Ultrasonic Estimation of Effective Scatterer Diameter for Micro-Macro Parameter Based Improved Breast Lesion Classification

Navid Ibtehaj Nizam, Sharmin R. Ara, and Md. Kamrul Hasan

Abstract—Effective scatterer diameter (ESD) is an important quantitative ultrasound (QUS) based micro-parameter for ultrasonic tissue characterization. However, classifying between benign and malignant breast lesions from a broad data set, using ESD, still remains a challenge. In this work, we propose a new technique for the estimation of effective scatterer diameter (ESD) of breast tissues from the diffuse component of backscattered radio-frequency (RF) data. This allows us to combine ESD with mean scatter spacing (MSS) and other ultrasonic macro-parameters for binary classification of breast lesions. In order to separate the diffuse component from the coherent component of the backscattered data, ensemble empirical mode decomposition (EEMD) is performed which decomposes the backscattered data into their intrinsic mode functions (IMFs). A non-parametric Kolmogorov-Smirnov (K-S) test is used to automatically select the IMFs responsible for diffuse scattering. To ensure proper minimization of the system effects, a multi-step process is adopted where the RF data is deconvolved and filtered as the first two steps prior to normalization using a reference tissue-mimicking phantom (TMP). As the ESD is supposed to have a fairly consistent value over a small tissue subregion, it is estimated, for an interrogated block, from the slope of the average regression line, computed from the exponential weighted average of lines, fitted over the normalized log-power spectra of the neighboring blocks, both in the axial and lateral directions. Using the proposed method, the average ESD for malignant lesions is estimated to be 123.05±8.85 µm and that for fibroadenomas is estimated to be 98.88±9.48 µm. Average ESD values estimated for inflammatory breast tissues (∼76 µm) are close to normal tissues (∼74 µm). When the estimated ESD values are used to classify between 159 benign and malignant lesions, we obtain sensitivity, specificity, and accuracy values of 91.07%, 96.11%, and 94.34%, respectively. ESD is also combined with the mean scatterer spacing (MSS) and 27 previously reported macro-parameters, estimated from the ultrasound B-mode images and ultrasound elastography, for binary classification. This results in very high sensitivity, specificity, and accuracy values of 96.43%, 98.06%, and 97.48%, respectively. The proposed classification based on this hybrid feature set of micro- and macro-parameters, therefore, show great potential to be used as a computer-aided diagnosis (CAD) tool for breast lesion classification.

Index Terms—Quantitative ultrasound, effective scatterer diameter, breast tissue characterization, mean scatterer spacing, macro-parameters.

I. INTRODUCTION

Quantitative ultrasound (QUS) has been employed for the parametrization of tissue micro-structures with a view to developing a diagnostic modality which is less subjective to operator settings and interpreter variability compared to conventional ultrasound imaging [1]. The inherent idea behind the use of QUS is that the backscattered radio frequency (RF) signals from tissue micro-structures and the interfaces of tissue regions with different acoustic properties reveal more information about the interrogated tissues than conventional B-mode imaging [2], [3]. The heterogeneity in tissue structure manifests itself in the frequency dependent scattering of ultrasound waves [4], [5]. QUS methods exploit the frequency dependence of backscattered RF signals for the estimation of attenuation of sound waves in tissues and model the backscattered data to estimate scatterer micro-parameters such as the size, shape, spacing and number density in tissues.

The potential of effective scatterer diameter (ESD) as a micro-parameter for characterizing pathological tissues have been reported extensively in the literature. The ESD of ocular tumors [6], liver [7], [8], renal tissues [10], [11], glomerular tissues [12], kidney [13], prostate [14], human aortae [15], and the uveal melanomas [16] have been reported. Additionally, ESD has been successfully used to distinguish between benign and malignant lesions in the eyes [17] and in the lymph nodes [18], [19]. Ultrasonic characterization of human breast tissues have been reported in [20] and in [21]. However, in the method discussed by Golub et al. [20], no specific micro-parameters have been used while Tadayyon et al. [21] used micro-parameters for tumor grading rather than breast lesion classification. Our group previously used mean scatterer spacing (MSS) for classification of breast lesions on a small data set [22]. But, to the best of our knowledge, no previous work has been reported on a broad dataset for classifying between benign and malignant breast lesions using ultrasonic micro-parameters like ESD and MSS, either individually or combined. In addition, macro-parameters derived from the ultrasound B-mode image and ultrasound elastography has been successfully employed for classification of breast lesions [23], [24], [25]. However, no previously reported technique has attempted to combine the micro- and macro-parameters for ultrasonic tissue characterization.

Previously, it has been established that the backscattered RF data consists of a coherent component and a diffuse component [26]. In [27], it has been reported that, for the estimation of ESD, the coherent component behaves as interference and hence, needs to be suppressed. The generalized spectrum (GS) and Rayleigh envelope statistics have been used to suppress the coherent component in [28]. Hanning tapers

Q...
have also been used to separate the diffuse echoes from the backscattered RF data [29]. These proposed algorithms [28], [29], however, remain untested on human tissue. Recently, ensemble empirical mode decomposition (EEMD) has been successfully employed to separate the coherent and diffuse components from deconvolved backscattered RF data for MSS estimation of female breast tissues [22]. Furthermore, the existing techniques use large 2-D spatial signal blocks to generate a stable block power spectra [5], [30] which is often an unrealistic approach because the tissue pathology inside a large spatial region must be considered uniform, a requirement that is too idealistic for a heterogeneous tissue medium. To combat this problem, an attenuation estimation technique has been developed in [31] which avoids the need of large spatial signal blocks by calculating the 1-D block power spectra as a weighted average of power spectra of the neighboring blocks.

In this work, we propose a new ESD estimation technique for breast tissues based on EEMD and we combine ESD with MSS and ultrasound macro-parameters to propose a unique hybrid of micro- and macro-feature set for breast lesion classification. A multi-step system effect minimization process is employed where deconvolution of the raw RF data is carried out followed by filtering. The deconvolution technique has been previously proposed in [32]. An ideal bandpass filter of frequency range $2 - 13$ MHz is applied on the RF data to eliminate aberrations due to diffraction and other noises [33]. EEMD is used to decompose the deconvolved and filtered RF data into its intrinsic mode functions (IMFs). Since information about the diffuse component is contained in only some of the IMFs, a non-parametric Kolmogorov-Smirnov (K-S) [34] test is performed to automatically select the IMFs that show significant diffuse scattering. Finally, a reference tissue-mimicking phantom (TMP) is used to minimize the residual system effects in the EEMD domain since a reference TMP is shown to be suitable for the linear array transducers used in our system [35]. The reference phantom-based system effect minimization approach has been previously employed for system effect minimization for clinical ultrasound data [21], [36]. Additionally, we consider that the local ESD of a heterogeneous soft tissue is similar in a small sub-region which derives from the assumption that several small homogeneous sub-regions can model a heterogeneous tissue [37] and ESD is measured from the slope of a regression line fitted to the power spectra of an effectively small 1-D signal block. The average ESD of each investigated block on a particular scan line is calculated by using its neighbors in the axial and lateral directions. The spectral normalization-based average attenuation estimation (SNAAE) method proposed in [31] is used in this paper to obtain the attenuation coefficient (AC) values which are required for attenuation compensation while estimating the log power spectra. In order to check the efficacy of ESD as a diagnostic tool, we use it for classifying between benign and malignant lesions of 159 patients. We also obtain the classification results by combining ESD with MSS. The MSS is obtained by a little modification of our recently proposed algorithm [23]. Furthermore, ESD is combined with MSS and 27 previously reported QUS macro-parameters to produce a unique hybrid of ultrasound micro- and macro-parameter feature set for classifying between benign and malignant lesions.

