IEEE Xplore

Notice to Reader

"*Breast tumor detection using empirical mode decomposition features*,"
by Hongchao Song, Aidong Men, Zhuqing Jiang,
published in ***IEEE Access*** **Early Access**

Digital Object Identifier: 10.1109/ACCESS.2017.2737633

It has been recommended by the authors and Editor-in-Chief of *IEEE Access* that the article should not be considered for citation purposes. It was later realized that simulation results were not valid for certain conditions and configurations. Therefore the reported results in this article might be unreasonable. Further research needs to be conducted.

We regret any inconvenience this may have caused.

Michael Pecht
Editor-in-Chief
IEEE Access

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2737633, IEEE Access

Breast tumor detection using empirical mode decomposition features

Hongchao Song, Aidong Men and Zhuqing Jiang

***Abstract*—Breast cancer is the second leading cause of cancer deaths among women worldwide. Microwave-based breast cancer detection has attracted increasing attention over the past two decades. Rather than recovering the image of the breast area and accurately determining the tumor location, machine-learning-based algorithms concentrate on detecting the existence of a tumor. Feature extraction is a key step in machine learning, and this step strongly impacts the final detection accuracy. Principal component analysis (PCA) is one of the most widely used feature extraction methods; however, PCA is negatively impacted by signal misalignment. This paper presents an empirical mode decomposition (EMD)-based feature extraction method that is more robust to signal misalignment. The statistical features are extracted from the decomposed subbands of the original signal. The experimental results obtained from clinical data indicate that the detection accuracy is improved by the combination of features from EMD and PCA.**

***Index Terms*—Microwave breast cancer detection, Empirical mode decomposition, Principal component analysis, Support vector machine.**

I. Introduction

A. Background and motivation

Breast cancer is the most commonly diagnosed cancer in women. According to statistics from the American Cancer Society (ACS), approximately 252,710 new cases of invasive breast cancer will be diagnosed in women, and approximately 40,610 women will die from breast cancer [1]. Statistics from the ACS also show that the 5-year survival rate of breast cancer is approximately 90% with early detection. Various techniques have been developed over the past decades for breast cancer detection, such as X-ray mammography, ultrasound imaging and magnetic resonance imaging (MRI). However, each of these techniques has some drawbacks. X-ray mammography is the standard clinical technique for breast cancer detection, but it requires uncomfortable compression and exposure to ionizing radiation [2]. This radiation can also increase the probability of developing cancer [3]. Furthermore, distinguishing the dense breast tissue from cancer is difficult using X-ray mammography because both of them appear as white regions in the mammogram. Although MRI can detect some cancers that are not observed in a mammogram, this technique also suffers from a high false positive rate and high cost. Ultrasound is inexpensive, but it is less accurate and can struggle to distinguish between benign and malignant lesions [4]. Microwave-based breast cancer detection is a potential alternative modality, and it has been intensively explored over the past two decades [5].

The motivation for microwave-based breast tumor detection is that significant differences in the dielectric properties exist between malignant and healthy breast tissues in the microwave frequency range [6]. The estimated malignant-to-normal breast tissue contrast is approximately 2:1 to 10:1 depending on the density of the normal tissue [3]. Machine learning algorithms are used to capture the differences that exist between normal and malignant tissues. Feature extraction is an important step in achieving good breast cancer detection performance. Misalignment among breast scans, which is caused by equipment movement and the system's intrinsic jitter in the clock and sampling oscilloscope, results in different features being extracted from different scans of the same patent. Therefore, creating features that are insensitive to the time shift existing among scans is important.

B. State-of-the-art

Researchers have investigated various algorithms for microwave-based breast cancer detection, and these algorithms can be grouped into two categories: imaging methods [7]–[9] and machine learning methods [10]–[12]. Imaging methods aim to reconstruct the image of the breast area and accurately determine the location of the tumor. Considering the differences in imaging theory, breast imaging algorithms fall into two categories: tomography [13], [14] and radar (backscatter) methods [15]–[17]. The goal of microwave breast tomography is to recover the dielectric properties of the entire breast area using inverse scattering algorithms. This is an ill-posed problem, leading to challenges in robust inference and a high computational cost [18]. Rather than recovering the dielectric properties of the entire breast, radar methods attempt to estimate the locations of significant microwave scatterers using the measured reflected signals. The scatterers generally indicate the region with significant contrast in microwave frequency, which is likely to be a malignant tissue [16], [17]. Alternative imaging algorithms, such as those based on Bayesian modeling and time reversal techniques, have also been proposed [19], [20].

Another direction is the application of machine learning techniques. Rather than identifying the location of a tumor, machine learning methods generally predict the existence of a tumor. The performance of machine learning algorithms is highly related to the quality of features and the classifier architecture. The feature extraction method should be robust to the characteristics of the original data. The most widely

Hongchao Song was with the Information and Telecommunication engineering University , Beijing University of Posts and Telecommunications, Beijing, China, e-mail: shch@bupt.edu.cn

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2737633, IEEE Access

used feature extraction method for breast cancer detection is principal component analysis (PCA), which reduces the data dimensionality by converting a set of observations of possibly correlated variables into a subset of values of linearly uncorrelated variables [21]–[23]. However, PCA is sensitive to the misalignment that commonly exists among recorded signals. Local discriminant basis (LDB) was used in [24], which seeks to find orthonormal principal components for the data with localized time-frequency characteristics. The discrete cosine transform (DCT) represents the original signal as a sum of infinitely many cosine basis functions in different frequencies with varying amplitudes, and DCT was adopted in [10].

C. Contribution

In this paper, we propose a novel method that combines EMD-based features with PCA-based features in microwave breast tumor detection and utilizes the strengths of both feature extraction approaches. In EMD-based feature extraction, several instinct mode functions (IMFs) are first obtained, and then different features are extracted from the IMFs. These features are combined with PCA scores to represent the input signal. The EMD-based feature extraction method focuses on the characteristics of a single measurement, whereas the PCA-based feature extraction method focuses on the variances among different samples. Thus, we are motivated to combine these two extraction methods. The performance of the proposed method is tested, and the results are discussed.

