Comparison of Microwave Breast Cancer Detection Results with Breast Phantom Data and Clinical Trial Data: Varying the Number of Antennas

Yunpeng Li, Adam Santorelli, Mark Coates

Dept. of Electrical and Computer Engineering, McGill University, Montréal, Québec, Canada

yunpeng.li@mail.mcgill.ca, adam.santorelli@mail.mcgill.ca, mark.coates@mcgill.ca

Abstract—Over the last two decades, microwave breast cancer detection has attracted an increasing amount of research attention and various research groups have recently commenced clinical testing. Much of the past assessment of algorithms and hardware has relied on breast phantoms. We illustrate in this paper through experimental data collected with breast phantoms that classification with breast phantoms can be a much easier task than that with clinical trial data (with numerically-simulated tumour responses). One antenna pair is sufficient to achieve almost perfect classification on the phantom data but has little success with the clinical trial data. The results demonstrate that we should exercise caution when evaluating classifier performance based solely on breast phantoms, and highlight the importance of validating microwave breast cancer detection algorithms with clinical trial data.

Index Terms—microwave breast cancer detection, clinical trial, breast phantom, ensemble classifier.

I. INTRODUCTION

Microwave breast cancer detection techniques have been proposed as a complementary modality for breast cancer screening. They promise non-invasive screening with low-cost system fabrication and operation. These scans are conducted without breast compression and can be repeated frequently since no ionizing radiation is used.

Research on the signal processing aspect of microwave breast cancer detection has primarily focused on generating images or building detection techniques for specific system setups. Numerous imaging approaches [1], [2], [3], [4] were initially validated with numerical breast models. Subsequently, research groups across the globe built a variety of microwave prototypes with different configurations. A 16-element monopole antenna clinical prototype was built for tomographic imaging at Dartmouth College [5]. At University of Bristol, multiple generations of microwave systems with 16, 31, and 64-element antenna arrays were constructed for phantom testing and clinical trials [6], [7]. A 16-antenna time-domain microwave system [8] was developed at McGill University, tested with breast phantoms and clinical trials. Instead of being equipped with an increasing number of antennas, some recent microwave prototypes were built with one monostatic transceiver [9], [10] or one antenna pair [11]. In the last decade, microwave breast cancer detection algorithms based on machine learning techniques were developed. But

they have been mainly evaluated using experimental data derived using breast phantoms [12], [13].

Because different teams developed their own prototypes with different system characteristics, it is difficult to compare results from different groups. [9] reports 90% accuracy on benign and malignant tumour classification with a set of breast and tumour phantoms and a single monostatic antenna, and [11] reports close to 100% tumour detection rate with two antennas. Two questions arise naturally: (i) do the phantom-based experiments reported in [9], [11] sufficiently represent a clinical scanning scenario to provide a meaningful test of a classification technique? (ii) is it really necessary to develop systems with more than 2 antennas? The answers to these questions are important. Using a smaller number of antennas leads to more cost-effective equipment solutions with smaller form factors. If the classification task with breast phantoms is too simple compared to the task on clinical trial data, then we should strive to develop more realistic phantoms, devise a different experimental procedure, or interpret phantom-based results with care and conduct algorithm validation using clinical trial data sets.

Recently, together with the other members of the McGill University microwave breast cancer detection research group, we conducted a set of experiments involving both breast phantoms and clinical trials. The experimental methods and primary analysis of the results were reported in [14], [15]. For the experiments with breast phantoms, we made the data acquisition process as close to the practical clinical trial as possible, by collecting different scans on different days, and removing and re-inserting phantoms for each scan. This is in contrast to the fixed phantom position scenario used in the experiments reported in [9], [11]. The clinical trial data were collected over a 8-month span [14]. Each volunteer was scanned multiple times, with a minimal interval between scans of one month. This data was collected from healthy patients, so in order to use it for the evaluation of classifier algorithms, we incorporate numerically-simulated tumour responses.