This paper is divided into four sections. Section II explains the proposed method. Section III presents the results obtained from our proposed method and compares them with some existing techniques. Section IV discusses some of the important aspects of the proposed method and the breast lesion classification results. Finally, Section V concludes the paper with some comments.

II. MATERIALS AND METHOD

A. Patient Data

The in vivo patient data used in this paper have been obtained at the Bangladesh University of Engineering and Technology (BUET) Medical Center with the help of a SonixTOUCH Research (Ultrasonix Medical Corporation, Richmond BC, Canada) scanner integrated with a L14–5/38 linear probe. The probe was operating at 10 MHz with 65% bandwidth at a sampling rate of 40 MHz. The pulse length was approximately $0.4$ mm and the beam width was approximately $2.2$ mm. The pulse length has been calculated from the full-width at half-maximum (FWHM) of the emitted pulse, estimated using the multiple input-output inverse theorem [38].

The study has been conducted on 179 patients, with their prior written consent, and approved by the institutional review board (IRB) of BUET. From the 179 patients, 245 RF data files have been recorded. Out of the 245 data records, 56 show instances of malignant tumors, 79 show benign growths, 24 show breast abscess, 42 show cysts and 44 data records do not show any growths. The age range of all patients was 13-75 years (mean: $35.27$ years). The patients having masses underwent fine-needle aspiration cytology (FNAC) and/or excision biopsy according to the suggestion of their physicians. All patients having FNAC diagnosis positive for malignancy underwent surgery. Therefore, diagnoses of malignant and some benign lesions were confirmed by histopathology, and diagnoses of the remaining lesions, by cytopathology.

B. TMP Data

A homogenous TMP, namely A, of Computerized Imaging Reference Systems Inc. (CIRS) of dimension $3 \times 4$ cm$^2$ is used as the reference phantom. The speed of sound in the TMP is around $1530$ m/s. The TMP data was acquired at the Bangladesh University of Engineering and Technology (BUET) Medical Center, Dhaka, Bangladesh, by using a SonixTOUCH Research scanner integrated with a L14–5/38 probe operating at 10 MHz with 65% bandwidth and at a sampling rate of 40 MHz. For the estimation of average ESD we have used two homogeneous CIRS TMPs, namely A and B, which are inclusion-free, of dimensions $3 \times 4$ cm$^2$ and one heterogeneous TMP, namely C, of dimensions $4.5 \times 4$ cm$^2$, having an inclusion of diameter $1.4$ cm. The actual average ESD, as supplied by the manufacturer, i.e. CIRS, are used as gold standards for performance evaluation of the ESD estimators. The ESD description of the experimental phantom datasets are presented in Table I.
TABLE I
DESCRIPTION OF EXPERIMENTAL TMPs

| TMP Dataset | Description | Average ESD ($\mu m$) |
|-------------|-------------|----------------------|
| A           | Homogeneous | 45                   |
| B           | Homogeneous | 45                   |
| C           | Heterogenous| 45                   |

C. Preprocessing

A block diagram of our proposed ESD estimation algorithm is shown in Fig. 1. At first, 2-D regions of interest (ROI) are selected from suitable frames of the recorded RF data. In case of selecting the ROI within a breast lesion, the B-mode image is observed, and a suitable 2-D region within the border of the lesion is taken as the ROI. Additionally, a second ROI is selected outside the lesion to compare the ESD values inside and outside the lesion. It is ensured that the selected ROI is away from the edges of the imaging plane. The dimension of each ROI is approximately $(10-12) \times (6.25-9.40) \text{ mm}^2$. The size of the ROI is approximately 25 pulse lengths axially and 4 beam widths laterally. For estimating the average ESD from TMP datasets A and B, twenty-five ROIs are selected, each of dimension $1 \times 1 \text{ cm}^2$, while for TMP dataset C, ten ROIs are selected from outside the inclusion and five ROIs are selected within the inclusion. In TMP dataset C, we have also selected five heterogenous ROIs across the border of the inclusion such that the ROIs encompass both the inclusion and the background. As illustrated in Fig. 1, deconvolution and filtering of the raw RF data serve as preprocessing steps. It is to be noted that, in order to accurately estimate micro-parameters like ESD from the backscattered RF data, system effects, e.g., the effect of the system point spread function (PSF) and diffraction need to be minimized [33], [39].

In this paper, a multi-step system effect minimization scheme has been proposed. The deconvolution, and filtering of the raw RF data serve as both preprocessing and system effect minimization steps. The last step of system effect minimization is normalization using a reference TMP as shown on the block diagram of Fig. 1. The multi-step system effect minimization process is discussed in detail in later sections. The RF data is deconvolved using a blind multi-channel least-mean square algorithm previously proposed in [32]. After that, an ideal bandpass filter of frequency range $2-13 \text{ MHz}$ is applied on the deconvolved data. The rationale behind choosing a frequency range of $2-13 \text{ MHz}$ is to reduce the impact of diffraction prominent below $2 \text{ MHz}$ and eliminate other acquisition noises present at higher frequencies [33].

D. Coherent Component Suppression using EEMD

The backscattered RF data consist of two major components, a coherent component and a diffuse component [26]. The diffuse component represents incoherent scattering, i.e., scattering from aperiodic scatterers. Although the coherent component has been shown to be of diagnostic importance in the estimation of other QUS micro-parameters, such as the MSS [40], [41], accurate estimation of ESD requires the suppression of the coherent component [28]. This is because, the coherent component has been shown to increase the bias and variance of ESD estimates [28]. It is to be also noted that the backscattered RF data is produced by the convolution of the echo signal produced by the tissue scatterers with the system PSF [32]. Therefore, if the effect of the system PSF is removed through deconvolution, only the coherent and the diffuse components remain in additive form and hence, can be separated by a suitable signal decomposition technique [22].

Empirical mode decomposition (EMD) decomposes a signal into its IMFs in additive form [12]. These IMFs are known to contain information about the different frequency components present in the signal. The main advantage of EMD over other data-driven approaches is that it requires no pre-selection of basis functions and is suitable for both stationary and non-stationary signals [32]. Hence, EMD can be used to extract the diffuse component from the selected ROI from the backscattered signal after deconvolution and filtering has been carried out. However small perturbations can adversely effect EMD and a completely different set of IMFs may result each time it is performed. This would increase the variance of any micro-parameter estimated from these IMFs [43]. In order to produce stable IMF estimates, EEMD is usually performed [44], [45]. In EEMD, an ensemble of random Gaussian noise is added to the backscattered signal to produce an ensemble of signals. The EMD algorithm is applied to each signal in the ensemble to produce an ensemble of IMFs. These ensemble of IMFs are then averaged to produce a new set of IMFs. The resulting IMFs are referred to as the IMFs produced by EEMD. As stated before, EEMD allows separation of the diffuse and the coherent components from the backscattered data [22] and thereby, helps in reducing the effect of the coherent scatterers on the diffuse scatterer components.