II. Methods

A. Feature extraction

A set of high-dimensional features can negatively affect the performance of pattern or image recognition systems. Too many features occasionally reduce the classification accuracy of the recognition system since some of the features may be redundant and non-informative [25]. The length of the recorded signal is 1024 in our dataset, and it is essential to reduce the dimension of the input signal.

1) Empirical mode decomposition: EMD is an iterative procedure, as described in Figure 1 [26]. During each iteration, a decomposition signal is constructed. This decomposition signal is identified as an IMF if it satisfies two conditions:

- *C1*: The difference between the number of extrema (maxima and minima) and the number of zero crossings must be no greater than one;
- *C2*: The local mean, defined as the mean of the upper and lower envelopes, must be zero.

The original signal can be represented as the sum of IMFs and the residual signal:

$$x = \sum_{k=1}^{K} d_k + r_K \tag{1}$$

where d_k represents the k-th IMF and r_K represents the residual signal in the K-th level.

An example of the EMD of an antenna pair recording is shown in Figure 2. The IMFs can be viewed as a set of components of the signal, and each IMF captures different time-scale characteristics and local structure [26].

Input: Signal x. Maximum number of IMFs: N_{IMF}.
1: Set $j = 0$ and $r_0 = x$. Set $z = x$.
2: **while** $j < N_{\mathrm{IMF}}$ **do**
3: Find all extrema of z.
4: Interpolate between minima (maxima) of z to obtain the lower (upper) envelope $e_{\min}$ ($e_{\max}$).
5: Compute the mean envelope $m = (e_{\min} + e_{\max})/2$.
6: Compute the IMF candidate $d_{j+1} = z - m$.
7: **if** d_{j+1} satisfies *C1* and *C2* **then**
8: Save d_{j+1};
9: compute the residual $r_{j+1} = x - \sum_{i=1}^{j} d_i$.
10: Set $z = r_{j+1}$.
11: **else**
12: Set $z = d_{j+1}$.
13: **end if**
14: **end while**

Fig. 1. Empirical mode decomposition

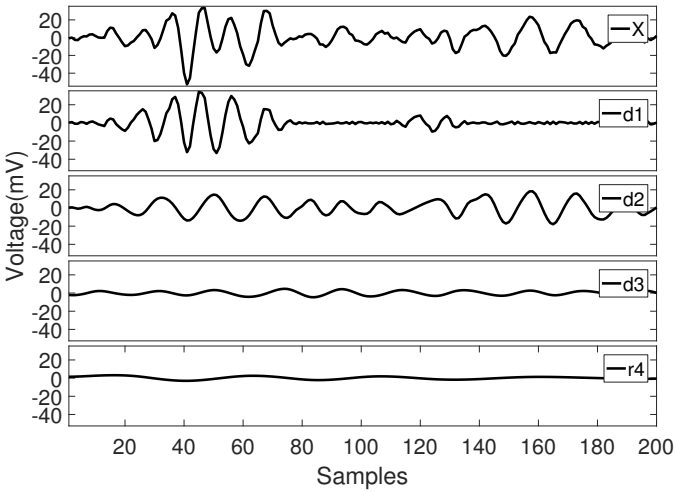


Fig. 2. Example of EMD. X is a healthy measurement collected by the antenna pair A16A15 from the left breast of volunteer two during her first visit, $d1 - d3$ are the first three IMFs (modes), and $r4$ is the residual.

2) Statistical features : EMD has been successfully applied in electroencephalography (EEG) signal processing [27], [28]. It decomposes the signal into different subbands according to the properties of the input signal without requiring the predefined parameters. The decomposition is completely data driven. EMD decomposes each signal without considering the difference among different signals, and it only relates to the envelopes and the number of zero-crossing points. If we shift a signal by a few time samples, the IMFs obtained from the shifted signal will be time-shifted copies of those obtained from the original signal.

We calculate the first 5 IMFs of the EMD for each measurement. For each IMF, 4 features are calculated. These features are the mean absolute value (μ), the standard deviation (σ), the kurtosis (κ) and the mean absolute successive difference (S). The mean absolute successive difference is used to measure the intensity of signal changes. The features for the k-th IMF,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2737633, IEEE Access

of length L time samples, are as follows:

1) Mean absolute value (μ_k):

$$\mu_k = \frac{1}{L}\sum_{i=1}^{L}|d_k(i)| \ , \tag{2}$$

2) Standard deviation (σ_k):

$$\sigma_k = \sqrt{\frac{1}{L}\sum_{i=1}^{L}(d_k(i) - \frac{1}{L}\sum_{i}^{L} d_k(i))^2} \, , \tag{3}$$

3) Kurtosis (κ_k):

$$\begin{aligned} M_k &= \frac{1}{L}\sum_{i=1}^{L} d_k(i) \\ \kappa_k &= \frac{\frac{1}{L}\sum_{i=1}^{L}(d_k(i) - M_k)^4}{(\frac{1}{L}\sum_{i=1}^{L}(d_k(i) - M_k)^2)^2} \end{aligned} \tag{4}$$

4) Mean absolute successive difference (S_k):

$$S_k = \frac{1}{L-1}\sum_{i=1}^{L-1}|d_k(i+1) - d_k(i)| \ . \tag{5}$$

The magnitudes of the features derived from the IMFs obtained later in the decomposition procedure are most often smaller. For the measurements that are collected by the same antenna pair, we rescale each feature of the training data into the range $[0, 1]$. The features of the testing data are scaled by the same ratio. The process is repeated for each antenna pair.

We also perform PCA on the signals derived from each antenna pair. Thirty principal scores are retained, as in [29]. The PCA scores are also normalized by scaling the first principal components to the range $[0, 1]$. The PCA scores of the other dimensions are rescaled by the same ratio. For the PCA features, a coefficient matrix is calculated from the training data. The training data are formed from the measurements collected by the same antenna pair from all volunteers, and both healthy measurements and tumor-response-injected measurements are included. In contrast, the EMD features only consider the single measurement itself. For combined features, each measurement is represented by a 50-dimensional feature vector that combines the 30 PCA features and the 20 EMD features.

