\mathbf{p})$. Typically, $\mathcal{P}_{\text{t}}(t)$ is limited to the connected region surrounding $\mathbf{p}^{\text{max}}$ only and other points above the threshold not connected to this region are included in $\mathcal{P}^{\text{clut}}$. Taking the example in fig. 2b, if $\mathcal{P}_{\text{t}}(t)$ with $t = 1.5$, the two points with magnitude 2.0 would not be included in that set, but just the four points with magnitude 3.0. In other cases, a region based on *a priori* knowledge of the signal region (typically the center or centroid of the tumor region, coming from a segmented image such as in fig. 2a) is used:

- $\mathcal{P}_{\text{sph}}(L) = \left\{\mathbf{p} \in \mathcal{P}' \;\middle|\; \mathbf{p} - \mathbf{p}^{\text{tum}} \leq L\right\}$;
- $\mathcal{P}_{\text{cub}}(L) = \left\{\mathbf{p} \in \mathcal{P}' \;\middle|\; \max_{\mathbf{d}\in\{\mathbf{i},\mathbf{j},\mathbf{k}\}} |\mathbf{p}\cdot\mathbf{d} - \mathbf{p}^{\text{tum}}\cdot\mathbf{d}| \leq {}^{L}\!/_{2}\right\}$;

where $\mathbf{p}^{\text{tum}}$ is the chosen *a priori* point. Simply put, $\mathcal{P}_{\text{sph}}(L)$ is a sphere of radius L centered as a point $\mathbf{p}^{\text{tum}}$ and $\mathcal{P}_{\text{cub}}(L)$ is a cuboid with sides of length L centered at the point $\mathbf{p}^{\text{tum}}$.

Typically, the summary metric is either:

- the mean: $\langle\cdot\rangle$;
- the maximum: $\max$;
- or the median: med;

intensity of the set of the points.

Based on the literature search described above, metrics where the full definition was available are listed below. A number of common metrics are identifiable, albeit sometimes under different names and these are summarised in table I. Although all very similar, several different definitions and naming conventions have been adopted across the systems. In most cases, these metrics are only calculated for the designers and are not part of the information given to the radiologist for image review, however, the different names and definitions make it challenging to compare across systems. In most cases, the final result is given in dB.

The Signal-to-Mean ratio is commonly defined as:

$$\text{SMR} = \hat{\Phi}^{\max}_{\langle\cdot\rangle}\left(\mathcal{P}_{\text{max}}\right) \tag{4}$$

and is known as the SMR or sometimes the M2AVG, MAX/AVG or PEAK/MEAN metric. A similar metric is also sometimes used with the same name:

$$\text{SMR} = \hat{\Phi}^{\langle\cdot\rangle}_{\langle\cdot\rangle}\left(\mathcal{P}^{\text{sig}}\right) \tag{5}$$

although the exact definition of $\mathcal{P}^{\text{sig}}$ is not always given. The Signal-to-Mean Ratio ($\hat{\Phi}^{\max}_{\langle\cdot\rangle}\left(\mathcal{P}_{\text{max}}\right)$) is used in [1], [14]–[28].

A related metric is the Signal-to-Clutter Ratio (SCR) which is typically calculated as:

$$\text{SCR} = \Phi^{\max}_{\max}\left(\mathcal{P}^{\text{sig}}\right) \tag{6}$$

A variety of definitions of $\mathcal{P}^{\text{sig}}$ are given in these circumstances:

- $\mathcal{P}_{\text{sph}}(L)$ where the radius L is the tumor radius or the tumor radius plus a margin such as $5\,\text{mm}$;
- $\mathcal{P}_{\text{cub}}(L)$ where the length of the cube is $1.5\,\text{cm}$ or $3\,\text{cm}$;
- $\mathcal{P}_{\text{sph}}(L)$ but centered at $\mathbf{p}^{\text{tum}}$ instead of $\mathbf{p}^{\text{max}}$ where the radius L is based on the FWHM or maximum dimension of the region identified by $\mathcal{P}_{\text{t}}(t)$. Typically, 50% is chosen as the threshold t.

A variant of the SCR metric is used in [14]–[18], [21], [26], [28]–[31]. under the name of SCR or SMXR. Where a reference image without the abnormality is available as described above, a similar metric often called the SCR has also been used which compares the maximum intensity of the signal region $\mathcal{P}^{\text{sig}}$ to the maximum intensity of the same region in the referent.

Localization Error (LE) is determined in two primary ways depending on the information available:

- by qualitative descriptions based on breast quadrant or clock face position in an image slice;
- quantitatively compared to the known tumor location.

In the quantitative cases, the Euclidean distance is used:

$$\text{LE} = \left|\mathbf{p}^{\text{max}} - \mathbf{p}^{\text{tum}}\right| \tag{7}$$

as used in [14]–[18], [21], [28]–[30]. In [23], a fractional equivalent is used relative to the width of the breast. However, the LE in this formulation is not a good reflection of imaging performance. Both the response in the image and the tumor itself are typically much larger than a point source, and the location of $\mathbf{p}^{\text{max}}$ may be skewed towards one side of the response. Furthermore, due nature of radar-based imaging, the image response would not always be expected to overlap the tumor location exactly, but may be located to one side.

The metrics in this section are typically applied to individual images as a measure of quality or used in comparison studies when proposing or comparing new imaging algorithms such as [26] as one of many. In other cases, these metrics have been used as part of an optimisation step of the imaging parameters such as [31]. However, several other metrics have also been proposed for these use cases such as [32], [33] which do not directly measure image quality as the metrics in this section do. Recent work from the literature would suggest that the use of individual, image-based metrics is not useful for comparing algorithms in a clinical context, but that database metrics such as those reviewed in the following section would be more beneficial [34]–[36].

TABLE I: Summary of the quantitative image-based metrics reviewed for this work.