Selection of IMFs is a crucial issue in any algorithm involving EEMD [22], [46]. This is because, a few of the IMFs will contain information about the coherent scatterers while the others will contain information about the diffuse scatterers [22]. In order to identify the IMFs exhibiting diffuse scattering, a non-parametric K-S test is performed on the IMFs estimated using EEMD. The method used is described in [34]. The K-S classifier assumes that diffuse scatterers generate Gaussian statistics and any deviation is a result of coherent scattering. Those IMFs that show deviation from Gaussian statistics (at
5% significance level) are excluded and the remaining IMFs are summed up and used for further processing.

E. Proposed ESD Estimation Technique

The gated backscattered RF signal intensity, \( W(f, z) \), at the transducer face can be expressed in the frequency domain as \[ W(v) = T(f) \cdot D(f, z) \cdot A(f, z) \cdot S(f, D_{\text{eff}}, n_z), \] where \( T(f) \) represents the combined effect of the transmit pulse and the transducer sensitivity (electro-acoustic and acousto-electric transfer functions); \( D(f, z) \) is the effect of diffraction; \( A(f, z) \) is the cumulative attenuation in the soft tissue; \( S(f, D_{\text{eff}}, n_z) \) represents the scattering properties of the tissue, including the effective scatterer size \( (D_{\text{eff}}) \), and acoustic concentration \( (n_z) \); \( z \) is the depth of the gated segment from the transducer face; and \( v \) = \( \{ f, D_{\text{eff}}, z, n_z \} \) is the set of variables on which the backscattered RF signal depends \[ (47), \] \[ (48). \]

We acquire RF signals from the tissue sample and the reference TMP. The backscattered RF signal is then deconvolved and filtered. The resulting signal, in the frequency domain, is given by \( W'(v) \). The cumulative attenuation, \( A(f, z) \), in soft tissues is a function of frequency \( f \) and depth \( z \), and can be expressed as \[ A(f, z) = e^{-4\rho(f)z} = 10^{-2\rho(f)z/10}, \] where \( \rho(f) \) denotes the AC in unit Nepers/cm. It is reported in \[ (49) \] that \( \rho(f) \) demonstrates a linear frequency dependence. Therefore, it can be written as \( \rho(f) = \beta \cdot f \), where \( \beta \) denotes the AC in Nepers/CMHz. By compensating for the effect of frequency dependent attenuation we get the compensated backscattered RF signal in the frequency domain as \[ W'_{\text{comp}}(v) = W'(v) A_c(f, z) = T(f) \cdot D(f, z) \cdot S(f, D_{\text{eff}}), \] where \( A_c(f, z) \) is the frequency dependent attenuation compensation function defined as \[ A_c(f, z) = A^{-1}(f, z); \] the AC, \( \beta \), and the ESD of the TMP are known and the sound velocity is approximately the same in both sample and the TMP. The method for estimating the AC is described in \[ (31). \]

To perform EEMD of the attenuation-compensated RF data, an ensemble of \( N_E \) random Gaussian noise, \( g_p(n) \) \( (p=1, \ldots, N_E) \), maintaining a signal-to-noise ratio of 30 dB with the attenuation-compensated signal in the time domain, \( w'_{\text{comp}}(n) \), is added to \( w'_{\text{comp}}(n) \). That is, the ensemble is given by \[ (50). \]

After that, the EMD algorithm \[ (42) \] is applied to each of the signals in the ensemble to decompose them into a sum of their IMFs, \( c_{pj}(n), j=1, \ldots, K \), where \( K \) is the number of IMFs and a residue, \( r_p(n) \), given by \[ (50). \]

Finally, the IMFs using EEMD are obtained from the ensemble average \[ (50). \]

The IMFs responsible for diffuse scattering, \( c_d(n) \), \( d=1, \ldots, M \), where \( M \) is the number of IMFs (normalized) responsible for diffuse scattering, are then identified by the K-S test \[ (34). \]

The IMFs are normalized using their amplitude. Normalization of the IMFs is done to ensure that no undue weight is given to any one of the IMFs. The attenuation-compensated RF data, for both the sample and the TMP, are then replaced by the summation of the IMFs responsible for diffuse scattering as \[ (8). \]

where, \( w'_i'_{\text{comp}}(n) \) is the IMF-replaced signal.

The power spectrum of the sample tissue at a particular depth is normalized by the temporally averaged power spectrum of the reference TMP for the same depth to minimize the system effects. Exploiting this system effect minimization process, we propose a spectral normalization-based method where an average linear regression line is fitted over the normalized half-power bandwidth of the power spectrum and some spectral parameters such as average spectral slope, and average intercept are estimated. The average regression line is computed from the weighted average of regression lines fitted over the nearest neighbourhood (NN) blocks in order to better represent the tissue structure and hence, reduce the variance of estimates.

Now, the attenuation compensated intensity spectra for the IMF replaced signals for the reference and sample tissues can be rewritten as \[ (9), \] \[ (10). \]

where \( W'_i'_{\text{comp}}(v) \) \( = T(f) \cdot D_R(f, z) \cdot S_R(f, D_{\text{eff}}, n_z, R) \) \( (9) \) \( W'_i'_{\text{comp}}(v) \) \( = T(f) \cdot D_S(f, z) \cdot S_S(f, D_{\text{eff}}, n_z, S) \) \( (10) \). where the subscripts \( R \) and \( S \) denote the reference and sample, respectively. For the same average sound speed in the reference and sample tissues, the diffraction terms can be considered as \( D_R(f, z) = D_S(f, z) \). It is to be noted that here diffraction refers to any residual aberrations due to diffraction remaining after the pre-processing steps. Finally, dividing \[ (10) \] by \[ (9) \], the normalized spectra, \( W_n(v) \), is obtained as \[ (11). \]

Taking the logarithm on both sides of \[ (11) \], we get \[ (12). \]

A model for tissue scattering, \( S(f, D_{\text{eff}}, n_z) \), in \( (1) \) in the frequency domain was developed in \[ (31) \] for clinical array systems given by

\[ S(f, D_{\text{eff}}, n_z) = (185Lq^2(D_{\text{eff}}f^6n_zf^4 \left[ 1 + 2.66(fq(D_{\text{eff}}f^2y^2)) \right] e^{-12.159f^2(D_{\text{eff}}f^2/2)}}). \]
with \( L \) the gate length (mm), \( q \) the ratio of aperture radius to distance from the region of interest, \( f \) the frequency in MHz, and \( D_{\text{eff}} \) the effective scatterer diameter in mm. The model was derived using a Gaussian form factor model and it has been previously established [5], [52], that a Gaussian form factor model is the most reliable representation of tissue structure. The quantity, \( n_z \), is termed the acoustic concentration and is the product of the number of scattering particles per unit volume, and the fractional change in the impedance between the scattering particles and the surrounding medium. For sample and reference, the tissue scattering is then represented as

\[
S_S(f, D_{\text{eff}, s}, n_z, s) = \frac{185Lq^2(D_{\text{eff}, s})^6n_z,Sf^4}{[1 + 2.66(fq(D_{\text{eff}, s})^2)]e^{-12.15f^2(D_{\text{eff}, s})^2}},
\]

\[
S_R(f, D_{\text{eff}, R}, n_z, R) = \frac{185Lq^2(D_{\text{eff}, R})^6n_z,Rf^4}{[1 + 2.66(fq(D_{\text{eff}, R})^2)]e^{-12.15f^2(D_{\text{eff}, R})^2}}.
\]

