

BREAST CANCER DETECTION BASED ON DIFFERENTIAL ULTRAWIDEBAND MICROWAVE RADAR

D. Byrne^{*}, M. O'Halloran, M. Glavin, and E. Jones

College of Engineering and Informatics, National University of Ireland Galway, University Road, Galway, Ireland

Abstract—Ultrawideband (UWB) microwave imaging is a promising emerging method for the detection of breast cancer. Fibroglandular tissue has been shown to significantly limit the effectiveness of UWB imaging algorithms, particularly in the case of premenopausal women who may present with more dense breast tissue. Rather than trying to create an image of the breast, this study proposes to compare the UWB backscattered signals from successive scans of a dielectrically heterogeneous breast, to identify the presence of cancerous tissue. The temporal changes between signals are processed using Support Vector Machines to determine if a cancerous growth has occurred during the time between scans. Detection rates are compared to the results from a previous study by the authors, where UWB backscatter signals from a single scan were processed for cancer detection.

1. INTRODUCTION

The primary goal for any breast cancer screening methodology is to determine the presence of cancerous tissue at the earliest stage of development. In the U.S., the mortality rate for breast cancer has reduced by 28% between 1991 and 2006, which may be attributed to improvements in prevention, early stage detection and treatment [1].

UWB radar is a non-ionizing breast imaging technology which locates dielectric scatterers within the breast [2]. A UWB pulse is transmitted into the breast and electromagnetic reflections are generated, due to the dielectric contrast between certain tissues, particularly cancerous and fatty tissues. These reflections are recorded by the receiving antennas and processed using an image formation algorithm to create an energy profile of the breast [3–9].

Received 8 August 2011, Accepted 6 September 2011, Scheduled 8 September 2011

^{*} Corresponding author: Dallan Byrne (dallanbyrne@gmail.com).

Ex-vivo tissue samples examined by Lazebnik et al. [10] indicate the dielectric contrast between fibroglandular and cancerous tissues may no longer be sufficient to clearly identify cancerous regions within UWB images of a dielectrically heterogeneous breast [8, 9].

A recent study by the authors has demonstrated the potential of a Support Vector Machines (SVM) based cancer detection algorithm [11] which processes UWB microwave backscatter signals to determine if a cancerous region exists in the breast. While the algorithm has proven robust to variations in dielectric heterogeneity and tumor size, the detection process becomes much more difficult for smaller, non-palpable tumors. The method proposed in this paper seeks to address this problem.

Rather than examining the UWB backscattered signals from a single scan, the authors propose to use signal data taken from two consecutive scanning scenarios, to attempt to detect any changes in cancerous tissue. The majority of the reflections from the breast remain constant between the successive scans, so any significant change could potentially suggest the presence of cancer. 3D anatomically accurate Finite-Difference Time-Domain (FDTD) breast models are used to generate simulation signal data. Heterogeneous tissue density is varied within these models to evaluate the robustness of the SVM detection algorithm. This variation is based on the reported fluctuations in breast density between the luteal (days 15–28) phase and follicular (Days 1–14) phase of the menstrual cycle [12].

The paper is organised as follows: Section 2 describes the cancer detection algorithm. The details regarding the numerical breast model and the test setup are given in Section 3. Results are analysed and discussed in Section 4, and finally, conclusions are drawn in Section 5.

2. DETECTION ALGORITHM

This section outlines the UWB cancer backscatter detection algorithm. A differential signal is obtained from two consecutive UWB scans, principal components are extracted and applied to an SVM classifier.

2.1. Differential Signals

In order to detect a change in the composition of the breast (i.e., the growth of cancerous tissue), the difference between the UWB backscattered signals from two successive scans are compared. Let $X1_{(i,j)}$ represent the backscattered signal from the first scan, transmitted from antenna i and received at antenna j . Similarly, let $X2_{(i,j)}$ represent the corresponding multistatic [7–9] signal from the

second scan. The difference signal, $\delta_{(i,j)}$, can then be calculated as follows:

$$\delta_{(i,j)} = X1_{(i,j)} - X2_{(i,j)} \tag{1}$$

This process is illustrated in Figure 1, where the first breast model contains a 4 mm tumor (surrounded by two glandular tissue regions), while the second breast model illustrates tumour growth with increased volumes of fibroglandular tissue and a 10 mm cancerous inclusion.