Name	Other names	Variant	Uses
Signal-to-Mean ratio (SMR)	M2AVG, MAX/AVG, PEAK/MEAN		
		$\Phi^{\max}_{\langle\cdot\rangle}(\mathcal{P}_{\max})$	[1], [14]–[21], [24]–[28]
		$\Phi^{\langle\cdot\rangle}_{\langle\cdot\rangle}(\mathcal{P}^{\text{sig}})$	[22], [23]
Signal-to-Clutter ratio (SCR)	SMXR		
		$\Phi^{\max}_{\max}(\mathcal{P}^{\text{sig}})$: sphere tumour radius + 5 mm	[29]
		$\Phi^{\max}_{\max}(\mathcal{P}^{\text{sig}})$: cube of side 1.5 cm	[17], [31]
		$\Phi^{\max}_{\max}(\mathcal{P}^{\text{sig}})$: cube of side 3 cm	[16]
		$\Phi^{\max}_{\max}(\mathcal{P}^{\text{sig}})$: twice FWHM of image response	[15], [26], [28]
		between images with and without target	[14], [21]
		other or undefined	[18]
Localization Error (LE)			[14], [15], [17], [18], [21], [23], [28]–[30]

Furthermore, no standardized set of metrics or implementations for individual image assessment exist, however, a number of open-source toolboxes and analysis tools have been proposed with specific implementations [37]–[40]. While using different data formats, structures and addressing different use cases, these toolboxes help to publish and standardise implementations and work flows for tasks such as image segmentation and analysis.

III. Metrics for Establishing Clinical Efficacy

Performance metrics are needed to assess overall performance of MWI across sets of scans. In other words, performance metrics are those which can be established only from clinical studies and which provide data for clinical acceptance of the technology. Such assessment could include the study of image-based metrics across many scans, for example, by examining the average localization error of tumors or reporting the mean, worst-case, and best-case SCRs. Study of these image-based metrics would likely have most value to the engineering communities in understanding and optimizing prototypes; however, they do not allow straightforward comparison to standard or competing technologies such as mammography [12]. To this end, use of common performance metrics that quantify clinical efficacy is needed. Therefore, in this section, various performance metrics are defined and discussed.

It is important to note that there is no standard quantitative definition for clinical efficacy. Clinical efficacy is typically described as the performance of a technology in achieving the targeted goal, as measured under controlled (ideal) conditions [41]. Similarly, effectiveness is the performance under real-world or typical operating conditions [41]. Careful study design is critical to investigating both efficacy and effectiveness [41]. No single metric quantifies efficacy or effectiveness, and the choice of how to study these parameters is study-dependent and may require investigation of multiple types of metrics.

In the context of a clinical study for MWI, we emphasize that the choice of metric(s) to be used is highly dependent on the proposed use-case application - for example, a study of MWI for breast tumor screening would likely focus on sensitivity and specificity metrics, while a study of MWI-based treatment monitoring might examine metrics like 5-year survival rate and extended life expectancy.

Four terms are used to support the definition of the metrics reported in this section:

- true positive (TP): when a tumor is detected and the patient does have a tumor;
- true negative (TN): when no tumor is detected and the patient does not have a tumor;
- false positive (FP): when a tumor is detected, but in fact no tumor is present in the breast;
- false negative (FN): when no tumor is detected but a tumor is actually present in the breast.

(These terms can be applied for any type of disease. The word 'tumor' is used here as breast cancer is typically the focus of MW imaging studies of the breast).

It is clear that in order to define any metrics based on these four terms, data for each patient is required from: i) a standard technology as the ground truth, in order to define true positive and true negative; and ii) the proposed technology (i.e., microwave imaging prototype) in order to study how close the results are to the ground truth. The choice of technology to use as a standard reference ground truth can itself impact determination of metric values. For example, if using another imaging technology like mammography for assessment of sensitivity and specificity of a MW prototype, it must be kept in mind that mammography has its own sensitivity and specificity limitations [42], [43]. Therefore, in the case of breast cancer, the ground truth diagnosis typically requires a biopsy with tissue histopathology [43], [44].

A. Sensitivity and Specificity

Sensitivity and specificity are two of the most common metrics used in describing the clinical efficacy of a disease screening technology [43], [44]. These metrics are calculated across groups of patients that include those both with and without breast cancer.

The sensitivity, S_e, and specificity, S_p, are defined as:

$$\text{sensitivity} = S_e = \frac{\text{TP}}{(\text{TP} + \text{FN})}, \tag{8}$$

$$\text{specificity} = S_p = \frac{\text{TN}}{(\text{TN} + \text{FP})}. \tag{9}$$

Thus, the sensitivity is the number of patients correctly identified as having a tumor out of the total that actually do have a tumor. It can also be referred to as 'true positive rate'. Similarly, specificity is the proportion of patients correctly identified as not having a tumor when they are indeed cancer-free. Specificity may also be called 'true negative rate'. In other words, the sensitivity is a measure of the ability of a screening technology to identify a positive; while specificity is the ability to identify a negative. Ideally, FN $= 0$ and FP $= 0$, in which case $S_e = 1$ and $S_p = 1$ [45]. However, practically, sensitivity and specificity tend to be inversely related, and increasing one leads to a decrease in the other [46]. This relationship is given by the receiver operating characteristic (ROC) curve, which plots sensitivity vs. $1 - S_p$, or, equivalently, true positive rate vs. false positive rate. In an ideal scenario, the area under this curve (AUC) is 1. The shape of the ROC and the value of AUC can both be indicators of the quality of a screening or diagnostic tool [45], [47], [48].

Importantly, in screening applications, it is vital to avoid false negatives, as then patients will not receive appropriate follow-ups to potentially diagnose and treat the disease. Whereas, with a false positive, further testing will identify that there is no tumor. Therefore, high sensitivity is desirable in screening technologies. However, in diagnostic technologies the opposite is true, and high specificity is preferred in order to definitively confirm disease presence [42].

B. Positive and Negative Predictive Value

Positive and negative predictive values (PPV and NPV, respectively), are measures that provide confidence in the test results. The PPV is the likelihood of having a tumor given a positive (+) test result:

$$\text{PPV} = P(\text{tumor} \,|\, +) = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

Similarly, the NPV is the likelihood of not having a tumor given a negative (-) result:

$$\text{NPV} = P(\text{healthy} \,|\, -) = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{11}$$

An ideal technology will have NPV and PPV of 1 (i.e., no FN or FP). These values are dependent on the prevalence, $(^{(\text{TP}+\text{FN})}/_{\text{total population}})$, of breast cancer within the population [47], and are particularly informative for screening technologies as it would be expected that the majority of scans are of patients without cancer.