Dividing (14) by (15), and considering the fact that \( 2.66(fq(D_{\text{eff}, s})^2) \ll 1 \), we get the normalized tissue scattering as

\[
\frac{S_S(f, D_{\text{eff}, s}, n_z, s)}{S_R(f, D_{\text{eff}, R}, n_z, R)} = \frac{D_{\text{eff}, s}^6n_z,S}{D_{\text{eff}, R}^6n_z,R}e^{-3.09375(D_{\text{eff}, s}^2 - D_{\text{eff}, R}^2)f^2}.
\]

It is established that normalization using a reference phantom removes the effect of the system PSF as well as aberrations due to diffractions [21]. However, normalization by a reference TMP may not be sufficient for complete removal of system effects. This has been illustrated in the results and discussion section by showing that the ESD estimation accuracy and classification performance can be further improved if a multi-stage system effect normalization technique is adopted. Therefore, normalization using a reference TMP serves as one of the three steps of our system effect normalization process, along with deconvolution and filtering discussed in the preprocessing section. It is to be noted that for TMP datasets B and C, TMP dataset A has been used as reference while for TMP dataset A, TMP dataset B has been used as reference. Next, taking logarithm on both sides of (16) yields

\[
10\log\frac{S_S(f, D_{\text{eff}, s}, n_z, s)}{S_R(f, D_{\text{eff}, R}, n_z, R)} = 10\log\frac{D_{\text{eff}, s}^6n_z,S}{D_{\text{eff}, R}^6n_z,R} - 13.20 \times (D_{\text{eff}, s}^2 - D_{\text{eff}, R}^2)f^2.
\]

To estimate ESD, we fit a regression line through the usable (i.e., \( 6 \, \text{dB} \)) bandwidth of the normalized log scattering power spectrum. From (12), this is equivalent to fitting a line through the usable (i.e., \( 6 \, \text{dB} \)) bandwidth of the normalized log power spectrum. Assuming \( f^2 = x \), the regression line can be expressed as

\[
y = mx + c
\]

with

\[
y = 10\log\frac{S_S(f, D_{\text{eff}, s}, n_z, s)}{S_R(f, D_{\text{eff}, R}, n_z, R)},
\]

\[
m = -13.20 \times (D_{\text{eff}, s}^2 - D_{\text{eff}, R}^2),
\]

\[
c = 10\log\frac{D_{\text{eff}, s}^6n_z,S}{D_{\text{eff}, R}^6n_z,R}.
\]

Using average block power spectra generated from spatial signal blocks of sufficiently large size, ESD can be estimated from (20). On the other hand, probability that a single gated RF block includes regions of heterogeneous tissues increases with the block size. Hence to trade-off between homogeneity in the spatial blocks and consistency of the estimated power spectrum, we propose an ESD estimation technique by using weighted nearest neighbors. We assume that the ESD of the neighboring tissues as well as the acoustic concentration of the scattering particles in the neighboring blocks are almost the same for their physical proximity.

Therefore, to improve the reliability of the estimate, we calculate an average regression line as the weighted average of the regression lines of the neighboring blocks as

\[
Y(i_s, j_s) = \frac{\sum_{i_o=i_s-L_a}^{i_o+L_a} \sum_{j_o=j_s-L_t}^{j_s+L_t} (\log w(i_o, j_o))(i_o, j_o)}{\sum_{i_o=i_s=L_a}^{i_o+L_a} \sum_{j_o=j_s=L_t}^{j_s+L_t} w(i_o, j_o)}
\]

where \( Y(i_s, j_s) \) denotes the weighted average value of \( y(i_s, j_s) \), and \( w(i_o, j_o) \) is the exponential weight function for an interrogated point \( (i_s, j_s) \) on the 2-D ESD map, defined as

\[
w(i_o, j_o) = e^{-|\lambda_0(i_o-i_s)-|\lambda_1(j_o-j_s)|},
\]

where \( \lambda_0 \) and \( \lambda_1 \) denote the weighting factors in the axial and lateral directions, respectively, and \( i_s - L_a \leq i_o \leq i_s + L_a \) and \( j_s - L_t \leq j_o \leq j_s + L_t \). \( L_a \) and \( L_t \) are the NN factors for the axial and lateral directions, respectively. From [23] it is evident that \( w(i_o, j_o) \) has the maximum value (unity) at \( (i_o, j_o) = (i_s, j_s) \). We define \( w(i_o, j_o) \) in a way such that in the averaging process, the logarithm of measured power spectrum of a neighboring window is properly weighted to have less contribution with increasing distance from the interrogated block. A 2-D weighted exponential neighborhood having \( L_a = L_t = 5 \) is illustrated in Fig. 2. The values on the weight axis are arbitrary but show an exponential decay as we move away from the interrogated window both axially and laterally. To minimize the noise, an average regression line can be estimated using the proposed weighted nearest neighbors and hence, average spectral parameters (i.e., average slope and average spectral intercept) can be estimated from this average regression line.

Substituting the value of \( y \) from (18) into (22), we get

\[
Y(i_s, j_s) = \frac{A}{B},
\]

where

\[
A = \sum_{i_o=i_s-L_a}^{i_s+L_a} \sum_{j_o=j_s-L_t}^{j_s+L_t} (m(i_o, j_o)x + c(i_o, j_o))w(i_o, j_o),
\]

\[
b = \sum_{i_o=i_s-L_a}^{i_s+L_a} \sum_{j_o=j_s-L_t}^{j_s+L_t} w(i_o, j_o),
\]

\[
c = 10\log\frac{D_{\text{eff}, s}^6n_z,S}{D_{\text{eff}, R}^6n_z,R}.
\]
and

\[
B = \sum_{i_0=s-L_a}^{i_0+s-L_a} \sum_{j_0=s-L_i}^{j_0+s-L_i} u_j(i_0, j_0). \tag{26}
\]

If we define weighted average value of the slope as

\[
M(i_s, j_s) = \frac{\sum_{i_0=s-L_a}^{i_0+s-L_a} \sum_{j_0=j_s-L_i}^{j_0=j_s-L_i} m(i_0, j_0) u_j(i_0, j_0)}{\sum_{i_0=s-L_a}^{i_0+s-L_a} \sum_{j_0=j_s-L_i}^{j_0=j_s-L_i} u_j(i_0, j_0)}, \tag{27}
\]

and weighted average value of the intercept as

\[
C(i_s, j_s) = \frac{\sum_{i_0=s-L_a}^{i_0+s-L_a} \sum_{j_0=j_s-L_i}^{j_0=j_s-L_i} c(i_0, j_0) u_j(i_0, j_0)}{\sum_{i_0=s-L_a}^{i_0+s-L_a} \sum_{j_0=j_s-L_i}^{j_0=j_s-L_i} u_j(i_0, j_0)}, \tag{28}
\]

then \(24\) can be written in the form of a regression line given by

\[
Y(i_s, j_s) = M(i_s, j_s) x + C(i_s, j_s). \tag{29}
\]