As well as changes in tumor size, the volume of breast glandular tissue may increase or decrease during the menstrual cycle [12], generating unwanted microwave reflections. Difference signals are processed by an SVM classification algorithm which mitigates the effects of changes in the scanning environment and detects the growth of a cancerous region.

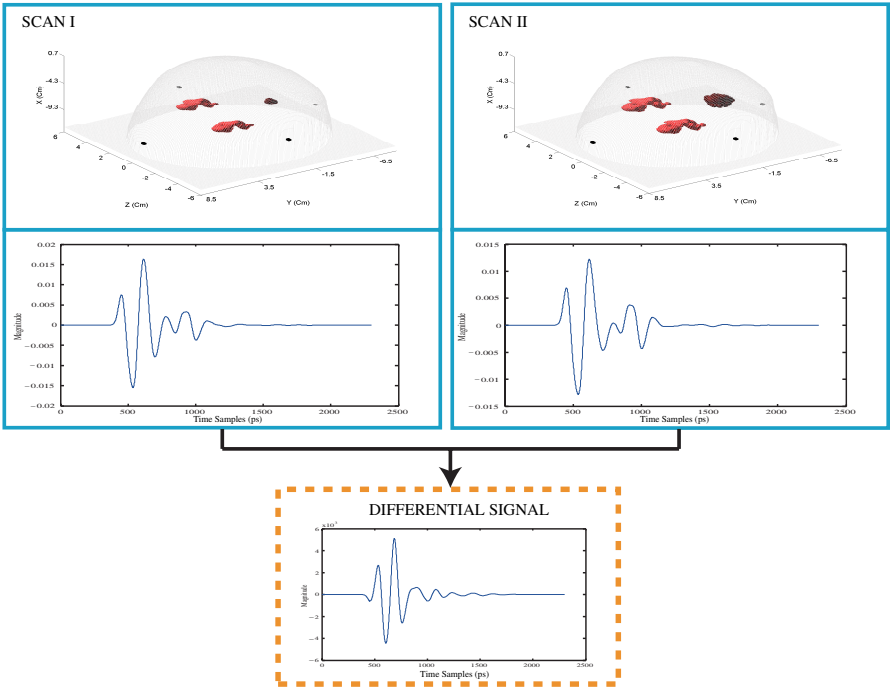


Figure 1. SVM-based Cancer Detection System. Two glandular pieces and a tumor located at $(-9.4\text{ cm}, -0.4\text{ cm}, -2.5\text{ cm})$ are illustrated in both scans. Lesion diameter is 4 mm in the first scan and 10 mm in the second, while glandular tissue volume increases by 120% between scans. Differential signals are created by comparing the UWB backscattered signals from two successive scans. Note the variations in scales used for the scan signals and the differential signal.

2.2. Support Vector Machines

The breast cancer detection system presented here is based on Support Vector Machines (SVM). An SVM is a learning algorithm, where training vectors are mapped to a high (or infinite) dimensional feature space labeled a hyperplane [13]. A discriminant plane is chosen to correctly classify the data into two distinguishable sets. The maximum distance between the parallel supporting planes is known as the margin. SVM maximizes the margin between the discriminant plane and the examples in the training set. In order to account for linearly inseparable cases, soft margins [14] are implemented.