C. Diagnostic Odds Ratio

The Diagnostic Odds Ratio (DOR) is a metric used in assessing the effectiveness of diagnostic tools. The DOR, given by:

$$\text{DOR} = \frac{^{\text{TP}}/_{\text{FN}}}{^{\text{FP}}/_{\text{TN}}} = \frac{^{\text{TP}}/_{\text{FP}}}{^{\text{FN}}/_{\text{TN}}} \tag{12}$$

quantifies the tool's performance as a ratio of the odds of the test showing a positive when the patient has the disease relative to the odds of showing a positive when the patient does not have the disease. Unlike PPV and NPV, it does not depend on prevalence [47]. DOR can also be expressed in terms of sensitivity and specificity:

$$\text{DOR} = \frac{S_e S_p}{(1 - S_e)(1 - S_p)} \tag{13}$$

and in terms of PPV and NPVs:

$$\text{DOR} = \frac{\text{PPV} \times \text{NPV}}{(1 - \text{PPV})(1 - \text{NPV})}. \tag{14}$$

D. Accuracy

Unlike the other metrics, the term 'accuracy' may be defined differently in different contexts. Commonly used definitions include:

1) 'Diagnostic accuracy' may be used as an umbrella term for other parameters (such as S_e, S_p, predictive values, odds ratios, AUC, etc.) which together are indicators of the accuracy of a diagnostic tool in differentiating healthy versus disease cases [46], [47].
2) 'Diagnostic accuracy' may be used as a quantitative metric defined by the proportion of correct results to the total number of results (i.e., $^{(\text{TP}+\text{TN})}/_{(\text{TP}+\text{TN}+\text{FP}+\text{FN})}$) [43], [47]. In this way, the resulting accuracy is affected by disease prevalence in the underlying population [47].
3) '(Classification) Accuracy' may be used when machine learning or binary classification algorithms are applied to images to classify whether they represent healthy or diseased breasts. There are two ways accuracy in classification is commonly defined in this field: i) as an average of the sensitivity and specificity, and ii) as the overall proportion of correctly classified outputs to the total number of outputs [49]–[52]. If the test set is balanced (equal disease and healthy observations), then the definitions in i) and ii) are equivalent.

Metrics of accuracy can be extremely sensitive to the study design, and as a result can be misleading and difficult to compare. Efforts have been made to standardize reporting of diagnostic accuracy in order to improve the quality of results, thereby providing more consistent evidence bases for patient care [47], [53].

E. Comment on Other Metrics

Further metrics for assessing performance include:

- recall (or call back) rate, when the patient is asked to return for repeat or further imaging [54]–[56];
- morality reduction rate (e.g., thanks to screening followed by early intervention) [12], [44], [56];
- survival rate (typically, after 5 or 10 years of diagnosis) [44], [57];
- life expectancy (i.e., as extended through screening) [56].

However, in clinical studies for microwave imaging, we believe that these types of metrics are likely to be of secondary concern, with primary study endpoints more likely to focus on the key metrics in subsections A-D—as high-quality outcomes in relation to those metrics are unquestionably vital to a new technologies success.

Additionally, the evaluation of some metrics can be significantly impacted by the technology's usage (e.g., the screening protocols and guidelines) and not just the technology itself [57]: for instance, mortality reduction rate may be affected by how often patients are recommended to submit to screening, as well as the therapy that follows diagnosis. Furthermore, assessment of mortality reduction, survival rate, and life expectancy requires a longer-term study than one that might aim solely to quantify sensitivity and specificity, presenting another challenge for clinical studies examining these metrics with new technologies.

IV. Pre-Clinical Assessment

Pre-clinical testing, assessment, and validation of microwave systems are vital steps required prior to testing with human subjects, as confirmation of safety and some evidence of efficacy of a prototype is necessary for ethical human-based clinical studies.

MWI systems are still highly varied in design and implementation (e.g., number/type of antennas, array layout) and further research is needed to identify the optimal configuration(s) prior to large clinical trials, or else clinical trials for each design type may be needed before they can be commercialized. As clinical trials are often time-consuming and costly, consolidation of designs into optimal configuration(s) at the experimental testing phase may be beneficial.

A tool that may help identify optimal configuration(s) and establish clinically reliable prototype systems is the development and acceptance of a standard phantom or set of phantoms that provide a 'ground truth' for measurements and images. Standard phantoms have been developed for other well-established imaging technologies, including magnetic resonance imaging (MRI), computed tomography (CT), Positron Emission Tomography (PET), and Single-Photon Emission Computed Tomography (SPECT) [64]–[67]. The standardization efforts, primarily led by the National Institute of Standards and Technology (NIST, U.S. Department of Commerce), have lead to traceable phantoms that are now used with, and often sold with, common imaging systems [64]–[67]. These standard phantoms provide a basis for ensuring image data is consistent and reliable, and importantly, that the information from images is meaningfully quantitative regardless of the specific piece of imaging equipment that collected the data. They can be useful for measuring and calibrating for drift over time, evaluation of parameters such as resolution and contrast, and comparison between multi-site studies [64], [67].

In microwave breast imaging, use of standard phantoms may prove useful for optimizing prototypes, comparing between prototypes especially with image-level metrics, calibrating and quantifying measurement data, and achieving consistent and repeatable data over time and place. Several numerical and experimental breast phantoms sets have been created and proposed as standards for microwave breast imaging [68]–[71], although none have achieved widespread adoption.

A leading standard phantom, the GeePs-L2S (Supelec) phantom, an anthropomorphic breast phantom fabricated from MRI-derived 3-D prints filled with dielectric tissue-mimicking liquids, was proposed and made available for other laboratories to use to test their prototypes [70], several of which have reported the associated results [72]–[74]. However, with the diversity in MWI prototypes as mentioned above, it is still a significant challenge to develop a phantom or series of phantoms that can be used with all prototypes, and that cover all relevant patient-based parameters (e.g., breast density, size, etc.). Therefore further work is needed for the community to achieve a consensus on the design and usage of such phantoms.