In light of the experimental results reported in [9], [11], for this paper we revisited our experimental data. We examined how the performance of classification algorithms changes when they process measurements from only subsets of the 16 antennas available in the microwave system. This allows us to assess how many antennas are really necessary before
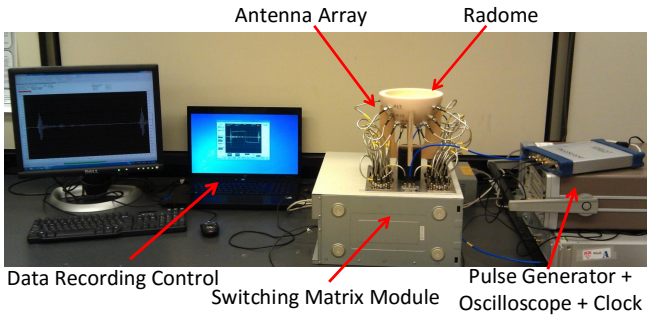
Fig. 1. The experiment system we use to collect measurements.

classification becomes viable. We compare the results for the phantom data with those from the clinical data in order to assess whether the phantom-based testing is adequately representing the clinical measurement setting.

The rest of the paper is organized as follows: Section II presents the microwave system, and the data acquisition process. Detection algorithms are presented in Section III. We present and discuss experimental results in Section IV. Concluding remarks are made in Section V.

II. SYSTEM OVERVIEW AND DATA COLLECTION

A. System overview

The system (Figure 1) uses a 16-element antenna array to transmit and record pulses to and from the breast. The antennas are resistively loaded sensors which are designed for microwave breast cancer detection [16]. The antenna array is embedded on the hollowed-out dielectric radome which also houses the breast or the breast phantom. When a breast scan starts, a short-duration Gaussian-modulated pulse concentrated in the 2-4GHz range [17] is propagated into the breast and collected by a receiving antenna specified by a switching matrix. An equivalent-time sampling oscilloscope records and saves the pulse to a computer. This process is repeated for each antenna pair – a total of 240 pulses are obtained for one breast scan.

B. Breast phantom data set

We constructed 9 breast phantoms with varying dielectric properties. Three are heterogeneous and contain mixtures that mimic the dielectric properties of skin, gland, fat and tumour. The construction process is described in [15]. Those three phantoms are rotated by $120°$ and $240°$ to mimic 6 new phantoms, as different structures are presented to the measurement system after the rotation. Thus, we have 15 phantoms in total. 14 of these phantoms contain a plug position so we can insert either the fat plug or the tumour plug to mimic the tumour-free or the tumour-bearing breasts. We collected 10 sets of scans on tumour-free phantoms and 10 sets of scans on tumour-bearing phantoms. Each phantom is removed from the radome once a scan is completed. Scans on the same phantom are performed in different days with the system being re-initialized each day. These steps are used to introduce variations common in a real clinical trial data collection process.

C. Clinical trial data set

We performed breast scans on 12 healthy volunteers over an 8 month span. Their ages ranged from 21 to 77, and the bra cup sizes varied from A to D. Each volunteer visited a minimum of two and a maximum of six times, with the visit interval being at least one month. As scans are performed on both breasts, a total of 96 scans were collected in 48 volunteer visits.

Since all volunteers are healthy, we insert numerically simulated tumour responses into the clinical trial data set. The detailed description of the data-adaptive tumour response construction that factors in the heterogeneous propagation environment inside the breast is described in [18]. Tumour responses with attenuation factor 1 are inserted in the data set we analyze in this paper.

III. CLASSIFICATION METHODS

In this section, we describe the classification algorithms that we compare, and provide details concerning the antenna configurations that are examined.

1) Imaging-based detection algorithm: The delay-multiply-and-sum (DMAS) algorithm [2] is a popular method to generate images from microwave breast cancer screening data. It first extracts backscatter from pulses received in each antenna pair. For each voxel position, the time delays with each antenna pair are estimated and the time-aligned backscatter are multiplied in a pairwise fashion. This is followed by a summing and integration operation which outputs the image intensity at each voxel. We proposed in [19] to use the maximum image intensity as the classifier input for the detection task. The maximum image intensity is compared with a threshold to decide whether a tumour is present.