Now, from the slope, \(M\), of the regression line that fits \(29\), we can estimate the ESD (mm) at the point \((i_s, j_s)\) as

\[
D_{eff,S} = \sqrt{-\frac{M(i_s, j_s)}{13.20} + D_{eff,R}^2}. \tag{30}
\]

In order to estimate the block power spectra of an interrogated block with higher resolution, the block is divided into 1-D segments with consecutive window segments in a block having 50% axial overlapping. The windowed segments are gated by the Hamming window to reduce the leakage artifacts. The length of the segments are chosen so that the stationarity assumption is satisfied. The block power spectra are calculated using the Welch method \[54\]. We select \(L_a = L_i = 5\) as the NN factors to estimate the local ESD for a particular interrogated block. The impact of varying the NN factors on the ESD estimation is discussed in the next section.

F. MSS Estimation from Breast Tissues

Another important micro-parameter which has previously been successfully used for breast tissue characterization is MSS \[21\], \[22\], \[40\]. In this paper, we have also estimated MSS to use multiple micro-parameters along with multiple macro-parameters for breast lesion classification. In \[22\], an EEMD based method for estimating MSS was proposed. The MSS is estimated from the coherent component of the backscattered data. However, the method relied on an empirical technique for the selection of IMFs and the K-S test was simply used for validation of the selected IMFs. In this work, the MSS for the 245 RF data records are estimated using the automatic IMF selection scheme proposed in this paper for ESD estimation. In addition, the backscattered data is also filtered using the ideal bandpass filter discussed earlier for minimization of the impacts of diffraction. The estimated MSS values for the different types of tissue are presented in the results section.

G. Micro and Macro Parameter Based Classification of Breast Lesions

Previous works on a broad dataset employing QUS based breast tissue characterization involved the use of parameters extracted from the power spectrum of backscattered echoes \[20\]. QUS based micro-parameters like the ESD and MSS have also been successfully employed for breast tumor grading \[21\]. A benign-malignant breast lesion classification technique based on MSS has been recently proposed in \[22\] on a limited dataset. However, to the best of our knowledge, no previous work on breast lesion classification has been carried out using ultrasound micro-parameters like the ESD and MSS on a broad dataset. Very recently, in \[67\], bi-modal (ultrasound B-mode and ultrasound elastography) ultrasound macro-parameters have been successfully used for classification of breast lesions. However, to the best of our knowledge, no breast lesion classification method has been reported in the lit-
In this work, we attempt to further improve the classification accuracy achieved in [67] by extending the feature space with the micro-parameters extracted here to produce a unique hybrid feature set of micro- and macro-parameters. Bi-modal QUS based macro-parameters have been previously reported in the literature [23], [25], [55]– [66]. The macro-parameters are mainly composed of shape, boundary, margin, contrast, and echo-pattern based features and are estimated from the ultrasound B-mode images and from ultrasound elastography. Out of the full-set of macro-parameters, 27 features that have shown promise in breast lesion classification are listed in Table II. In this paper, these features are combined with ESD and MSS for benign-malignant classification of breast lesions. The normalized feature set is passed through a wrapper-based feature reduction scheme and classified by using a K-fold cross-validation technique described in [68], with the help of support vector machine (SVM), K-nearest neighbor (KNN), linear discriminant analysis (LDA), multinomial logistic regression (MNR), and Naïve Bayes classifiers. The classification performance of the 27 macro-parameters are also measured using the same scheme.

In addition, classification performances of ESD and MSS alone, and the combination of ESD and MSS (both normalized) are also evaluated directly using the above mentioned classifiers. In these cases no feature reduction scheme is used.

### III. Results

In this section, we present the results of estimating the ESD using our proposed method. We validate the accuracy of our proposed algorithm by estimating the ESD from TMPs whose actual ESD is known. We also present the results of ESD estimated using some of the conventional techniques that employ a Gaussian form factor model [5], [53]. Furthermore, we present the classification results obtained using our ESD estimates and compare that classification performance with the classification results obtained using ESD estimated from the other techniques [5], [53]. Moreover, we present the classification results obtained using MSS as well as the combination of ESD and MSS, the 27 macro-parameters derived from ultrasound bi-modal images (B-mode and elastography), and finally, the proposed hybrid feature set of ESD and MSS with the 27 macro-parameters.

#### TABLE III

**ESTIMATED ESD VALUES (IN µm) WITH SD (IN BRACKET) FROM EXPERIMENTAL TMPs**

| Method                     | TMP Dataset | Average ESD of Background (µm) (±SD) | MAPE (%) | Average ESD inside Inclusion (µm) (±SD) | MAPE (%) |
|----------------------------|-------------|--------------------------------------|----------|----------------------------------------|----------|
| MASD Based Method [5]      | A           | 34.82 (±8.88)                        | 21.82    | -                                      | -        |
|                            | B           | 52.18 (±8.09)                        | 15.96    | -                                      | -        |
|                            | C           | 50.91 (±7.99)                        | 13.13    | 74.91 (±7.87)                          | 7.01     |
| Frequency Domain Method [53]| A           | 53.84 (±7.11)                        | 19.64    | -                                      | -        |
|                            | B           | 50.10 (±7.07)                        | 11.33    | -                                      | -        |
|                            | C           | 48.97 (±7.39)                        | 8.82     | 73.61 (±7.09)                          | 5.15     |
| Proposed Method            | A           | 47.02 (±5.89)                        | 4.49     | -                                      | -        |
|                            | B           | 47.76 (±6.01)                        | 6.13     | -                                      | -        |
|                            | C           | 47.78 (±5.85)                        | 6.18     | 73.02 (±6.33)                          | 6.71     |

#### TABLE IV

**ESTIMATED VALUES OF ESD WITH SD (IN BRACKET) USING DIFFERENT METHODS FOR DIFFERENT TISSUE TYPES**

| ESD (±SD) (µm)          | Malignant | Fibroadenoma | Abscess | Normal |
|-------------------------|-----------|--------------|---------|--------|
| Method                  | in        | out          | in      | out    |
| MASD based method [5]   | 100.67 (±21.13) | 86.12 (±8.41)     | 94.09 (±18.17) | 82.10 (±7.24) | 88.31 (±11.34) | 90.35 (±10.01) | 76.13 (±7.01) |
| Frequency domain method [53] | 109.21 (±17.34) | 80.04 (±8.21)    | 101.41 (±13.12) | 79.31 (±8.54) | 81.24 (±9.81) | 82.45 (±11.01) | 75.88 (±6.74) |
| Proposed Method         | 123.05 (±8.85) | 74.90 (±4.19)    | 98.71 (±9.55)  | 74.89 (±4.11) | 75.72 (±4.09) | 75.77 (±4.07) | 75.12 (±4.01) |