B. Classification

We adopt the cost-sensitive ensemble selection-based classifier proposed in [29]. This classifier exhibits better classification performance than imaging-based classifiers and other ensemble classification structures. The ensemble selection-based classifier constructs base classifiers by selecting different antenna pairs and different hyperparameters in a cost-sensitive 2ν-support vector machine (SVM) [30]. The 2ν-SVM is formulated as follows:

$$\min_{w,b,\xi,\rho} \frac{1}{2}||w||^2 - \nu\rho + \frac{\omega_+}{n}\sum_{i\in I_+}\xi_i + \frac{1-\omega_+}{n}\sum_{i\in I_-}\xi_i \tag{6}$$

$$subject\ to \qquad y_i(k(w, x_i) + b) \le \rho - \xi_i \tag{7}$$

$$\xi_i \ge 0 \tag{8}$$

$$\rho \ge 0 \tag{9}$$

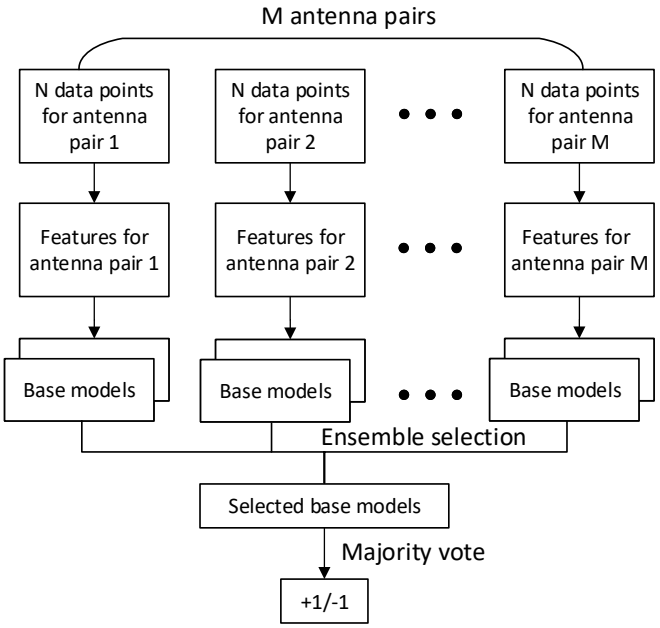


Fig. 3. The combined feature extraction and classification procedure.

where I_+ and I_- denote the set of data elements with labels $+1$ and -1, respectively. ω_+ controls the relative weight of the penalties. The parameters ν and ω_+ can be represented as follows:

$$\nu = \frac{2\nu_+\nu_-n_+n_-}{(\nu_+n_+ + \nu_-n_-)n} \tag{10}$$

$$\omega_+ = \frac{\nu_-n_-}{\nu_+n_+ + \nu_-n_-} = \frac{\nu n}{2\nu_+n_+} \tag{11}$$

where $\nu_+ \in [0, 1]$, $\nu_- \in [0, 1]$, and n_+ and n_- denote the numbers of positive and negative labeled data, respectively.

The hyperparameters ν_+ and ν_- control the importance associated with the two types of classification errors (false positive and false negative), while the hyperparameter γ is the kernel width of the SVM. Classifier pools with different types of errors are obtained by varying the values of hyperparameters. The model selection stage chooses the base models with the smallest Neyman-Pearson measure $\hat{e}$ [31].

$$\hat{e} = \frac{1}{\alpha}\max\{\hat{P}_F - \alpha, 0\} + \hat{P}_M \, , \tag{12}$$

where $\hat{P}_M$ is the empirical miss probability, $\hat{P}_F$ is the empirical false positive rate, and α is the upper bound of the target false positive rate. The classification algorithm is depicted in Figure 3.

C. Experimental data

The experimental data were collected using a multi-static radar system with $R = 16$ antennas [32] constructed by the research team at McGill University. For one scan, each of the R antennas transmitted an ultra-wideband pulse into the breast, and the other antennas recorded the backscattered signals. This process was repeated until one measurement was recorded by each of the $M = R(R-1) = 240$ different antenna pairs.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2737633, IEEE Access

The dataset was collected from 12 healthy volunteers. Table I lists the number of visits for each volunteer. The clinical trial lasted 8 months, and it involved 48 patient visits, with each volunteer visiting at most once per month [33]. Measurements of the left breast and the right breast from the same person were recorded during each visit. Thus, $48 \times 2 = 96$ sets of measurements were collected. Since all volunteers in the clinical trial were healthy, there were only tumor-free measurements in our original dataset. We adopt the method outlined in [29] to simulate tumor responses for each volunteer based on the transmitted pulses from the antennas and the dielectric properties of breast tissue. For one antenna pair and a tumor position p_0, the frequency-domain representation of the tumor response $R^t(p_0, \omega)$ is modeled as follows:

TABLE I
Number of visits for each volunteer

Volunteer index	1	2	3	4	5	6
Number of visits	3	3	4	5	2	6
Volunteer index	7	8	9	10	11	12
Number of visits	6	4	4	4	3	4

$$R^t(p_0, \omega) = \Gamma R(\omega) e^{-j(k_{im}(d^t_{im} - d^d_{im}) + k_{br}(d^t_{br} - d^d_{br}))}, \quad (13)$$

where $R(\omega)$ is the frequency-domain representation of the received signal and d^d_{im} and d^d_{br} are the lengths of the direct path for this antenna pair through the immersion medium (ultrasound gel) and breast tissue, respectively. d^t_{im} and d^t_{br} are the lengths of the shortest path between the antenna pair via the tumor position p_0 in the immersion medium and the breast tissue, respectively. Γ is a constant that can be used to introduce additional attenuation in the tumor response. In this paper, we concentrate on the case of $\Gamma = 1$. k_{im} and k_{br} are the wavenumbers for the immersion medium and breast tissue, respectively, and these have the following expressions:

$$k_{im}(\omega) = \frac{2\pi}{\lambda_{im}(\omega)} = \sqrt{\epsilon_{im}(\omega)}\frac{\omega}{c}, \quad (14)$$

$$k_{br}(\omega) = \frac{2\pi}{\lambda_{br}(\omega)} = \sqrt{\epsilon_{br}(\omega)}\frac{\omega}{c}. \quad (15)$$