Given a training set of sample-label pairs $(\mathbf{x}_i, y_i)$, with features $\mathbf{x}$, label y and $i = 1, \dots, M$, the Support Vector Machines are constructed from the following mathematical optimization procedure:

$$\begin{aligned} & \underset{\mathbf{w}, b, \xi}{\text{minimize}} && \left[\frac{1}{2}(\mathbf{w}^T \mathbf{w}) + C \sum_{i=1}^M \xi_i \right] \\ & \text{subject to} && y_i \left(\mathbf{w}^T \phi(\mathbf{x}_i) + b \right) \geq 1 - \xi_i, \\ & && \xi_i \geq 0. \end{aligned} \tag{2}$$

where $\mathbf{w}$ is the decision plane orientation vector, b is the bias, ξ_i represents the margin slack variable, ϕ is the mapping function and C is the penalty parameter of the error term. When a linear discriminant function cannot sufficiently separate each supporting plane, a quadratic function is used. To avoid over-fitting and excessive computation, each training vector is mapped to an infinite dimensional feature space by means of a kernel function: $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$. A Radial Basis Function (RBF) kernel is implemented in this study and is described as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp \left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2 \right), \quad \gamma > 0 \tag{3}$$

where γ is a kernel scaling factor. The RBF kernel implements nonlinear mapping to project samples into a higher dimensional space, and can emulate Linear and Sigmoid kernels using specific parameters [15]. Previous work by the authors has shown the RBF to be an effective kernel when partitioning UWB backscatter data [11].

2.3. Feature Extraction and Classifier Optimization

Principal Component Analysis (PCA) is applied to the difference signals to reduce the dimensionality of the data [16]. 50 components are chosen as this offered the best trade-off between computational load and performance [11]. Prior to SVM application, the dataset is

scaled between $[+1, -1]$ and the the penalty parameter (C) and RBF kernel scaling parameter (γ) are determined through a grid searching process [15].

2.4. Linear Discriminant Analysis

A Linear Discriminant Analysis (LDA) classifier is used as a baseline to examine the performance and robustness of the SVM classifier [16, 17]. LDA is an effective classification technique when the groups being discriminated have multivariate normal distributions and the same covariance matrix.

The pooled within-group covariance matrix is calculated and used to determine the discriminant function which will allow classification. The mean, or centroid for each group is obtained from the discriminant scores for all objects within a group, which represents the location of an object from a particular group. A linear separator is then derived based on the centroid and the distributions of the discriminant functions [16, 18].

3. EXPERIMENTAL SETUP

In this section, the details regarding the numerical breast phantoms, tissue heterogeneity and the test architecture are described.

3.1. Numerical Breast Model

The FDTD models of the breast used in this study are based on anatomically accurate MRI-derived phantoms [19]. Spatial dimensions within the breast model correspond to the three axes, where the X axis signifies the depth of the breast and the Y and Z represent the span and breadth of the breast respectively.

The volume of the modeled FDTD grid is approximately 3.3 million cubic cells with a spatial resolution of 1 mm in all three dimensions (dx, dy, dz). The time step dt is 1.66 ps ($dx/2c$), where c represents the speed of light. A 12 layer Uniaxial Perfectly Matched Layer boundary condition [20] is used to minimize edge reflections.

The dispersive properties of breast tissue are incorporated into the FDTD model using a single-pole Debye model [21]. The dielectric properties of adipose and fibroglandular tissue are based on the results presented by Zastrow et al. [19]. Skin Debye parameters are obtained from data published by Gabriel et al. [22] while the tumor Debye parameters are taken from Shea et al. [23].

Tumor models are generated using the Gaussian Random Spheres (GRS) method [16, 24–26] to simulate realistic tumor shapes and

surface textures. Tumors with diameters of: 2 mm, 4 mm, 5 mm, 10 mm and 15 mm are placed at three distinct locations within the FDTD breast model $((-9.4 \text{ cm}, 0 \text{ cm}, 0 \text{ cm}), (-9.4 \text{ cm}, -0.4 \text{ cm}, -2.5 \text{ cm})$ and $(-8.4 \text{ cm}, -4 \text{ cm}, 0.5 \text{ cm}))$.