#### TABLE V

**ESTIMATED VALUES OF MSS WITH SD (IN BRACKET) FOR DIFFERENT TISSUE TYPES**

| Tissue Type  | MSS (± SD) (mm) |
|--------------|-----------------|
| Abscess      | 0.729 (± 0.040) |
| Fibroadenoma | 0.750 (± 0.035) |
| Malignant    | 0.793 (± 0.040) |
| Normal       | 0.689 (± 0.032) |
A. Experimental Phantom Results

To check the accuracy of our proposed average ESD estimation technique, we use three CIRS experimental TMPs for which the average ESD values in the inclusion and background are available from the manufacturer as gold standard. The average ESD of these phantoms are also estimated using the mean average square deviation (MASD) minimization based method [5], and the frequency domain method [53] in order to compare the ESD estimates with techniques that also employ a Gaussian form factor model. Table III presents the actual average ESD and the average ESD estimated using our proposed algorithm as well as the other techniques for the experimental phantoms. The average ESD represents the average of the ESD estimated from each of the ROIs within a particular TMP. It is evident that our proposed algorithm estimates the ESD for all three TMP data with a relatively high degree of accuracy compared to the other methods as reflected by a lower mean absolute percentage error (MAPE) compared to the other methods [5], [53]. Moreover, our method also shows a lower SD of estimates compared to these methods.

We have also estimated the ESD from the heterogenous ROIs of TMP C using our proposed method. A scatter plot showing how the ESD varies as we move laterally from a region just outside the inclusion across the border of the inclusion to a region within the inclusion is shown in Fig. 3. It is evident from the figure that our proposed method is able to reliably estimate the ESD in the homogenous regions inside and outside the inclusion with a mean value of approximately 48 µm outside the inclusion (represented by a red line) and a mean value of approximately 74 µm within the inclusion (represented by a green line). There is a sharp change in the average ESD values across the border of the inclusion, as expected.

Now, in order to study how the different steps in the proposed algorithm impacts the accuracy of the ESD estimates, the ESD was estimated for several ROIs of TMP datasets A, B, and C by removing the different steps shown in the flow chart of Fig. 1 one by one while retaining the others. The results are presented in Fig. 4 in the form of a bar plot. It can be seen that the overall estimation accuracy is most adversely impacted on removing the EEMD step. There is also a slight rise in the SD of estimates on removing the EEMD step. Removal of the any one of the system effect minimization steps, i.e., deconvolution, filtering, and normalization using a reference TMP, also noticeably impacts the ESD estimation accuracy, with a slight increase in the SD of ESD estimates in each case. The weighted neighborhood step seems to have the least impact on the overall accuracy. However, the SD of the estimates is seen to rise more significantly from the proposed method. The first column of the bar plot presents the performance of the proposed algorithm on the same datasets for comparison. It can be inferred from the plot that the EEMD step has the most impact on the accuracy of the ESD estimation. Moreover, it also justifies the use of a multi-step system effect minimization technique since removal of any one of these steps has detrimental effect on the overall ESD estimation accuracy.

B. ESD Estimation Results on in vivo Breast Tissues

The results of estimating the ESD of in vivo breast tissues using various techniques are presented in Table IV. It is evident that, using our proposed method, the estimated ESD values for fibroadenoma and malignant breast tissues fall within the range of ESD values previously reported in the literature [21]. According to the best of our knowledge, the ESD values of breast abscess and cyst have not been previously reported. The ESD values in the region outside the lesions are consistent in the range between 70 – 80 µm. The results obtained from normal breasts tissues agree with the results obtained from regions outside the lesions of pathology. The ESD estimates for abscess or inflammatory tissues show little or no deviation from this range. The estimated ESD values of cyst are rather erratic and show a high standard deviation (SD) and hence, are not presented. This is true for all the techniques used to obtain the ESD estimates in Table IV. Therefore, it can be concluded that ESD estimates of cyst are of no diagnostic importance. This is in good concordance with the anatomy of the cysts as they are largely fluid-filled sacs [69] and thus, are likely to produce an inconsistent scattering signature. Furthermore, it is seen that the average ESD value of malignant lesions is

![Fig. 3. Scatter plot showing the variation of ESD across different windows on moving laterally from the edge of the ROI in the TMP background to the edge of the ROI within the inclusion.](image)

![Fig. 4. Bar plot showing the impact on the percentage error and SD of the ESD estimates from the experimental TMPs on removing the different steps from our proposed algorithm.](image)
greater than that of benign tumors which also conform with the previously reported results [21]. The estimated average ESD values using the MASD-based method [5] is presented in the first row of Table IV. We see that the SD of ESD estimates is significantly higher than that of our proposed method and the separation between the average ESD values for malignant lesions and fibroadenomas is also smaller.

Next, the results presented in the second row of Table IV obtained from the original frequency domain method proposed in [53] show a larger SD as well as a smaller separation between the average ESD values of fibroadenoma and malignant tumors than that of our proposed method.

To show the correlation between the ESD estimates and the histology images, the representative hematoxylin and eosin (H & E) histology sections of fibroadenomas, malignant tissues and abscess are presented in Figs. 5(a)–(c), respectively. In the histology images, the microlobules are purple stained while the stroma are pink stained and the fat is white. It is evident that the stromal content decreases in the malignant lesion section and the size of the micro-lobules are larger. The sizes of the lobules in the malignant tumor section also vary more widely than in the fibroadenoma section. The section for abscess is typical of what is expected for normal tissues. On measuring the average sizes of the glandular structures in the fibroadenoma and malignant sections, with the help of a microscope, they are found to be 95±17 µm and 149±24 µm, respectively. In comparison, using our proposed method, the average ESDs corresponding to fibroadenomas and malignant tissues are found to be 98.71±9.55 µm and 123.05±8.85 µm, respectively. Hence, the values obtained from the sole histology image of each type are in fairly good concordance with the values estimated from our proposed method.

The MSS values estimated from the RF data are presented in Table V. The MSS values estimated for fibroadenomas, malignant tumors and normal breast tissues fall within the range of values as reported previously [21], [40]. As in the case of ESD, the MSS values of cysts and abscesses have not been reported before. The estimated MSS values for abscesses are slightly higher than those for normal tissues while the MSS values for cysts are inconsistent, as expected, and the algorithm often fails to produce any reasonable estimates similar to that of ESD. Therefore, the average MSS value of cysts are not presented in the table.

An important factor that has to be taken into consideration while estimating ESD from in vivo tissues is the impact of the kernel size of the weighted exponential neighborhood. The algorithm involves the use of a 5 × 5 weighted exponential neighborhood. The use of such a neighborhood allows the modeling of tissue homogeneity over a small region rather than a large spatial block where the tissue becomes more heterogenous. In order to investigate how the ESD varies over the neighborhood, it is estimated for each of the blocks in a representative 5 × 5 neighborhood of a normal tissue. The results are represented in the form of a box plot in Fig. 6. It is evident from the figure that the deviation of the values is small and there are no outliers. Hence, the values of ESD in the neighborhood is fairly consistent and this trend is evident in almost all of the neighborhoods for a particular type of tissue. Therefore, the use of an exponential weighted averaging for a neighborhood is justified. To further investigate the impact of the size of the neighborhood on the ESD estimates, the ESD is estimated using our proposed algorithm for normal tissues for no neighborhood, a neighborhood of size 3 × 3, and a neighborhood of size 8 × 8. It has already been stated that the original results are produced for a neighborhood of size 5 × 5.