2) Ensemble selection algorithm: The ensemble selection algorithm for microwave breast cancer detection is proposed in [18]. Principal component analysis (PCA) is first applied to measurements from each antenna pair for feature extraction. 2μ-support vector machines (SVM) [20] with different hyper-parameter values are then trained on the extracted features from each antenna pair. These trained "base" models form a model "library" – the base models include 2ν-SVMs trained using different antenna pairs and those for the same antenna pair but with different hyper-parameter values. The ensemble selection algorithm selects the best base models by evaluating the achieved cross validation error, to create an ensemble of more effective models. They are used to classify the test data and reach a final detection decision using a majority vote across the decisions of all model in the ensemble. The process is described in Figure 2.

3) Antenna configuration: Our current system consists of 16 antennas. To investigate the impact of the number of antennas on classification error, we perform classification using the measurements recorded by different subsets of the 16 antennas. Configurations with $16, 8, 4$ or 2 antennas are investigated. Experiments are designed so that different relative geometric relations between antennas are explored in different experiments.
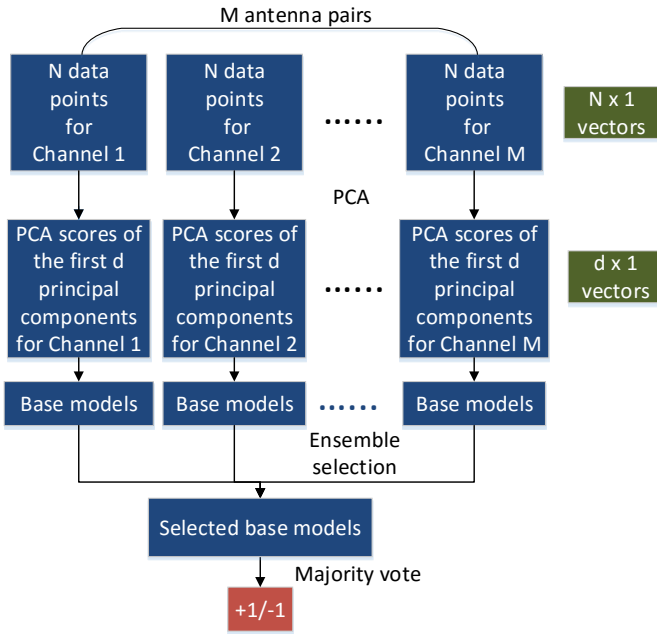
Fig. 2. The classification procedure based on ensemble selection.

To facilitate the grouping of antennas, we categorize antennas into four groups, using the antenna indices shown in Figure 3: $Z_1 = \{1, 5, 9, 13\}, Z_2 = \{2, 6, 10, 14\}, Z_3 = \{3, 7, 11, 15\}, Z_4 = \{4, 8, 12, 16\}$. Antennas in the same group have the same z coordinates. The antennas used in each experiment are listed in Table II.
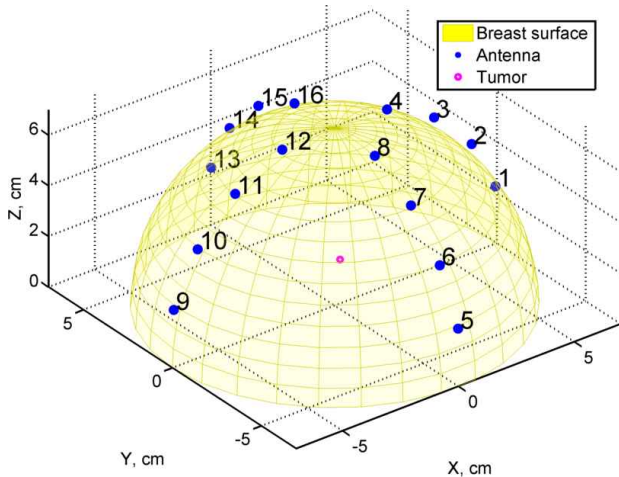

Fig. 3. Positions of antennas A1 to A16 (blue) correspond approximately to the locations in our experimental system.