Here, ϵ_{im} is the relative permittivity of the immersion medium, c is the speed of light, and ϵ_{br} is the average breast tissue relative permittivity. The latter is specified by the Debye model [34]:

$$\epsilon_{br}(\omega) = \epsilon_\infty + \frac{\Delta\epsilon}{1 + j\omega\tau} + \frac{\sigma_s}{j\omega\epsilon_0}, \quad (16)$$

where $\epsilon_0 = 8.854 \times 10^{-12}$ F/m is the permittivity of free space and ϵ_∞ is the dielectric constant of the material at infinite frequency. $\Delta\epsilon = \epsilon_s - \epsilon_\infty$, and ϵ_s is the static dielectric constant. The pole relaxation constant is τ, and the static conductivity is σ_s. The model parameters are chosen to approximate the dielectric properties of breast tissue.

Note that this is not a model that assumes homogeneous breast tissue. By including the *received* signal $R(w)$ rather than the transmitted signal, the model incorporates distortions and delays caused by inhomogeneous tissue. In Figure 4, we present one example of a tumor-response-injected signal.

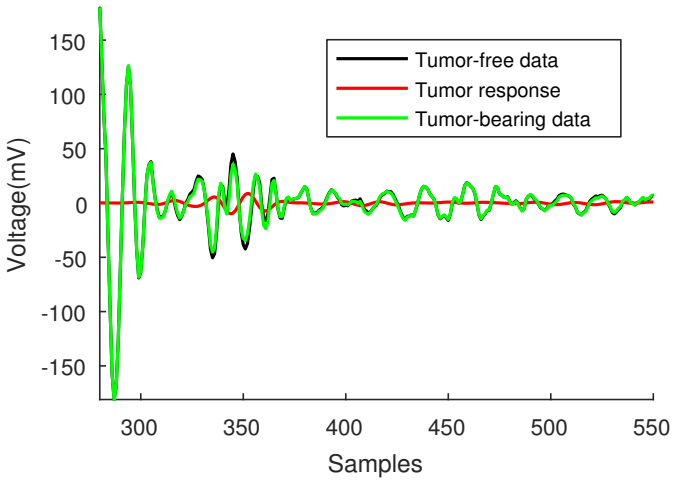


Fig. 4. An example of received pulses before and after tumor response injection. The pulse is collected by the antenna pair A1A2 from the left breast of volunteer one during her first visit.

To evaluate the classification performance with different feature extraction methods, we generated 50 datasets with different breast relative permittivity values and different tumor positions for the same volunteer. For each dataset, half of each volunteer's visits were randomly selected as tumor-bearing visits. If the number of visits N_j for patient j was an odd number, then $\frac{N_j - 1}{2}$ visits were selected as tumor-bearing visits. In these tumor-bearing visits, we randomly selected one breast as the tumor-bearing breast and randomly sampled a tumor position in the upper-quadrant region of the breast, closer to the armpit, which was reported to be the most likely breast tumor region [35]. The Debye model parameters were randomly generated for different volunteers. The breast permittivities were different for different volunteers.

For evaluating the classification performance, we divided each dataset into $\binom{12}{2} = 66$ training-testing pairs by using data from 10 volunteers to form the training data and data from 2 volunteers to form the testing data. To construct the testing data, we further added tumor-bearing scans for each tumor-free scan such that the ratio of tumor-free scans and tumor-bearing scans was 1:1 in the testing data. This process allowed us to more extensively test classification performance. For this purpose, we randomly generated tumor positions for each scan such that different tumor positions might exist for the same volunteer. However, we kept the same relative permittivity for the same volunteer in the training and testing sets. We did not add more tumor-bearing data in the training data; thus, the ratio of tumor-free scans and tumor-bearing scans in the training data was close to 3:1. This formed an imbalanced training dataset to train the classifier and mimicked the real scenario where the number of tumor-free scans is greater than the number of tumor-bearing scans. A graphical illustration for one training-testing pair of one dataset is provided in Figure 5.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2737633, IEEE Access

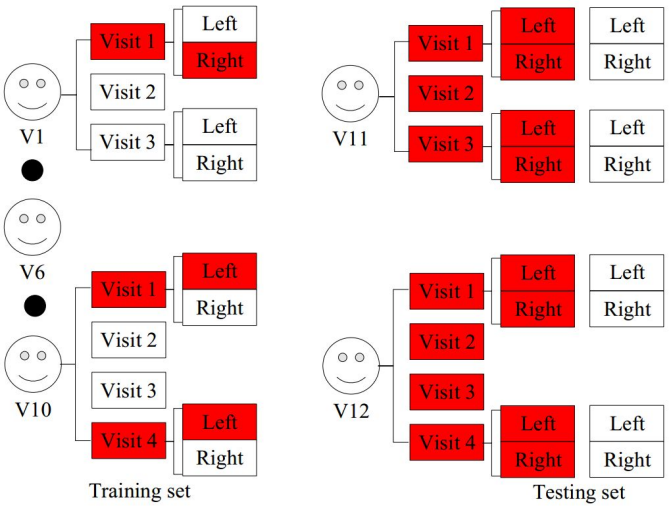


Fig. 5. One training-testing pair in which red blocks indicate randomly selected tumor visits and simulated tumor-bearing breasts, and white blocks indicate healthy visits and healthy breasts.

III. Experimental Results

A. Parameter settings

Signals collected by the antenna pairs located on the opposite sides of the breast can be highly distorted and varied significantly among different volunteer visits. Thus, we discarded signals from any antenna pair whose median peak amplitude of all the training data was less than a threshold of 20 mV. The length of the original signal was 1024, and the sample rate was 40 GHz [33]. We first windowed data from 60 samples before the peak amplitude of the measurement from antenna pair A1A2. Since antennas A1 and A2 were one of the closest antenna pairs, we did not expect any tumor responses to occur before the peak amplitude of the measurement from A1A2. Scans from the same antenna pair were then aligned based on maximal correlations with a reference scan and were windowed between the 61^{st} sample to the 604^{th} sample, corresponding to an actual time period of 13.6 ns. Excluding values outside this window reduced the noise during signal processing.