The results are shown in Table VI. It is clear from the table that the choice of neighborhood size does not greatly effect the average value of the ESD estimates. However, the SD of estimates seem to decrease significantly for a neighborhood size close to our selected one. A large neighborhood size or no neighborhood increases the SD of estimates and this could be attributed to the increased heterogeneity for a large spatial block of tissue. This observation is consistent to that obtained for experimental TMPs where removal of the weighted neighborhood step adversely impacted the SD of estimates.

The step-wise ablation technique employed for the ESD estimation from experimental TMPs cannot be adopted for the ESD estimation from in vivo tissues since the gold standard (i.e., histopathology) for all the data records are not available. However, it is to be noted that ablation has been applied for the classification results based on ESD estimates from in vivo tissues in the next subsection.

C. Classification Results

The classification results of breast lesions produced using the techniques discussed previously are presented in Table VII and Table VIII. The results include true positive (TP), true negative (TN), false positive (FP), false negative (FN), sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), the sum of the last five parameters, Sum5, and Matthew’s correlation coefficient (MCC) as quality metrics. TP refers to a diagnostic modality correctly identifying a malignant tumor, TN refers to a diagnostic modality correctly identifying a benign tumor, FP refers to a diagnostic modality incorrectly characterizing a benign tumor as a malignant tumor, while FN refers to a diagnostic modality incorrectly characterizing a malignant tumor as a benign tumor. From TP, TN, FP, and FN, sensitivity, specificity, accuracy, PPV, NPV and MCC can be computed as [70], [71], [72].

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (31)
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (32)
\]

\[
\text{Accuracy} = \frac{TP + TN}{TN + FN + TP + FP} \quad (33)
\]
### TABLE VII
BREAST TUMOR CLASSIFICATION RESULTS USING ESD FOR DIFFERENT METHODS

| Method                  | TP  | TN  | FP  | FN  | Sensitivity (%) | Specificity (%) | Accuracy (%) | PPV (%) | NPV (%) | Sum      | MCC  |
|-------------------------|-----|-----|-----|-----|-----------------|-----------------|--------------|---------|---------|----------|------|
| MASD [5]                | 20  | 93  | 10  | 36  | 33.71           | 90.32           | 71.06        | 66.66   | 72.09   | 335.83   | 0.3175|
| Frequency Domain [53]   | 29  | 85  | 18  | 27  | 51.79           | 82.52           | 71.69        | 61.70   | 75.89   | 343.60   | 0.3591|
| Proposed Method         | 51  | 99  | 4   | 5   | 91.07           | 96.12           | 94.34        | 92.73   | 95.19   | 469.44   | 0.8755|

### TABLE VIII
BREAST LESION CLASSIFICATION RESULTS OBTAINED FOR THE DIFFERENT MICRO- AND MACRO-PARAMETERS USING THE PROPOSED METHOD

| Parameters Used    | TP  | TN  | FP  | FN  | Sensitivity (%) | Specificity (%) | Accuracy (%) | PPV (%) | NPV (%) | Sum      | MCC  |
|--------------------|-----|-----|-----|-----|-----------------|-----------------|--------------|---------|---------|----------|------|
| MSS                | 42  | 72  | 31  | 14  | 75              | 69.90           | 71.70        | 57.53   | 83.72   | 357.86   | 0.4304|
| ESD and MSS        | 54  | 98  | 2   | 2   | 96.43           | 95.15           | 91.60        | 91.33   | 98.00   | 476.89   | 0.7054|
| Macro              | 53  | 94  | 9   | 3   | 94.24           | 91.26           | 92.45        | 85.48   | 96.91   | 461.75   | 0.7841|
| Hybrid (micro and macro) | 54  | 101 | 2   | 2   | 96.43           | 98.06           | 97.48        | 96.43   | 98.05   | 486.46   | 0.9449|

### TABLE IX
BREAST LESION CLASSIFICATION RESULTS OBTAINED USING ONLY ESD BY REMOVING THE DIFFERENT STEPS IN THE PROPOSED METHOD

| Condition           | TP  | TN  | FP  | FN  | Sensitivity (%) | Specificity (%) | Accuracy (%) | PPV (%) | NPV (%) | Sum      | MCC  |
|---------------------|-----|-----|-----|-----|-----------------|-----------------|--------------|---------|---------|----------|------|
| Proposed Method     | 51  | 99  | 4   | 5   | 91.07           | 96.12           | 94.34        | 92.73   | 95.19   | 469.44   | 0.8755|
| No Filtering        | 49  | 97  | 6   | 7   | 87.50           | 94.17           | 91.82        | 89.09   | 93.27   | 455.86   | 0.8202|
| No Normalization    | 51  | 54  | 9   | 5   | 91.07           | 91.26           | 91.99        | 85.00   | 94.95   | 453.48   | 0.7666|
| No Deconvolution    | 49  | 96  | 7   | 7   | 87.50           | 91.20           | 91.19        | 87.30   | 95.20   | 452.80   | 0.8070|
| No Neighborhood     | 39  | 100 | 3   | 17  | 69.64           | 97.08           | 87.42        | 92.86   | 85.47   | 432.48   | 0.7230|
| No EEMD             | 43  | 81  | 22  | 13  | 76.78           | 78.64           | 77.99        | 66.15   | 86.17   | 385.74   | 0.5385|

Fig. 5. Histology images of (a) fibroadenoma, (b) malignant and (c) abscess breast tissue. Scale bar: 100 µm in (a) and 200 µm in (b) and (c).

Higher values of sensitivity, specificity, accuracy, PPV, and NPV indicate better classification performance. For MCC, +1 indicates a perfect prediction, 0 indicates a uniform random prediction, and −1 indicates an inverse prediction [71]. Table VII presents the classification results based on only ESD for the proposed method, MASD-based method [5], and the frequency domain method [53]. For the proposed method, a classification scheme based on ESD alone produces sensitivity, specificity, and accuracy values of 91.07%, 96.12%, and 94.34%, respectively which is significantly better than the other techniques used for comparison in this paper. Additionally, the Sum and MCC values of 469.44 and 0.8755, respectively are clearly more superior compared to the other techniques. When the Fisher’s exact test [73] is used for testing statistical significance, the proposed method achieved significantly better quality metrics than the MASD-based method [5] and the frequency-domain method [53] with p value equal to 0.005, and 0.00001, respectively. Table VIII presents the classification performance of MSS, a combination of ESD
and MSS, the macro-parameters, and the hybrid micro- and macro-parameter feature set. The classification performance of MSS on this dataset is not satisfactory. However, it is to be noted that the classification has been carried out on a much larger dataset than that used in our previous work [22]. But, on combining ESD with MSS, we obtain improved sensitivity, specificity, accuracy, and MCC values of 96.43%, 95.15%, and 95.60%, and 0.9054, respectively. In addition, binary classification based on only the macro-parameters yields sensitivity, specificity, accuracy, and MCC values of 94.64%, 91.26%, 92.45%, and 0.8413, respectively. Finally, on combining the micro- and macro-parameters, we obtain sensitivity, specificity, accuracy, and MCC values of 96.43%, 98.06%, 97.48%, and 0.9449, respectively. The very high values of the classification metrics, using this hybrid feature set of micro- and macro-parameters, illustrates its potential to be used as a CAD tool.