IV. RESULTS

For the breast phantom data set, we use all of the data from one phantom as the test data; for the clinical trial data set, we use all data from one volunteer as the test data, We use all of the measurements from the other phantoms or volunteers as the training data. So, 15 training-testing pairs are constructed for the breast phantom data set, and 12 for the clinical trial data

set. The number of principal components retained is 30, and the number of best base models retained in ensemble selection is set to $Q = 100$.

We perform parameter selection for the 2ν-SVM using cross validation over values listed in Table I. The number of folds of cross validation is 14 for the breast phantom data set and 11 for the clinical trial data set, both being equal to the number of phantoms or volunteers in the training data set. Each fold contains all data from one breast phantom or one volunteer. We do not need to specify the value of the intensity threshold for the DMAS-based classification algorithm as we examine its performance when varying the value of the threshold.

TABLE I
CANDIDATE PARAMETER VALUES USED IN THE ENSEMBLE SELECTION
ALGORITHM.

γ	$2^{-15}, 2^{-13}, \ldots, 2^5$
ν_+	$1 \times 10^{-5}, 3 \times 10^{-5}, 1 \times 10^{-4}, 3 \times 10^{-4},$ $0.001, 0.003, 0.01, 0.03, 0.1, 0.2, 0.3 \ldots, 1$
ν_-	$1 \times 10^{-5}, 3 \times 10^{-5}, 1 \times 10^{-4}, 3 \times 10^{-4},$ $0.001, 0.003, 0.01, 0.03, 0.1, 0.2, 0.3 \ldots, 1$
r	$-0.4, -0.3, \ldots, 0.4$

Table II presents the area under curve (AUC) values of the average receiver operating characteristic (ROC) among all training-test pairs, for classification with different antenna configurations. An AUC with value 1 corresponds to perfect classification, while 0.5 indicates a performance similar to a random guess.

With breast phantoms, we observe very high detection rate using the ensemble selection algorithm for any combination of antennas tested in the experiment. When the number of antennas used is 2, ensemble selection algorithm does not select antennas since only one antenna pair is available. This shows that almost every antenna pair records measurements that contain enough information for the detection algorithm to make very accurate classification. The DMAS-based algorithm on breast phantoms has poor performance, and its AUC values are not significantly different than that of random guess for most antenna configurations tested. This is probably due to poor model matching and data alignment challenges. We have constructed heterogeneous phantoms, and attempted to make the data collection process similar to that of the clinical setting, by conducting different scans of the same phantom in different days, removing and replacing the breast phantoms from the radome, and re-initializing the microwave system before each day's collection. But even with all these efforts to mimic the complexities encountered when working with human patients, the data set with breast phantoms suggests that sufficient information is obtained by a single antenna pair for the ensemble selection algorithm to achieve almost perfect classification.

For the clinical trial data set, the ensemble selection algorithm achieves an AUC of 0.81 using all 16 antennas. Occasional decrease of detection performance is observed for antenna configurations with 4 or 8 antennas. The difference