V. Clinical Assessment: Studies, Needs, and Challenges

A systematic search was done for reports of MW imaging used in clinical studies. Search terms were: "Microwave" and "breast" and ("patient" or "clinical" or "human") and only journal papers that included MW imaging (as opposed to data-driven analyses without imaging) were included. Additionally, we focus on the largest and most recent clinical study to date for each prototype, unless earlier studies investigated different clinical questions. The results are summarized in table II.

As seen from the table, only four studies have reported performance metrics that quantitatively assess MW imaging on patients. These studies include four that report measures of sensitivity [1], [2], [4], [7] and only one that reports specificity [1]—all published since 2019.

A. Sensitivity and Specificity of Microwave Imaging Systems

In these works, sensitivity is broadly quantified as the correct detection rate, i.e., ratio of positive detection to total number of abnormal scans. In [2], an overall sensitivity of 63% was found for the SAFE system, with 64% for benign lesions, 59% for malignant lesions and 88% for high-risk (uncertain malignant potential) lesions. In [4], overall sensitivity was found to be 76% for the MARIA system, with 76% for benign and 74% for malignant. For the Mammowave system, empirically chosen decision thresholds were chosen to establish sensitivity (and specificity) [1]. The sensitivity was found to be 74% overall (70% for benign, 71% for malignant). In [7], the overall sensitivity for the Wavelia system was found to be 87.5% ($^{21}/_{24}$ patients; $^{12}/_{13}$ benign and $^{9}/_{11}$ malignant) with 17 of these also being approximately correctly localized ($^{7}/_{11}$ malignant and $^{10}/_{13}$ benign). To summarize, the overall sensitivity across all of these MW imaging clinic studies ranges from 63–87.5%, with 64–93% for benign lesion detection and 59–81.8% for malignant lesion detection.

Some studies have also examined sensitivity in relation to patient age, breast size, and breast density. In [2], it was found that sensitivity was lower in smaller breasts than larger ones (51% vs. 74%). In [4], higher sensitivity was found for ages 51–80 than ages 31–50 (benign: 84%, malign: 78%, vs. 75% and 66%, respectively). Interestingly, two studies have shown increased sensitivity with denser breasts; while a third study found the opposite. In [1], overall sensitivity increased to 82% (from 74%) when considering only dense breasts (benign sensitivity 78% for dense only, as compared to 70% overall; malignant sensitivity 85% for dense only, compared to 71% overall). Similarly, in [4], sensitivity of 79% was found

TABLE II: Summary of MW Breast Imaging Clinical Studies.
Note: DC = Dartmouth College; MARIA = Multistatic Array Processing for Radiowave Image Acquisition; TSAR = Tissue Sensing Adaptive Radar;SUST = Southern University of Science and Technology; HU = Hiroshima University; MU = McGill University; SU = Shizuoka University; SAFE = Scan and Find Early; GU = Goethe University of Frankfurt. Norm. = normal, Ben. = benign, Mal. = malignant. MG = mammogram; MR = magnetic resonance imaging; US = ultrasound; RR = radiologist review.

MW Prototype	Application	Patient Group(s)			Size	Performance Metrics	Reference Diagnosis
		Norm.	Ben.	Mal.			
DC [58]	non-specific	✗			43	-	RR (MG)
DC [3]	non-specific	✗	✗	✗	130	-	Pathology
DC [6]	therapy monitoring			✗	8	-	MR
MARIA® M5 [4]	non-specific		✗	✗	225	diagnostic sensitivity	RR (MR, MG, US)
TSAR [59]	complementary breast imaging	✗	✗	✗	8	-	MG, MR, Pathology
HU [8]	screening, detection			✗	5	-	Histopathology
SUST [60]	complementary screening, detection	✗			11	-	-
MU [61]	health monitoring	✗			13	-	-
SU [62]	early detection			✗	2	-	MR
Wavelia [7]	non-specific		✗	✗	24	sensitivity	Pathology, MG
MammoWave [1]	screening	✗	✗	✗	103	sensitivity, specificity	RR (MG, US, MR, Pathology)
SAFE [2]	screening, early detection		✗	✗	115	sensitivity	Pathology
GU [63]	complementary diagnosis			✗	2	-	MG

for dense breasts for both benign and malignant detection, with only 71% for malignant and 62% for benign for low-density breasts. Lastly, in [2], sensitivity was found to decrease with increasing density: fatty breast (86% sensitivity), normal fibroglandular tissue (75%), heterogeneously dense (65%), and dense (48%). In all cases, the densities were defined by their BI-RADS (Breast Imaging-Reporting and Data System) value: A (non-dense, mostly fatty), B (scattered fibroglandular tissue), C (heterogeneously dense), D (extremely dense); where "dense" includes groups C and D [1], [2], [4].

For specificity, it was found in [1] that the specificity (true negative rate) for the Mammowave system was approximately 62% ($^{32}/_{52}$ correctly identified as negative). Additionally, in [51], the results from the same Wavelia clinical study as [7] were further analyzed using Quadratic Discriminant Analysis (QDA) to differentiate benign and malignant lesions. Accurate discrimination was achieved in 88.5% cases, and accurate lesion size estimate in 76.5% of cases. Although specificity typically refers to differentiating normal (healthy) cases from diseased cases, it is also valuable for a screening or diagnostic tool to be able to discriminate between types of lesions.

B. MW Imaging and Standard Technologies in Clinical Use

While the large-scale clinical trials of MW imaging in recent years have certainly resulted in valuable data, clinical efficacy data is still limited and the potential of MW imaging to meet or exceed performance of established breast imaging technologies (MG, MR, US) is yet to be concretely demonstrated. In order to achieve widespread clinical usage, MW technologies will need to demonstrate an aspect of superiority over standard care competitors. This may be in terms of low-cost, or increased sensitivity or specificity for at least a subset of the population, or it may fill a clinical gap where no imaging is currently part of the standard patient pathway. Therefore, this will highly depend on the proposed use-case application for the MW imaging prototype.