Four multistatic half-wavelength dipole antenna elements are arranged on two elliptical rings on the surface of the breast. Two antennas are located on each ring, adjacent to each other, with a uniform spacing of 20 mm between the two rings along the X axis. The input pulse is a 120 ps differentiated Gaussian pulse, with a center frequency of 3.8 GHz and a -3 dB bandwidth of 9 GHz. Prior to any signal processing, all FDTD signals are downsampled to 50 GHz.

3.2. Modeled Tissue Heterogeneity

Dielectric heterogeneity is incorporated in the numerical breast phantoms using several pieces of fibroglandular tissue at varying locations within the breast volume. Four distinct pieces of fibroglandular tissue are used, extracted from various MRI-derived breast phantoms which are taken from the UWCEM breast phantom repository, University of Wisconsin-Madison [19]. Each breast model contains either one or two pieces of fibroglandular tissue, where pieces are arranged in one of three positions, surrounding the tumor inclusion site. The remaining breast volume contains three variations of adipose (fatty) tissue. The modeled breast accounts for increased tissue density by rescaling the fibroglandular regions to 110% and 120% of their original size. Figure 2 illustrates the breast model with a tumor site at $(-9.4 \text{ cm}, 0 \text{ cm}, 0 \text{ cm})$ and various configurations of fibroglandular tissue regions.

3.3. Test Methodology

A total number of 2,880 UWB differential signal components are used to evaluate the SVM-based UWB breast cancer detection system. Half of the signals represent scenarios where cancerous and glandular growth has occurred in the time between scans, while the remainder represent scenarios where no tumor is present in either scan (but changes in glandular tissue composition occur). The test and training sets are allocated for each test and applied to an SVM classifier, to detect the presence of cancer growth. Results are compared to published cancer detection rates, where the dataset consisted of UWB backscattered data acquired from a single scan [11].

For each experiment, the dataset is shuffled and a test feature set, consisting of 25% of the entire set of features, is extracted. The remaining 75% is used to train the classifier. Features derived using

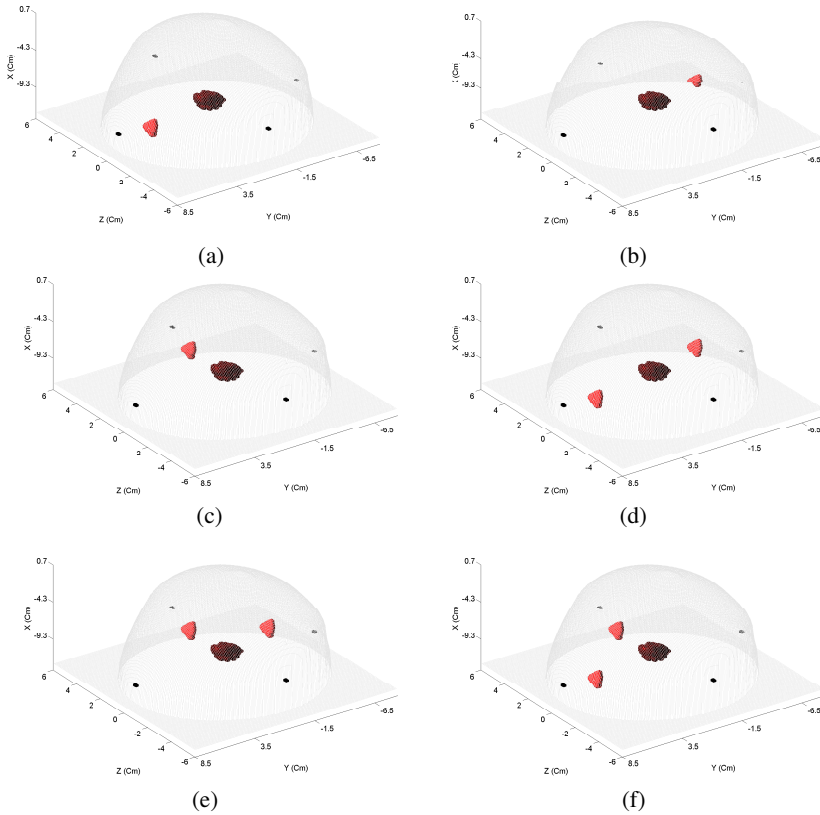


Figure 2. Breast tissue model with a malignant tumor inclusion at $(-9.4\text{ cm}, 0\text{ cm}, 0\text{ cm})$. A single piece of fibroglandular tissue in three positions is illustrated in (a), (b) and (c), while two fibroglandular regions are shown in (d), (e) and (f).