TABLE II
AREA UNDER CURVE (AUC) FOR EACH ANTENNA COMBINATIONS

Experiment indice	# antenna	Antenna indices	Phantom data		Clinical trial data	
			ensemble selection	DMAS	ensemble selection	DMAS
1	16	$\{Z_1, Z_2, Z_3, Z_4\}$	0.99	0.52	0.81	0.56
2	8	$\{Z_1, Z_2\}$	0.99	0.68	0.81	0.54
3		$\{Z_1, Z_3\}$	0.99	0.64	0.81	0.57
4		$\{Z_1, Z_4\}$	0.99	0.53	0.86	0.60
5		$\{Z_2, Z_3\}$	1.00	0.72	0.74	0.49
6		$\{Z_2, Z_4\}$	0.99	0.36	0.78	0.46
7		$\{Z_3, Z_4\}$	0.99	0.36	0.56	0.43
8		$\{1,2,7,8,9,10,15,16\}$	0.99	0.49	0.87	0.54
9		$\{1,3,6,8,9,11,14,16\}$	0.99	0.63	0.85	0.58
10		$\{1,4,6,7,9,12,14,15\}$	0.99	0.67	0.78	0.56
11	4	Z_1	1.00	0.43	0.82	0.57
12		Z_2	0.99	0.49	0.76	0.47
13		Z_3	0.98	0.61	0.58	0.47
14		Z_4	0.98	0.63	0.52	0.42
15		$\{1, 6, 9, 14\}$	0.98	0.45	0.80	0.53
16		$\{1, 7, 9, 15\}$	0.99	0.35	0.82	0.52
17		$\{1, 8, 9, 16\}$	0.99	0.65	0.85	0.62
18	2	$\{1,2\}$	0.99	0.54	0.52	0.55
19		$\{2,3\}$	0.97	0.52	0.46	0.48
20		$\{3,4\}$	0.96	0.49	0.46	0.46
21		$\{1,3\}$	0.97	0.52	0.41	0.47
22		$\{2,4\}$	0.97	0.53	0.49	0.52
23		$\{1,4\}$	0.96	0.52	0.54	0.54

is not significant, possibly due to the ability of ensemble selection algorithm to select the most informative antenna pairs from all of the available antenna pairs. However, when we only have access to data from a single pair of antennas, the AUC decreases to around 0.5, indicating that the detection performance is similar to that of a random guess. This shows that at least four antennas are needed for good classification performance, and the performance becomes more reliable as we increase the number to 8 or 16.

V. CONCLUSION

In this paper, we investigated the classification performance with data sets collected using breast phantoms and from a preliminary clinical trial. We show that classification performance with the ensemble learning algorithm is very good with the breast phantoms even when only two antennas are used, despite our attempts to mimic the practical clinical data acquisition process. This indicates that the variations in data collection, tissue, and breast shape are more complex and varied than we can currently mimic with (our state-of-the-art) phantoms. The classification performance achieved in an experiment using breast phantoms may not correspond to that obtained in a clinical setting.

We consider that it is important to continue to develop breast phantom construction techniques and to carefully design phantom experiments to incorporate as many of the clinical challenges as possible. The recent work towards 3D-printable phantoms [21] is a promising step in this direction. Our analysis of the clinical trial data suggests that a single antenna pair is insufficient for effective classification, and that performance becomes more robust as the number of antennas is increased to sixteen, the maximum available in our system.

REFERENCES

[1] E. C. Fear, X. Li, S. C. Hagness, and M. A. Stuchly, "Confocal microwave imaging for breast cancer detection: localization of tumors in three dimensions," *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 812–822, Aug. 2002.

[2] H. B. Lim, N. T. T. Nhung, E.-P. Li, and N. D. Thang, "Confocal microwave imaging for breast cancer detection: Delay-multiply-and-sum image reconstruction algorithm," *IEEE Trans. Biomed. Eng.*, vol. 55, pp. 1697–1704, June 2008.

[3] E. J. Bond, X. Li, S. C. Hagness, and B. D. Van Veen, "Microwave imaging via space-time beamforming for early detection of breast cancer," *IEEE Trans. Antennas Propagat.*, vol. 51, pp. 1690–1705, Aug. 2003.

[4] M. O'Halloran, E. Jones, and M. Glavin, "Quasi-multistatic MIST beamforming for the early detection of breast cancer," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 830–840, Apr. 2010.

[5] P. Meaney, M. Fanning, D. Li, S. P. Poplack, and K. Paulsen, "A clinical prototype for active microwave imaging of the breast," *IEEE Trans. Biomed. Circuits Syst. IEEE Trans. Microw. Theory Tech.*, vol. 48, pp. 1841–1853, Nov 2000.