Breast screening is a popular use-case, and most clinical studies to date have focused on this application. X-ray mammography is typically referred to as the 'gold standard' for breast cancer screening. For mammography, sensitivity and specificity estimates have broad ranges, approximately 44-86% and 90-94%, respectively [12], [44], [75]–[77], with significantly lower sensitivity values typically found for dense breasts, as high density tissues may mask the appearance of tumors [12], [44], [78], [79]. Ultrasound (US) and MRI may also be used in screening, typically at a secondary screening stage. The sensitivity and specificity of US are estimated at 80% and 88%, respectively [80]. MRI, which can image vascularity and peritumoral inflammation is therefore highly sensitive (near to 100%); however specificity is less clear with studies reporting values ranging from 37-97% [44], [81].

MW imaging is on dielectric contrasts between tumor and healthy tissues, which are predominantly based on tissue water content across the MW frequency range [12]. Therefore, like mammography, MW imaging may struggle to in scenarios where background tissues are highly dense. With dense background tissues, the expected contrast between tumor and tissue may be marginal, and moreover, MW propagation losses are high - therefore, both sensitivity and specificity of MW imaging may be limited in imaging high density breasts.

For MW imaging, we note significant differences in the sensitivity of MW imaging reported across systems, suggesting that, as well as being impacted by the study design (i.e., patients recruited), the sensitivity might be system- (i.e., prototype design: frequency range, antennas, array layout, etc) or post-processing- (i.e., filtering, noise correction, reconstruction or detection algorithm) dependent. One notable difference is also in how 'sensitivity' is defined: in [7], sensitivity values are given for correct detection with lower sensitivity values identified for combined correct detection and localization; in [4], detection in the correct location is considered; in [2], position, size, and shape are examined; and in [1], an approach based on identifying and scoring features followed by

thresholding to identify positive vs. negative findings enables calculation of sensitivity and specificity. It is very likely that calculation of sensitivity with these different approaches would lead to different values. However, importantly, these studies have been too small and too varied to assess with certainty the reasons behind the different sensitivity values. Further, only one study has addressed specificity. Therefore, from the achieved MW data to date, it is inconclusive if MW imaging will be advantageous for this application.

It is additionally important to note that studies of mammography screening examining sensitivity, specificity, and mortality reduction, have included study groups commonly including tens and hundreds of thousands of people, all the way up to millions, and have spanned across many years [56], [78], [82]. Therefore, demonstrating MW as screening tool or a complementary screening tool will require large trials which in turn will likely be costly and long-term efforts.

However, for other applications like breast health monitoring or treatment monitoring, there may not be any current standard approach so the burden of proof for efficacy of MW imaging may be more feasible to achieve in the near-term. However, regardless of proposed clinical application, it is apparent that in order to assess and quantitatively demonstrate clinical efficacy of MW imaging, there is still a significant need for clinical studies of MW breast imaging that are sufficiently powered and included large-scale patient groups.

VI. Conclusion

In recent years, microwave breast imaging has seen an increase in the number of systems trialled with patients, demonstrating the safety of the technique and helping increase understanding of the potential clinical applications of the modality. However, despite the substantial body of experimental and clinical studies, there is little consistency in reporting of results of either the image quality or of the overall efficacy observed. In this work, methods of assessing both image quality and overall trial efficacy are assessed.

Although very similar, many quantitative metrics have been proposed for analysing microwave breast imaging, both requiring *a priori* information and not. Although mostly standard and easy to understand, the multitude of definitions makes it both difficult to compare between studies and also to identify a set of useful metrics or features which may be helpful in evaluating efficacy in clinical trials. Additionally, the majority of imaging studies and comparative studies use these types of metrics alone for quality assessment and comparison, however, recent work would suggest that image-based metrics in isolation are not sufficient. Further evaluation of a set of objective quantitative metrics which are not sensitive to imaging resolution or other system-specific parameters may be useful.

Furthermore, despite the substantial increase in the both the amount and the size of clinical trials in the last five years, some key questions remain unanswered. Different trials use different definitions of common terms such as sensitivity, making it challenging to synthesise or compare trials. Furthermore, few trials have recruited from asymptomatic populations, limiting the understanding of the potential specificity which could be expected. In many cases, the expected clinical indication is not clear, and very large trials would be needed to establish enough clinical evidence to complement or replace existing modalities.

Due to these factors, further work is needed to establish a clear clinical need and route to proving efficacy.

References

[1] L. Sani, A. Vispa, R. Loretoni, *et al.*, "Breast lesion detection through mammowave device: Empirical detection capability assessment of microwave images' parameters," *PLOS ONE*, vol. 16, no. 4, pp. 1–17, Apr. 2021.

[2] A. Janjic, M. Cayoren, I. Akduman, *et al.*, "SAFE: A novel microwave imaging system design for breast cancer screening and early detection—clinical evaluation," *Diagnostics*, vol. 11, no. 3, 2021, Art. 533.

[3] S. P. Poplack, T. D. Tosteson, W. A. Wells, *et al.*, "Electromagnetic Breast Imaging: Results of a Pilot Study in Women with Abnormal Mammograms," *Radiology*, vol. 243, no. 2, pp. 350–359, May 2007.

[4] M. Shere, I. Lyburn, R. Sidebottom, *et al.*, "MARIA® M5: A multicentre clinical study to evaluate the ability of the Micrima radio-wave radar breast imaging system (MARIA®) to detect lesions in the symptomatic breast," *European Journal of Radiology*, vol. 116, pp. 61–67, Jul. 2019.

[5] S. P. Poplack, K. D. Paulsen, A. Hartov, *et al.*, "Electromagnetic Breast Imaging: Average Tissue Property Values in Women with Negative Clinical Findings," *Radiology*, vol. 231, no. 2, pp. 571–580, May 2004.

[6] P. M. Meaney, P. A. Kaufman, L. S. Muffly, *et al.*, "Microwave imaging for neoadjuvant chemotherapy monitoring: Initial clinical experience," *Breast Cancer Research*, vol. 15, no. 2, pp. 1–16, May 2013.

[7] B. Moloney, P. McAnena, S. Abd Elwahab, *et al.*, "Microwave imaging in breast cancer - results from the first-in-human clinical investigation of the Wavelia system," *Acad Radiol.*, 2022, [Epub ahead of print].