The candidate values of the 2ν-SVM hyperparameters used for cross validation are listed in Table II. We tested the detection performance with the γ value chosen from the candidate set $\gamma = \{2^{-15}, 2^{-13}, ..., 2^{5}\}$ using a small subset of the data, and we observed that the ensemble classifier almost always selected γ values of $\{2^{-1}, 2^{1}\}$. We further tested a range of fixed γ values between 2^{-2} and 2^{2} and observed that all values provided similar performance. To reduce the computational cost during training, we set $\gamma = 1$ for the experiments detailed here.

Thus, there were $1 \times 18 \times 18 = 324$ different 2ν-SVM hyperparameter combinations. These combinations were used to produce a model library consisting of $M \times 324$ base models, where M is the number of retained antenna pairs in each data set. The ensemble classifier selects 100 base models, choosing those with the smallest Neyman-Pearson measure when applied to the training data, to perform classification on the test data.

TABLE II
Candidate values of hyperparameters of 2ν-SVM in cross validation.

Parameter	Value
γ	1
ν_+	1×10^{-5},3×10^{-5},1×10^{-4},3×10^{-4}, $0.001, 0.003, 0.01, 0.03, 0.1, 0.2, 0.3, ..., 1$
ν_-	1×10^{-5},3×10^{-5},1×10^{-4},3×10^{-4}, $0.001, 0.003, 0.01, 0.03, 0.1, 0.2, 0.3, ..., 1$

B. Performance measurements

To measure the performance of the proposed method, the receiver operating characteristic curve (ROC curve) was adopted. The ROC curve was created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The TPR and FPR are formulated as follows:

$$FPR = \frac{FP}{FP + TN} \tag{17}$$

$$TPR = \frac{TP}{TP + FN} \tag{18}$$

where TP represents the number of correctly classified positive samples, TN represents the number of correctly classified negative samples, FP indicates the number of negative samples classified as positive, and FN is the number of positive samples classified as negative.

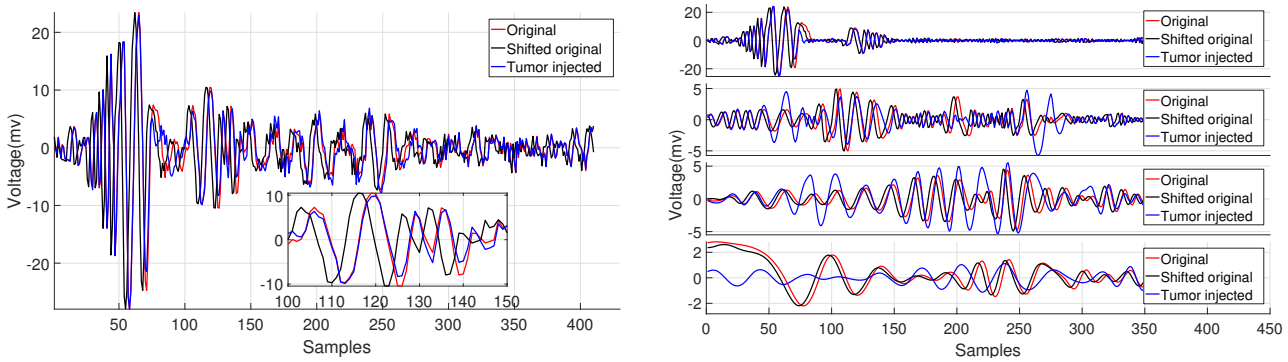
The area under the ROC curve, known as the AUC, was also calculated. The AUC of an ideal classifier is 1. A classifier with a higher AUC value generally performs better than those with lower AUC values.

C. Analysis of results

We first verify whether the EMD-based features are robust to signal misalignment. EMD is applied to individual signals; thus, time shifts in one signal due to misalignment only affect the IMFs of that signal. In addition, the IMFs derived by EMD are shift invariant in the sense that a shift in the input signal, provided that it is an integer number of samples, will result in an equivalent shift in the IMFs. The features that we have chosen to derive from them are insensitive to these shifts. An example is shown in Figure 6.

The EMD and PCA features extracted from the measurements in Figure 6a are shown in Figure 7. The figure shows that the PCA features extracted from the original measurement and those from the tumor-response-injected measurement are similar. The same observation applies for all other antenna pairs and tumor measurements. The classifier can perform substantially better if it is easier to discriminate between the features derived from the healthy and tumor-response-injected measurements. Indeed, for PCA, there appears to be a greater discrepancy between the features of the original and time-shifted signals. For the EMD feature extraction, there is less discrepancy between the features derived from the original

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2737633, IEEE Access



(a) Original data, time-shifted version and tumor-response-injected version

(b) Waveforms of original signal and IMF1-IMF3 (Top-Down)

Fig. 6. Comparisons of IMFs derived from different measurements, including the original measurement, the shifted original measurement and the tumor-response-injected measurement.

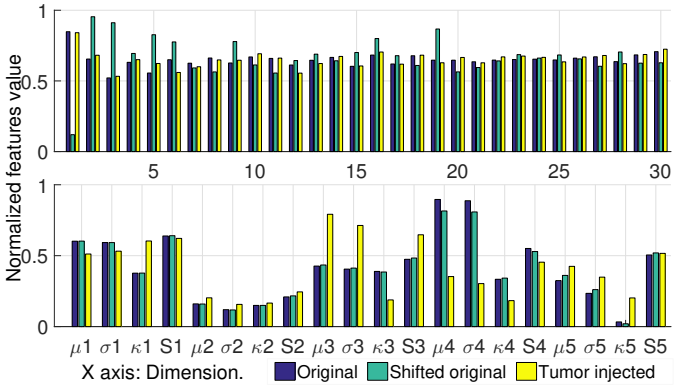


Fig. 7. Normalized EMD (bottom) and PCA features (top), where the X axis is the dimension index.

and time-shifted signals and a greater distinction between the original and tumor-response-injected measurements.