[6] M. Klemm, J. Leendertz, D. Gibbins, I. Craddock, A. Preece, and R. Benjamin, "Microwave radar-based differential breast cancer imaging: Imaging in homogeneous breast phantoms and low contrast scenarios," *IEEE Trans. Antennas Propag.*, vol. 58, pp. 2337–2344, July 2010.

[7] M. Klemm, D. Gibbins, J. Leendertz, T. Horseman, A. Preece, R. Benjamin, and I. Craddock, "Development and testing of a 60-element UWB conformal array for breast cancer imaging," in *Proc. European Conf. Antennas and Propag. (EuCAP)*, Apr. 2011, pp. 3077–3079.

[8] E. Porter, E. Kirshin, A. Santorelli, M. Coates, and M. Popović, "Time-domain multistatic radar system for microwave breast screening," *IEEE Antennas Wireless Propag. Lett.*, vol. 12, pp. 229–232, 2013.

[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. URSI General Assembly and Scientific Symposium (GASS)*, Beijing, China, Aug. 2014, pp. 1–4.

[10] E. Fear, J. Bourqui, C. Curtis, D. Mew, B. Docktor, and C. Romano, "Microwave breast imaging with a monostatic radar-based system: A study of application to patients," *IEEE Trans. Microw. Theory Tech.*, vol. 61, pp. 2119–2128, May 2013.

[11] S. A. AlShehri, S. Khatun, A. B. Jantan, R. S. A. R. Abdullah, R. Mahmud, and Z. Awang, "Experimental breast tumor detection using NN-based UWB imaging," *Prog. Electromagn. Res. (PIER)*, vol. 111, pp. 447–465, 2011.

[12] D. Byrne, M. O'Halloran, M. Glavin, and E. Jones, "Breast cancer detection based on differential ultrawideband microwave radar," *Prog. Electromagn. Res. (PIER)*, vol. 20, pp. 231–242, 2011.

[13] Y. Li, A. Santorelli, O. Laforest, and M. Coates, "Cost-sensitive ensemble classifiers for microwave breast cancer detection," in *Proc. Intl. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, Brisbane, Australia, Apr. 2015.

[14] E. Porter, M. Coates, and M. Popovich, "An early clinical study of time-domain microwave radar for breast health monitoring," *IEEE Trans. Biomed. Eng.*, vol. PP, no. 99, pp. 1–1, 2015, to appear 2015, EPUB ahead of print: http://www.ncbi.nlm.nih.gov/pubmed/26259214.

[15] 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 *European Conf. Antennas and Propag. (EuCAP)*, Lisbon, Portugal, Apr. 2015.

[16] H. Kanj and M. Popović, "A novel ultra-compact broadband antenna for microwave breast tumor detection," *Prog. Electromagn. Res.*, vol. 86, pp. 169–198, 2008.

[17] A. Santorelli, M. Chudzik, E. Kirshin, E. Porter, A. Lujambio, I. Arnedo, M. Popović, and J. D. Schwartz, "Experimental demonstration of pulse shaping for time-domain microwave breast imaging," *Prog. Electromagn. Res.*, vol. 133, pp. 309–329, 2013.

[18] Y. Li, E. Porter, A. Santorelli, M. Popovic, and M. Coates, "Microwave breast cancer detection via cost-sensitive ensemble classifiers: Phantom and patient investigation," submitted for publication.

[19] Y. Li, E. Porter, and M. Coates, "Imaging-based classification algorithms on clinical trial data with injected tumour responses," in *Proc. European Conf. Antennas and Propag. (EuCAP)*, Lisbon, Portugal, Apr. 2015.

[20] H.-G. Chew, R. E. Bogner, and C.-C. Lim, "Dual ν-support vector machine with error rate and training size biasing," in *Proc. Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, Salt Lake City, UT, May 2001, pp. 1269–1272.

[21] M. Burfeindt, T. Colgan, R. Mays, J. Shea, N. Behdad, B. Van Veen, and S. Hagness, "Mri-derived 3-d-printed breast phantom for microwave breast imaging validation," *IEEE Antennas Wireless Propag. Lett.*, vol. 11, pp. 1610–1613, 2012.