[8] H. Song, S. Sasada, T. Kadoya, *et al.*, "Detectability of Breast Tumor by a Hand-held Impulse-Radar Detector: Performance Evaluation and Pilot Clinical Study," *Scientific Reports*, vol. 7, no. 1, Dec. 2017, Art. 16353.

[9] M. Z. Mahmud, M. T. Islam, N. Misran, *et al.*, "Ultra-wideband (UWB) antenna sensor based microwave breast imaging: A review," *Sensors*, vol. 18, no. 9, 2018, Art. 2951.

[10] D. O'Loughlin, M. O'Halloran, B. M. Moloney, *et al.*, "Microwave breast imaging: Clinical advances and remaining challenges," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 11, pp. 2580–2590, 2018.

[11] M. A. Aldhaeebi, K. Alzoubi, T. S. Almoneef, *et al.*, "Review of microwaves techniques for breast cancer detection," *Sensors*, vol. 20, no. 8, 2020, Art. 2390.

[12] A. Modiri, S. Goudreau, A. Rahimi, *et al.*, "Review of Breast Screening: Towards Clinical Realization of Microwave Imaging," *Medical Physics*, vol. 44, no. 12, e446–e458, Nov. 2017.

[13] J.-C. Bolomey, "Crossed viewpoints on microwave-based imaging for medical diagnosis: From genesis to earliest clinical outcomes," *The World of Applied Electromagnetics: In Appreciation of Magdy Fahmy Iskander*, pp. 369–414, Jan. 2018.

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

[15] D. O'Loughlin, M. Elahi, B. Lavoie, *et al.*, "Assessing patient-specific microwave breast imaging in clinical case studies," *Sensors*, vol. 21, no. 23, 2021, Art. 8048.

[16] L. Kranold, M. Taherzadeh, F. Nabki, *et al.*, "Microwave breast screening prototype: System miniaturization with ic pulse radio," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 5, no. 2, pp. 168–178, 2021.

[17] H. Song, S. Sasada, N. Masumoto, *et al.*, "A two-stage rotational surface clutter suppression method for microwave breast imaging with multistatic impulse-radar detector," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9586–9598, 2020.

[18] D. Al-Zuhairi, A. Abed, J. Gahl, *et al.*, "Phase-based window function and cd-dmas beamforming for microwave breast cancer detection," *IET Microwaves, Antennas and Propagation*, vol. 14, no. 7, pp. 608–616, 2020.

[19] M. Islam, M. Samsuzzaman, S. Kibria, *et al.*, "Metasurface loaded high gain antenna based microwave imaging using iteratively corrected delay multiply and sum algorithm," *Scientific Reports*, vol. 9, no. 1, 2019.

[20] A. Vispa, L. Sani, M. Paoli, *et al.*, "UWB device for breast microwave imaging: Phantom and clinical validations," *Measurement: Journal of the International Measurement Confederation*, vol. 146, pp. 582–589, 2019.

[21] S. Shah Karam, D. O'Loughlin, B. L. Oliveira, *et al.*, "Weighted delay-and-sum beamformer for breast cancer detection using microwave imaging," *Measurement*, vol. 177, 2021, Art. 109283.

[22] L. Sani, N. Ghavami, A. Vispa, *et al.*, "Novel microwave apparatus for breast lesions detection: Preliminary clinical results," *Biomedical Signal Processing and Control*, vol. 52, pp. 257–263, 2019.

[23] S. Mukherjee, L. Udpa, S. Udpa, *et al.*, "A time reversal-based microwave imaging system for detection of breast tumors," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 5, pp. 2062–2075, 2019.

[24] S. Kibria, M. Samsuzzaman, M. Islam, *et al.*, "Breast phantom imaging using iteratively corrected coherence factor delay and sum," *IEEE Access*, vol. 7, pp. 40 822–40 832, 2019, Art. 8672573.

[25] M. Elahi, D. O'Loughlin, B. Lavoie, *et al.*, "Evaluation of image reconstruction algorithms for confocal microwave imaging: Application to patient data," *Sensors*, vol. 18, no. 6, 2018, 1678.

[26] M. Elahi, C. Curtis, B. Lavoie, *et al.*, "Performance of leading artifact removal algorithms assessed across microwave breast imaging prototype scan configurations," *Computerized Medical Imaging and Graphics*, vol. 58, pp. 33–44, 2017.

[27] D. Byrne, M. Sarafianou, and I. Craddock, "Compound radar approach for breast imaging," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 40–51, 2017.

[28] A. Santorelli, E. Porter, E. Kang, *et al.*, "A time-domain microwave system for breast cancer detection using a flexible circuit board," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 11, pp. 2986–2994, 2015.

[29] T. Reimer and S. Pistorius, "An optimization-based approach to radar image reconstruction in breast microwave sensing," *Sensors*, vol. 21, no. 24, 2021, Art. 8172.

[30] G. Liu, X. Xiao, H. Song, *et al.*, "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, 2021.

[31] X. Xiao, Y. Liu, H. Song, *et al.*, "Optimal microwave breast imaging using quality metrics and simulated annealing algorithm," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 30, no. 10, 2020, e22364.

[32] B. R. Lavoie, J. Bourqui, E. C. Fear, *et al.*, "Metrics for Assessing the Similarity of Microwave Breast Imaging Scans of Healthy Volunteers," *IEEE Transactions on Medical Imaging*, vol. PP, no. 99, pp. 1–1, 2018.

[33] D. O'Loughlin, B. L. Oliveira, M. A. Elahi, *et al.*, "Parameter Search Algorithms for Microwave Radar-Based Breast Imaging: Focal Quality Metrics as Fitness Functions," *Sensors*, vol. 17, no. 12, Dec. 2017, Art. 2823.

[34] D. O'Loughlin, B. L. Oliveira, A. Santorelli, *et al.*, "Sensitivity and specificity estimation using patient-specific microwave imaging in diverse experimental breast phantoms," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 303–311, Jan. 2019.

[35] D. O'Loughlin, B. L. Oliveira, M. Glavin, *et al.*, "Comparing radar-based breast imaging algorithm performance with realistic patient-specific permittivity estimation," *Journal of Imaging*, vol. 5, no. 11, 2019.