PCA (Section 2.3) are used for training and testing, and detection rates are obtained from the trained SVM classifier. This process is repeated over twenty iterations and a mean accuracy is obtained. The three experiments carried out are described as follows:

- (i) Overall detection performance.
The entire database is shuffled and test and training sets are extracted.
- (ii) Effect of heterogeneity on detection.
In order to examine the robustness of the SVM-based UWB breast cancer detection system to increasing levels of dielectric heterogeneity, two separate test sets are established, using data from successive scans which contain:

- only one piece of fibroglandular tissue (Hetero 1),
- two pieces of fibroglandular tissue (Hetero 2).

Training data is composed of equal distributions of Hetero 1 and Hetero 2 differential signal components.

(iii) Effect of tumor size on detection.

In order to examine the robustness of the SVM-based UWB breast cancer detection system to tumor size, test sets are extracted which are derived from consecutive scans between a healthy (no tumor) breast and a breast containing a tumor of specified size. Training sets are composed of consecutive scan data between a healthy breast and a breast which can contain any size of tumor bar the diameter specified for the test set.

4. RESULTS

4.1. Overall Detection Performance

Mean detection results are shown in Table 1. The differential SVM detection algorithm achieves an accuracy of 94.95%, 29.98% higher than the LDA method, and 12.83% more accurate than the equivalent test using data from the ‘Single Scan’ tests. The ‘Single Scan’ signals contain significant levels of noise, generated from skin and fibrous region reflections, whereas the detection of cancerous reflections within the difference signals is based on subtle changes in the shape of the scattering media within the breast.

4.2. Effect of Heterogeneity on Detection

Mean detection results are presented in Table 2. The differential SVM method has an accuracy of 96.32% for the Hetero 1 test scenario. Once a second glandular piece is introduced to the model, the performance of the algorithm drops to 93.04%. The increased number of glandular regions result in additional scattering reflections and unwanted noise in the resultant differential signals. This is a relatively small deterioration in the differential SVM detection scores, compared to corresponding

Table 1. Comparison of LDA and SVM breast cancer detection results.

Dataset	Differential LDA (%)	Single Scan SVM (%)	Differential SVM(%)
Full Dataset	64.97	83.66	94.95

Table 2. Effects of dielectric heterogeneity on breast cancer detection algorithm.

Dataset	Differential LDA (%)	Single Scan SVM (%)	Differential SVM (%)
Hetero 1	65.21	75.29	96.32
Hetero 2	65.74	71.13	93.04

Table 3. Effects of tumor size on breast cancer detection algorithm.

Dataset	LDA (%)	SVM (%)
15 mm	76.19	95.63
10 mm	72.61	96.03
5 mm	52.38	96.42
4 mm	55.95	96.82
2 mm	57.14	94.98

tests for the ‘Single Scan’ algorithm, which degrade 4.2% once the level of heterogeneity is increased. The LDA scores do not vary, but performance is well below ‘Single Scan’ SVM levels, finding just over 65% of tumors for both test cases.

4.3. Effect of Tumor Size on Classification

Mean results are shown in Table 3. The differential SVM method again proves robust against decreasing tumor diameters, finding 95.63% of 15 mm and 94.98% of 2 mm tumors. The highest score is achieved for the 5 mm tumor test sets, with an accuracy just under 1% higher than the 15 mm test set. Although this difference can be interpreted as statistically insignificant, it may be due to the proximity of the 15 mm tumor to the glandular tissue sites which can cause difficulties in discerning between tumor and glandular backscatter within the differential signal components. LDA struggles with tumor diameters of 5 mm and below, where the scores fall to between 50% and 60%. The smallest lesion dataset (2 mm) present the poorest ‘Single’ SVM score of 54.85%. The subtle changes in the ‘Single Scan’ backscatter due to the presence of a small lesion may not be sufficient to accurately diagnose the presence of cancer. This change is obviously adequate for the temporal algorithm to discern the influence of noise due to variations in glandular volume.