We evaluated the performance of the ensemble selection detection algorithm with EMD-extracted features, PCA scores, and combined feature vectors from EMD and PCA. The detection performance with different α values is shown in Table III. Irrespective of the features used, the ensemble selection detection algorithm almost always achieves an empirical false positive rate on the test dataset that lies below the target false positive rate specified by α. The use of combined features (EMD and PCA) leads to the smallest miss probability and also the smallest Neyman-Pearson error for all three values of α.

The upper bound of the target false positive rate in the training stage is controlled by α. By varying α over 101 evenly spaced values between 0 and 1, we obtain different false positive rates and true positive rates on the test data. These formed different operating points in the ROC curves. Figure 8 shows the ROC curves, averaged over the 50 datasets, for the three choices of features. The AUCs with EMD-extracted features, PCA scores and combined features are 0.83, 0.70, and 0.87, respectively.

We are also interested in examining whether the detection performance is consistent across volunteers. It is important

TABLE III
Average performance of three different types of features with different target false positive rates α. The 10% and 90% quantiles are shown in square brackets. The columns show the average false positive rate, the average false negative rate, the average error, and the average Neyman-Pearson error measure (Equation 12). Shaded entries indicate the smallest Neyman-Pearson error with the specific α.

α	Method	$\hat{P}_F$	$\hat{P}_M$	average error	$\hat{e}$
α=0.1	EMD	0.03 [0.00, 0.10]	0.65 [0.44, 0.86]	0.34 [0.25, 0.44]	0.70 [0.45, 0.93]
	PCA	0.01 [0.00, 0.07]	0.63 [0.42, 0.83]	0.32 [0.21, 0.42]	0.65 [0.43, 0.86]
	EMD PCA	0.02 [0.00, 0.10]	0.59 [0.38, 0.81]	0.31 [0.21, 0.42]	0.62 [0.39, 0.85]
α=0.3	EMD	0.09 [0.00, 0.22]	0.43 [0.21, 0.65]	0.27 [0.16, 0.37]	0.44 [0.22, 0.67]
	PCA	0.04 [0.00, 0.14]	0.57 [0.33, 0.79]	0.30 [0.18, 0.42]	0.57 [0.33, 0.78]
	EMD PCA	0.09 [0.00, 0.21]	0.40 [0.18, 0.60]	0.24 [0.14, 0.35]	0.40 [0.19, 0.62]
α=0.5	EMD	0.21 [0.00, 0.4]	0.27 [0.08, 0.50]	0.24 [0.14, 0.35]	0.28 [0.08, 0.5]
	PCA	0.14 [0.00, 0.36]	0.52 [0.29, 0.75]	0.33 [0.19, 0.46]	0.53 [0.30, 0.75]
	EMD PCA	0.18 [0.00, 0.36]	0.26 [0.07, 0.45]	0.22 [0.10, 0.32]	0.27 [0.07, 0.45]

to verify that the algorithms do not consistently fail to detect tumors for one of the volunteers. Figure 9 shows the detection error, averaged over the 50 constructed datasets, for three choices of α.

The final investigation that we conducted is to assess the performance of the combined features (EMD and PCA) on datasets with lower signal-to-noise ratios (SNRs). The lower SNR signals can be constructed by reducing the value of Γ in Equation (13). We generated 20 datasets with $\Gamma = 0.75$ and 20 datasets with $\Gamma = 0.5$. The ROC curves, averaged over the 20 datasets, are shown in Figure 10.

IV. Discussion

From the results reported in Section III, we observe that the use of EMD-extracted features results in significantly better detection performance compared to features based on PCA.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2737633, IEEE Access

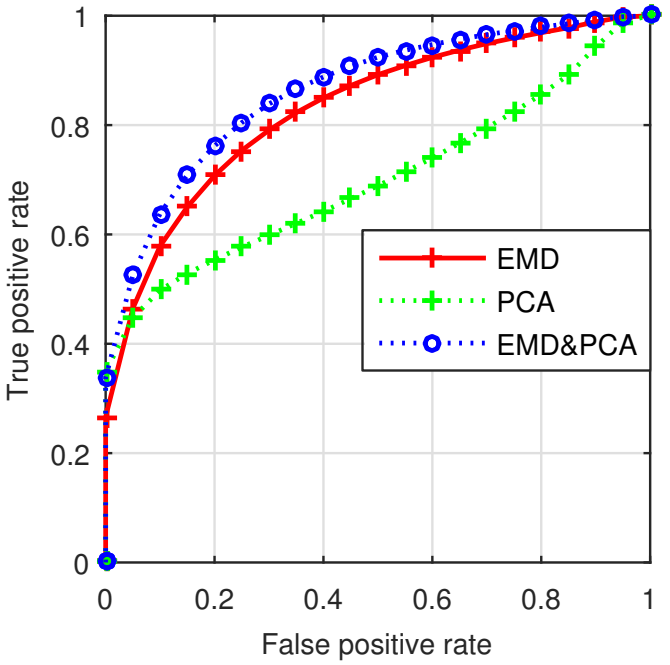


Fig. 8. Average receiver operating characteristic (ROC) curves for different features (EMD, PCA, and combined EMD and PCA) for the tumor-response-injected clinical data with $\Gamma = 1$.