[36] D. O'Loughlin, B. L. Oliveira, M. Glavin, *et al.*, "Advantages and Disadvantages of Parameter Search Algorithms for Permittivity Estimation for Microwave Breast Imaging," in *Proceedings of the 13th European Conference on Antennas and Propagation (EuCAP)*, Kraków, Poland: IEEE, Mar. 2019.

[37] N. K. Nikolova, "Microwave Biomedical Imaging," *Wiley Encyclopedia of Electric and Electronics Engineering*, pp. 1–22, Apr. 2014.

[38] D. O'Loughlin, M. A. Elahi, E. Porter, *et al.*, "Open-source software for microwave radar-based image reconstruction," in *Proceedings of the 12th European Conference on Antennas and Propagation (EuCAP)*, London, UK, Apr. 2018.

[39] T. Reimer, J. Krenkevich, and S. Pistorius, "An open-access experimental dataset for breast microwave imaging," 2020.

[40] D. Kurrant, N. Abdollahi, M. Omer, *et al.*, "MWSegEval—an image analysis toolbox for microwave breast images," *SoftwareX*, vol. 15, p. 100 728, 2021.

[41] A. G. Singal, P. D. R. Higgins, and A. K. Waljee, "A primer on effectiveness and efficacy trials," *Clin Transl Gastroenterol.*, vol. 5, no. 1, e45, 2014.

[42] J. J. Deeks, "Systematic reviews of evaluations of diagnostic and screening tests," *BMJ Systematic reviews in health care*, vol. 323, no. 7305, pp. 157–162, 2001.

[43] M. Zeeshan, S. B., Q. S. B. Khalid, *et al.*, "Diagnostic Accuracy of Digital Mammography in the Detection of Breast Cancer," *Cureus*, vol. 10, no. 4, e2448, 2018.

[44] J. V. Fiorica, "Breast Cancer Screening, Mammography, and Other Modalities," *Clinical Obstetrics and Gynecology*, vol. 59, no. 4, pp. 688–709, 2016.

[45] L. Maxim, R. Niebo, and M. Utell, "Screening tests: A review with examples," *Inhal Toxicol.*, vol. 26, no. 13, pp. 811–28, 2014.

[46] J. Shreffler and M. R. Huecker, "Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios," *StatsPearl Publishing*, pp. 1–8, 2022.

[47] A.-M. Šimundić, "Measures of diagnostic accuracy: Basic definitions," *Electronic Journal of the International Federation of Clinical Chemistry and Laboratory Medicine (EJIFCC)*, vol. 19, no. 4, pp. 203–211, 2009.

[48] B. L. Oliveira, D. Godinho, M. O'Halloran, *et al.*, "Diagnosing breast cancer with microwave technology: Remaining challenges and potential solutions with machine learning," *Diagnostics*, vol. 8, no. 2, 2018, Art. 36.

[49] A. Santorelli, E. Porter, E. Kirshin, *et al.*, "Investigation of classifiers for tumor detection with an experimental time-domain breast screening system," *Progress In Electromagnetics Research*, vol. 144, pp. 45–57, 2014.

[50] R. Conceição, H. Medeiros, D. Godinho, *et al.*, "Classification of breast tumor models with a prototype microwave imaging system," *Medical Physics*, vol. 47, no. 4, pp. 1860–1870, 2020.

[51] B. Moloney, P. McAnena, S. Elwahab, *et al.*, "The Wavelia microwave breast imaging system-tumour discriminating features and their clinical usefulness," *Medical Physics*, vol. 94, no. 1128, 2021.

[52] S. Rana, M. Dey, and G. Tiberi, "Machine learning approaches for automated lesion detection in microwave breast imaging clinical data," *Scientific Reports*, vol. 9, no. 10510, 2019.

[53] P. Bossuyt, "Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. standards for reporting of diagnostic accuracy," *Clinical Chemistry*, vol. 49, no. 1, pp. 1–6, 2003.

[54] D. Gur, "Changes in Breast Cancer Detection and Mammography Recall Rates After the Introduction of a Computer-Aided Detection System," *JNCI: Journal of the National Cancer Institute*, vol. 96, no. 3, pp. 185–190, Feb. 2004.

[55] G. H. Rauscher, A. M. Murphy, Q. Qiu, *et al.*, "The "sweet spot" revisited: Optimal recall rates for cancer detection with 2d and 3d digital screening mammography in the Metro Chicago breast cancer registry," *American Journal of Roentgenology*, vol. 216, no. 4, pp. 894–902, 2021.

[56] K. C. Oeffinger, E. T. Fontham, R. Etzioni, *et al.*, "Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society," *JAMA*, vol. 314, no. 15, pp. 1599–1614, 2015.

[57] N. Iragorri and E. Spackman, "Assessing the value of screening tools: Reviewing the challenges and opportunities of cost-effectiveness analysis," *Public Health Rev.*, vol. 39, no. 17, pp. 1–27, 2018.

[58] P. M. Meaney, M. W. Fanning, T. Raynolds, *et al.*, "Initial Clinical Experience with Microwave Breast Imaging in Women with Normal Mammography," *Academic Radiology*, vol. 14, no. 2, pp. 207–218, Feb. 2007.

[59] E. C. Fear, J. Bourqui, C. F. Curtis, *et al.*, "Microwave Breast Imaging With a Monostatic Radar-Based System: A Study of Application to Patients," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 5, pp. 2119–2128, May 2013.

[60] F. Yang, L. Sun, Z. Hu, *et al.*, "A large-scale clinical trial of radar-based microwave breast imaging for Asian women: Phase I," in *Proceedings of the International Symposium on Antennas and Propagation (APSURSI)*, San Diego, CA, USA, Jul. 2017, pp. 781–783.

[61] E. Porter, M. Coates, and M. Popović, "An Early Clinical Study of Time-Domain Microwave Radar for Breast Health Monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 530–539, Mar. 2016.

[62] Y. Kuwahara, "Microwave Imaging for Early Breast Cancer Detection," in *New Perspectives in Breast Imaging*, A. M. Malik, Ed., InTech, Oct. 2017.