5. CONCLUSIONS

This study has examined a differential SVM-based breast cancer detection method which uses signal data taken from two consecutive UWB microwave scanning scenarios. Various potential clinical scenarios are simulated which represent tumor and connective tissue growth within the breast using dielectrically and anatomically representative FDTD models. Differential signal components, containing the energy difference between these scans, are used to train and test an SVM learning algorithm to detect cancerous growth within the breast.

The algorithm has been shown to be robust against increasing levels of heterogeneity and cancer growth, detecting tumors as small as 2 mm in diameter with an accuracy of over 94%. The SVM can partition the feature space using soft margins and an RBF kernel to adequately train itself against these difficult test feature sets.

Future work will investigate the introduction of further noise on the differential signal components, due to increased fibrous growth. A future study will also investigate the effects of mismatched antenna locations between scans, which can occur in a clinical environment. Such displacement could manifest as artefacts in the differential signals, and may impede cancer detection rates.

REFERENCES

1. Jemal, A., R. Siegel, J. Xu, and E. Ward, "Cancer statistics," *CA: A Cancer Journal for Clinicians*, Vol. 60, No. 5, 277–300, 2010.
2. Hagness, S. C., A. Taflove, and J. E. Bridges, "Two-dimensional FDTD analysis of a pulsed microwave confocal system for breast cancer detection: Fixed focus and antenna array sensors," *IEEE Transactions on Biomedical Engineering*, Vol. 45, 1470–1479, 1998.
3. Guo, B., Y. Wang, J. Li, P. Stoica, and R. Wu, "Microwave imaging via adaptive beamforming methods for breast cancer detection," *Journal of Electromagnetic Waves and Applications*, Vol. 20, No. 1, 53–63, 2006.
4. Flores-Tapia, D., M. O'Halloran, and S. Pistorius, "A bimodal reconstruction method for breast cancer imaging," *Progress In Electromagnetics Research*, Vol. 118, 461–486, 2011.
5. Alshehri, S. A. and S. Khatun, "UWB imaging for breast cancer detection using neural network," *Progress In Electromagnetics Research C*, Vol. 7, 79–93, 2009.

6. Davis, S. K., E. J. Bond, X. Li, S. C. Hagness, and B. D. Van Veen, "Microwave imaging via space-time beamforming for the early detection of breast cancer: Beamformer design in the frequency domain," *Journal of Electromagnetic Waves and Applications* Vol. 17, No. 2, 357–381, 2003.
7. O'Halloran, M., M. Glavin, and E. Jones, "Channel-ranked beamformer for the early detection of breast cancer," *Progress In Electromagnetics Research*, Vol. 103, 153–168, 2010.
8. Byrne, D., M. O'Halloran, M. Glavin, and E. Jones, "Data independent radar beamforming algorithms for breast cancer detection," *Progress In Electromagnetics Research*, Vol. 107, 331–348, 2010.
9. Byrne, D., M. O'Halloran, E. Jones, and M. Glavin, "Transmitter-grouping robust capon beamforming for breast cancer detection," *Progress In Electromagnetic Research*, Vol. 108, 401–416, 2010.
10. Lazebnik, M., C. Zhu, G. M. Palmer, J. Harter, S. Sewall, N. Ramanujam, and S. C. Hagness, "Electromagnetic spectroscopy of normal breast tissue specimens obtained from reduction surgeries: Comparison of optical and microwave properties," *IEEE Transactions on Biomedical Engineering*, Vol. 55, No. 10, 2444–2451, Oct. 2008.
11. Byrne, D., M. O'Halloran, M. Glavin, and E. Jones, "Support vector machine-based ultrawideband breast cancer detection system," *Journal of Electromagnetic Waves and Applications*, Vol. 25, No. 13, 1807–1816, 2011.
12. White, E., P. Velentgas, M. T. Mandelson, C. D. Lehman, J. G. Elmore, P. Porter, Y. Yasui, and S. H. Taplin, "Variation in mammographic breast density by time in menstrual cycle among women aged 40-49 years," *Journal of the National Cancer Institute*, Vol. 90, No. 12, 906–910, 1998.
13. Bennett, K. P. and C. Campbell, "Support vector machines: Hype or hallelujah?" *SIGKDD Explorations Newsletter*, Vol. 2, No. 1, 1–13, 2000.
14. Cortes, C. and V. Vapnik, "Support-vector networks," *Machine Learning*, 273–297, 1995.
15. Hsu, C.-W., C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," April, 2010, [Online], Available: <http://www.csie.ntu.edu.tw/~cjlin/papers.html>.
16. Conceição, R. C., M. O'Halloran, E. Jones, and M. Glavin, "Investigation of classifiers for early-stage breast cancer based on radar target signatures," *Progress In Electromagnetics Research*, Vol. 105, 295–311, 2010.