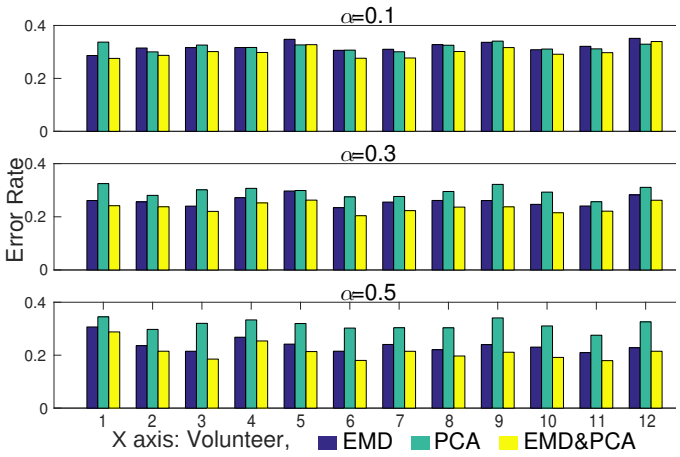


Fig. 9. The test set (generalized) detection error for each volunteer, averaged over 50 data sets, for three choices of α.

A slight improvement over EMD features can be achieved by using the combined EMD and PCA features.

The average detection error is similar for all volunteers, irrespective of the features employed, as shown in Figure 9. We have also observed (results not shown) that the detection algorithms have consistent performance with respect to tumor locations, which are randomly generated in the upper outer quadrant of the breast hemispherical model.

Figure 10 shows that the reduced SNR ratio leads to significantly poorer detection performance, but the performance degradation is smooth. When $\Gamma = 0.5$, the detection performance is only slightly better than a random guess. Although experiments with phantoms suggest that a choice of $\Gamma = 1$ is reasonable, the appropriate value is dictated by

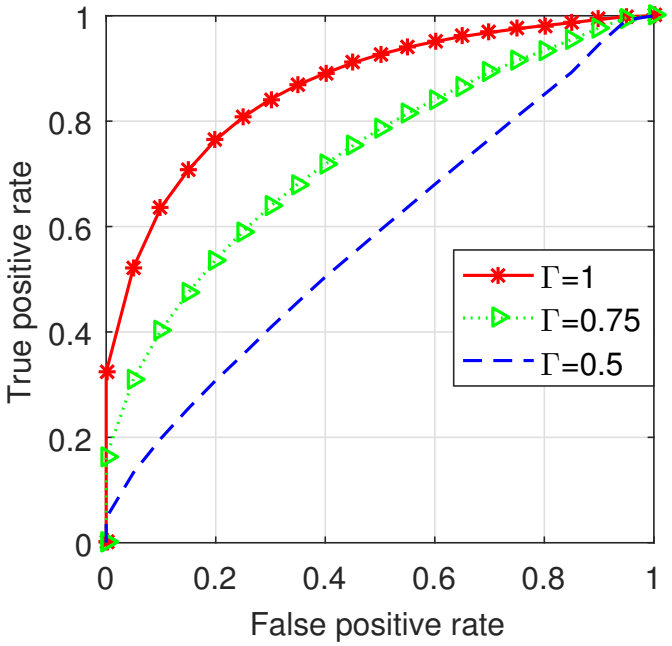


Fig. 10. The ROC curves, averaged over 20 datasets, for the case of combined EMD and PCA features for three different values of Γ.

the contrast between tumorous and healthy tissue, and there is still considerable uncertainty about the extent of this contrast. The performance deterioration motivates the development of feature extraction methods and detection algorithms that can better tackle these lower SNR scenarios.

We note that the detection rate still needs to be improved for clinical use. Future work will focus on further improving the detection performance and conducting more experiments as soon as more clinical data become available.

V. Conclusion

In this paper, we proposed and explored the use of an EMD-based feature extraction method for microwave breast tumor detection. We were motivated to explore the use of EMD features because of their potential robustness to the system jitter that is common in microwave breast cancer scans. We evaluated the detection performance using EMD-based features and the commonly used PCA features using clinical trial data collected over an eight-month period combined with a numerical tumor response construction method. We observed that the use of features based on EMD leads to significantly improved detection performance compared to PCA-based features. A marginal additional improvement can be achieved by using a combination of EMD and PCA features.

Acknowledgment

The authors gratefully acknowledge the McGill Microwave Breast Cancer Detection group, led by Prof. Milica Popović and Prof. Mark Coates, for providing experimental data and advice. The authors also appreciate the selfless help from Yunpeng Li.

REFERENCES

[1] A. C. Society, *Cancer News*. American Cancer Society, 2017, vol. 1.

[2] S. H. Heywang-Köbrunner, A. Hacker, and S. Sedlacek, "Advantages and disadvantages of mammography screening," *Breast care*, vol. 6, no. 3, pp. 199–207, 2011.

[3] E. C. Fear, S. C. Hagness, P. M. Meaney, M. Okoniewski, and M. A. Stuchly, "Enhancing breast tumor detection with near-field imaging," *IEEE Microwave*, vol. 3, no. 1, pp. 48–56, 2002.

[4] S. Hagness, K. Leininger, J. Booske, and M. Okoniewski, "Dielectric characterization of human breast tissue at microwave frequencies," Orlando,FL, 2000.

[5] A. J. Surowiec, S. S. Stuchly, J. R. Barr, and A. Swarup, "Dielectric properties of breast carcinoma and the surrounding tissues," *IEEE Trans. Biomed. Eng.*, vol. 35, no. 4, pp. 257–263, 1988.

[6] T. Sugitani, S.-i. Kubota, S.-i. Kuroki, K. Sogo, K. Arihiro, M. Okada, T. Kadoya, M. Hide, M. Oda, and T. Kikkawa, "Complex permittivities of breast tumor tissues obtained from cancer surgeries," *Appl. Physics Lett.*, vol. 104, no. 25, p. 253702, 2014.

[7] H. Song, H. Kono, Y. Seo, A. Azhari, J. Somei, E. Suematsu, Y. Watarai, T. Ota, H. Watanabe, Y. Hiramatsu *et al.*, "A radar-based breast cancer detection system using cmos integrated circuits," *IEEE Access*, vol. 3, pp. 2111–2121, 2015.

[8] S. Caorsi, G. L. Gragnani, and M. Pastorino, "Reconstruction of dielectric permittivity distributions in arbitrary 2-d inhomogeneous biological bodies by a multiview microwave numerical method," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 232–239, 1993.