[63] D. Wörtge, J. Moll, V. Krozer, *et al.*, "Comparison of X-ray-mammography and planar UWB microwave imaging of the breast: First results from a patient study," *Diagnostics*, vol. 8, no. 3, 2018, 54.

[64] K. F. Stupic, "A standard system phantom for magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 86, no. 3, pp. 1194–1211, 2021.

[65] K. E. Keenan, "Multi-site, multi-platform comparison of MRI T1 measurement using the system phantom," *PloS ONE*, vol. 16, no. 6, e0252966, 2021.

[66] Z. H. Levine, "A low-cost density reference phantom for computed tomography," *Medical Physics*, vol. 36, no. 2, pp. 286–288, 2009.

[67] B. Zimmerman and D. Bergeron, "The use of calibrated phantoms to assess the accuracy of PET-CT image quantification for 64Cu," vol. 58, no. Suppl.1, pp. 1318–1318, 2017.

[68] E. Zastrow, S. Davis, M. Lazebnik, *et al.*, *Database of 3d grid-based numerical breast phantoms for use in computational electromagnetics simulations (2008)*, http://uwcem.ece.wisc.edu/home.htm, Department of Electrical and Computer Engineering University of Wisconsin-Madison, 2008.

[69] M. Omer, P. Mojabi, N. Abdollahi, *et al.*, "Breast Imaging with Multiphysics Prior for Improved Tumour Detection and Localization," in *Proceedings of the 18th International Symposium on Antenna Technology and Applied Electromagnetics (ANTEM)*, Waterloo, Ontario, Canada, Aug. 2018, pp. 1–2.

[70] N. Joachimowicz, B. Duchêne, C. Conessa, *et al.*, "Anthropomorphic Breast and Head Phantoms for Microwave Imaging," *Diagnostics*, vol. 8, no. 4, pp. 1–12, 2018.

[71] B. L. Oliveira, D. O'Loughlin, M. O'Halloran, *et al.*, "Microwave Breast Imaging: Experimental tumour phantoms for the evaluation of new breast cancer diagnosis systems," *Biomedical Physics & Engineering Express*, vol. 4, no. 2, Feb. 2018, Art. 025036.

[72] T. Rydholm, A. Fhager, M. Persson, *et al.*, "A first evaluation of the realistic Supelec-Breast phantom," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 1, no. 2, pp. 59–65, 2017.

[73] M. Koutsoupidou, I. S. Karanasiou, C. G. Kakoyiannis, *et al.*, "Evaluation of a tumor detection microwave system with a realistic breast phantom," *Microwave and Optical Technology Letters*, vol. 59, no. 1, pp. 6–10, 2017.

[74] J. T. Vasquez, F. Vipiana, M. Casu, *et al.*, "Experimental assessment of qualitative microwave imaging using a 3-d realistic breast phantom," in *Proceedings of the 11th European Conference on Antennas and Propagation (EUCAP)*, Paris, France, 2017, pp. 2728–2731.

[75] C. D. Lehman, "Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection," *JAMA Internal Medicine*, vol. 175, no. 11, pp. 1828–1837, Nov. 2015.

[76] N. Ohuchi, A. Suzuki, T. Sobue, *et al.*, "Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): A randomised controlled trial," *The Lancet*, vol. 387, no. 10016, pp. 341–348, 2016.

[77] E. Pisano, R. Hendrick, M. Yaffe, *et al.*, "Diagnostic accuracy of digital versus film mammography: Exploratory analysis of selected population subgroups in DMIST," *Radiology*, vol. 246, no. 2, pp. 376–83, 2008.

[78] M. von Euler-Chelpin, M. Lillholm, I. Vejborg, *et al.*, "Sensitivity of screening mammography by density and texture: a cohort study from a population-based screening program in Denmark," *Breast Cancer Research*, vol. 21, no. 11, 2019.

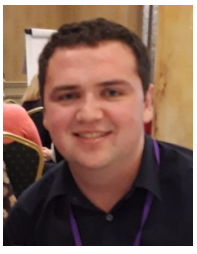
[79] E. Lynge, I. Vejborg, Z. Andersen, *et al.*, "Mammographic Density and Screening Sensitivity, Breast Cancer Incidence and Associated Risk Factors in Danish Breast Cancer Screening," *Journal of Clinical Medicine*, vol. 8, no. 11, pp. 1–11, 2021.

[80] R. Sood, A. F. Rositch, D. Shakoor, *et al.*, "Ultrasound for breast cancer detection globally: A systematic review and meta-analysis," *Journal of global oncology*, vol. 5, pp. 1–17, 2019.

[81] R. d. O. Pereira, L. A. d. Luz, D. C. Chagas, *et al.*, "Evaluation of the accuracy of mammography, ultrasound and magnetic resonance imaging in suspect breast lesions," *Clinics*, vol. 75, 2020.

[82] A. I. Mushlin, R. W. Kouides, and D. E. Shapiro, "Estimating the accuracy of screening mammography: A meta-analysis," *American Journal of Preventive Medicine*, vol. 14, no. 2, pp. 143–153, 1998.

Emily Porter Emily Porter received her B.Eng., M.Eng., and Ph.D. degrees in electrical engineering from McGill University, Montreal, QC, Canada, in 2009, 2010, and 2015, respectively. She is currently an Assistant Professor with the University of Texas at Austin. Her research interests include the measurement of dielectric properties of biological tissues and the development novel technologies for therapeutic and diagnostic applications of electromagnetic waves. In 2021, she received the IEEE Lot Shafai Mid-Career Distinguished Achievement Award for her research on these topics.



Declan O'Loughlin Declan O'Loughlin (M'15) received the BE and PhD degrees in Electronic and Electrical Engineerng from NUI Galway, Ireland in 2014 and 2019 respectively. His work focuses on the biomedical uses of electromagnetics for non-invasive sensing and monitoring. He is currently an Assistant Professor in Electronic and Electrical Engineering at Trinity College Dublin. He is a Senior Member of the International Radio-Science Union (URSI) and received the Young Scientist Award in 2020, and is a Member of the Institute of Engineers of Ireland.