17. O'Halloran, M., B. McGinley, R. C. Conceição, F. Morgan, E. Jones, and M. Glavin, "Spiking neural networks for breast cancer classification in a dielectrically heterogeneous breast," *Progress In Electromagnetics Research*, Vol. 113, 413–428, 2011.
18. Hair, J., W. Black, B. Babin, R. Anderson, and R. Tatham, *Multivariate Data Analysis*, 6th Edition, Prentice Hall, Upper Saddle River, NJ, 2006.
19. Zastrow, E., S. K. Davis, M. Lazebnik, F. Kelcz, B. D. Van Veen, and S. Hagness, "Development of anatomically realistic numerical breast phantoms with accurate dielectric properties for modeling microwave interactions with the human breast," *IEEE Transactions on Biomedical Engineering*, Vol. 55, No. 12, 2792–2800, Dec. 2008.
20. Sacks, Z., D. Kingsland, R. Lee, and J. Lee, "A perfectly matched anisotropic absorber for use as an absorbing boundary condition," *IEEE Transactions on Antennas and Propagation*, Vol. 43, No. 12, 1460–1463, 1995.
21. Lazebnik, M., M. Okoniewski, J. H. Booske, and S. C. Hagness, "Highly accurate Debye models for normal and malignant breast tissue dielectric properties at microwave frequencies," *IEEE Microwave and Wireless Components Letters*, Vol. 17, No. 12, 822–824, Dec. 2007, (fitting Debye to the Cole cole models).
22. Gabriel, C., S. Gabriel, and E. Corthout, "The dielectric properties of biological tissues: I. literature survey," *Phys. Med. Biol.*, Vol. 41, No. 11, 2231–2249, Nov. 1996.
23. Shea, J. D., P. Kosmas, B. D. Van Veen, and S. C. Hagness, "Contrast-enhanced microwave imaging of breast tumors: A computational study using 3D realistic numerical phantoms," *Inverse Problems*, Vol. 26, No. 7, 1–22, 2010.
24. Conceição, R. C., M. O'Halloran, M. Glavin, and E. Jones, "Support vector machines for the classification of early-stage breast cancer based on radar target signatures," *Progress In Electromagnetics Research B*, Vol. 23, 311–327, 2010.
25. Davis, S. K., B. D. Van Veen, S. C. Hagness, and F. Kelcz, "Breast tumor characterization based on ultrawideband backscatter," *IEEE Transactions on Biomedical Engineering*, Vol. 55, No. 1, 237–246, 2008.
26. Muinonen, K., "Introducing the gaussian shape hypothesis for asteroids and comets," *Astronomy and Astrophysics*, Vol. 332, 1087–1098, 1998.