[9] P. M. Meaney, M. W. Fanning, T. Raynolds, C. J. Fox, Q. Fang, C. A. Kogel, S. P. Poplack, and K. D. Paulsen, "Initial clinical experience with microwave breast imaging in women with normal mammography," *Academic radiology*, vol. 14, no. 2, pp. 207–218, 2007.

[10] S. A. AlShehri, S. Khatun, A. B. Jantan, R. S. A. Raja Abdullah, R. Mahmud, and Z. Awang, "Experimental breast tumor detection using NN-based UWB imaging," *Progress In Electromagnetics Research (PIER)*, vol. 111, pp. 447–465, 2011.

[11] R. C. Conceicao, H. Medeiros, M. O'Halloran, D. Rodriguez-Herrera, D. Flores-Tapia, and S. Pistorius, "Initial classification of breast tumour phantoms using a UWB radar prototype," in *Int. Conf. Electromagnetics in Advanced Appl. (ICEAA)*, Singapore, 2013, pp. 720–723.

[12] 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 Gen. Assem. Sci. Symposium (GASS)*, Beijing, China, Aug. 2014, pp. 1–4.

[13] P. M. Meaney, M. W. Fanning, D. Li, S. P. Poplack, and K. D. Paulsen, "A clinical prototype for active microwave imaging of the breast," *IEEE Trans. Microw. Theory Tech.*, vol. 48, no. 11, pp. 1841–1853, 2000.

[14] A. E. Souvorov, A. E. Bulyshev, S. Y. Semenov, R. H. Svenson, and G. P. Tatsis, "Two-dimensional computer analysis of a microwave flat antenna array for breast cancer tomography," *IEEE Trans. Microw. Theory Tech.*, vol. 48, no. 8, pp. 1413–1415, Aug. 2000.

[15] D. Byrne and I. J. Craddock, "Time-domain wideband adaptive beamforming for radar breast imaging," *IEEE Trans. Antennas Propag.*, vol. 63, no. 4, pp. 1725–1735, jan. 2015.

[16] 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, no. 4, pp. 830–840, Feb. 2010.

[17] J. Bourqui, J. M. Sill, and E. C. Fear, "A prototype system for measuring microwave frequency reflections from the breast," *J. Biomedical Imaging*, vol. 2012, p. 9, 2012.

[18] J. D. Shea, P. Kosmas, S. C. Hagness, and B. D. Van Veen, "Three-dimensional microwave imaging of realistic numerical breast phantoms via a multiple-frequency inverse scattering technique," *Medical physics*, vol. 37, no. 8, pp. 4210–4226, 2010.

[19] S. K. Davis, H. Tandradinata, S. C. Hagness, and B. D. Van Veen, "Ultrawideband microwave breast cancer detection: a detection-theoretic approach using the generalized likelihood ratio test," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 7, pp. 1237–1250, Jun. 2005.

[20] P. Kosmas and C. M. Rappaport, "FDTD-based time reversal for microwave breast cancer detection-localization in three dimensions," *IEEE Trans. Microw. Theory*, vol. 54, no. 4, pp. 1921–1927, Jun. 2006.

[21] R. Conceicao, H. Medeiros, M. O'Halloran, D. Rodriguez-Herrera, D. Flores-Tapia, and S. Pistorius, "Initial classification of breast tumour phantoms using a uwb radar prototype," in *Int. Conf. Electromagnetics in Advanced Applications (ICEAA)*. Torino,Italy: IEEE, 2013, pp. 720–723.

[22] 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 *General Assembly and Scientific Symposium (URSI GASS)*. IEEE, 2014, pp. 1–4.

[23] D. Byrne, M. O'Halloran, M. Glavin, and E. Jones, "Breast cancer detection based on differential ultrawideband microwave radar," *Progress In Electromagnetics Research M*, vol. 20, pp. 231–242, 2011.

[24] S. K. Davis, B. D. Van Veen, S. C. Hagness, and F. Kelcz, "Breast tumor characterization based on ultrawideband microwave backscatter," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 237–246, Dec. 2008.

[25] L. Bruzzone and C. Persello, "A novel approach to the selection of robust and invariant features for classification of hyperspectral images," in *IEEE Int. Geosci.& Remote Sensing Symp.*, vol. 1. IEEE, Jul. 2008, pp. I–66.

[26] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A: Math., Phys. and Eng. Sci.*, vol. 454, pp. 903–995, 1998.

[27] V. Bajaj and R. B. Pachori, "Classification of seizure and nonseizure EEG signals using empirical mode decomposition," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1135–1142, Dec. 2012.

[28] S. Li, W. Zhou, Q. Yuan, S. Geng, and D. Cai, "Feature extraction and recognition of ictal EEG using EMD and SVM," *Comput. in biology and medicine*, vol. 43, no. 7, pp. 807–816, 2013.

[29] Y. Li, E. Porter, A. Santorelli, M. Popović, and M. Coates, "Microwave breast cancer detection via cost-sensitive ensemble classifiers: Phantom and patient investigation," *Biomedical Signal Process. and Control*, vol. 31, pp. 366–376, 2017.

[30] H.-G. Chew, R. E. Bogner, and C.-C. Lim, "Dual ν-support vector machine with error rate and training size biasing," in *Int. Conf. Acoust., Speech, Singal Process.(ICASSP)*, Salt Lake City, USA, 2001, pp. 1269–1272.

[31] C. Scott, "Performance measures for Neyman-Pearson classification," *IEEE Trans. Inf. Theory*, vol. 53, pp. 2852–2863, Aug. 2007.

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

[33] E. Porter, M. Coates, and M. Popović, "An early clinical study of time-domain microwave radar for breast health monitoring," *IEEE T Biomed. Eng.*, vol. 63, no. 3, pp. 530–539, Aug. 2016.

[34] P. j. W. Debye, *Polar molecules*. New York: The Chemical Catalog CO., 1929.

[35] V. Y. Sohn, Z. M. Arthurs, J. A. Sebesta, and T. A. Brown, "Primary tumor location impacts breast cancer survival," *The American Journal of Surgery*, vol. 195, no. 5, pp. 641–644, 2008.