Now, we investigate how the different steps of the proposed algorithm impact the classification accuracy when only ESD is used for classifying the breast lesions. The results are shown in Table IX. It is clear from the table that removing the EEMD step has the most drastic impact on classification performance with a sharp fall in the sensitivity, specificity, and accuracy results from the proposed method. Furthermore, removal of the exponential neighborhood step has a significantly detrimental impact on the classification performance. This can be correlated to the earlier observations from experimental TEMPs and in vivo tissues that removing the weighted neighborhood step significantly increases the SD of the ESD estimates. It is also clear that removing any one of the system effect removal steps also impacts the classification performance adversely. This again confirms the idea that system effect minimization requires a multi-step approach.

IV. DISCUSSION

The objective of this study was to propose a new technique for accurate breast lesion classification. Also, as a part of this work, we modified our previously proposed MSS estimation scheme [22] to use MSS as another micro-parameter, which has previously shown promise in breast lesion classification, along with ESD, for breast lesion classification.

An important aspect of the proposed ESD estimation technique is the use of a multi-stage preprocessing technique for system effect minimization. Although, previous works on ESD estimation have tried to remove the system effects, they have used a single-step technique. It has been illustrated using the results presented in Fig. 4 and Table IX that the ESD estimation accuracy and the classification performance are sensitive to the preprocessing steps. It is clear from the results that the residual system effects have a detrimental impact on the ESD estimation accuracy. The work done in this paper, has, therefore, suggested that a multi-stage preprocessing technique should be employed for better system effect minimization.

In addition, it is clear from the results of Table VIII that ESD outperforms MSS in terms of the ability to classify breast lesions. However, the use of ESD along with MSS is shown to improve the classification performance. As discussed before, the classification performance of the micro-parameters (ESD and MSS) have been evaluated directly using SVM, LDA, MNR, KNN, and Naïve Bayes classifiers. In Table VIII, the reported Sum5 value of 476.09 is obtained for a LDA classifier and represents the best Sum5 value obtained out of all the above mentioned classifiers. The average Sum5 value (±SD), found by averaging the Sum5 values obtained from each of the mentioned classifiers, is 473.92 (±3.47). This indicates that the classification performance obtained using the micro-parameters is fairly stable. Additionally, in the wrapper-based feature reduction scheme, with K-fold cross validation, which has been used to produce the classification results for the macro-parameters and the hybrid feature set, the results are found to be sensitive to the value of K, the nature of feature reduction (forward/backward), and the type of classifier used. The classification results for the macro-parameters in Table VIII have been produced using a 10-fold backward elimination wrapper based feature reduction scheme using an MNR classifier with an average of 23.8 features selected amongst the 27 while the results for the hybrid feature set have been produced using a 9-fold backward elimination wrapper based feature reduction scheme using a LDA classifier with an average of 13 features selected. These classification results represent the best results obtained over different folds, different classifiers, and the two schemes for wrapper based feature reduction (forward/backward). The results are seen to deteriorate on changing these conditions. Future works should, therefore, look to reduce the sensitivity of the classification performance to these hyper-parameters of the wrapper based classification technique. Furthermore, the use of Sum5 to indicate the classification performance has been previously established in [22], [25], and [67]. In this paper, we have also used MCC to evaluate the classification performance which is shown to have a high degree of correlation with the Sum5 values. That is, a higher Sum5 value generally leads to a MCC value closer to +1 with a minor exception observed in the third and fourth rows of Table IX.
A drawback of the micro-parameters used in this paper is that they produce erratic values for cysts. This erratic behaviour has already been explained in the previous section through the idea that cysts are fluid-filled sacs which do not have a consistent scattering signature. Thus, cysts have been excluded while obtaining the classification results in this paper. As cysts are classified as benign lesions, a micro-parameter based classification scheme will be, therefore, unable to characterize cysts correctly. It is to be noted, however, that the macro-parameters can be accurately estimated for cysts since the macro-parameters are morphological features derived from the ultrasound bi-modal images. In fact, macro-parameters produce a better classification performance if cysts are included. A macro-parameter based classification scheme on the same dataset (with cysts included) produced a $\text{Sum}_5$ value of 476.50. Moreover, some strain imaging techniques have been developed which can successfully characterize cysts and hence, may be used in conjunction with the micro-parameters for breast lesion classification.

V. CONCLUSION

This paper has presented a novel method for estimating the ESD from the diffuse component of the backscattered RF data using EEMD. Breast lesion classification using a novel hybrid of micro- and macro-parameters feature set has also been proposed. The proposed ESD estimation technique has produced reliable ESD estimates as exhibited by a more accurate estimation of ESD from experimental TPMS, used in this work, compared to similar existing techniques using a Gaussian form factor model. Moreover, our proposed technique, estimates ESD from in vivo female breast tissues with a lower SD compared to the other techniques discussed in this paper. These ESD estimates are found to be in good agreement with the measurements made from histology images. Furthermore, when the ESD is fused with MSS and 27 other macro-parameters estimated from the ultrasound bi-modal images (B-mode image and ultrasound elastography) for classifying between 150 malignant and benign breast lesions, very high values of sensitivity, specificity, accuracy, and MCC are obtained. Therefore, our proposed ESD estimation technique, along with classification based on our hybrid micro- and macro-parameter based feature set, show promises to be used as computer-aided diagnosis (CAD) tool.

VI. ACKNOWLEDGEMENT

This work was supported by HEQEP UGC, Bangladesh, under Grant CPSF#96/BUET/Win-2/ST(EEE)/2017. The in vivo breast data were acquired at BUET Medical Center by Dr. F. Alam, Assistant Professor, Department of Radiology and Imaging, Bangabandhu Sheikh Mujib Medical University, Dhaka-1000, Bangladesh. The computer programs for the MASD and the frequency domain method were derived from the graphical user interface (GUI) developed by the authors of those papers. We would like to thank the authors for providing us with the GUI.

REFERENCES

[1] K. Nam, J. A. Zagzebski, and T. J. Hall, “Quantitative assessment of in vivo breast masses using ultrasound attenuation and backscatter,” Ultrasound. Imaging, vol. 35, no. 2, pp. 146–161, 2013.
[2] M. O’Donnell, J. Minms, and J. Miller, “Relationship between collagen and ultrasound backscatter in myocardial tissue,” J. Acoust. Soc. Am., vol. 69, pp. 580–588, 1981.
[3] K. K. Shung and G. A. Thieme, Ultrasound scattering in biological tissues. CRC press, 1992.
[4] R. F. Wagner, “Statistics of speckle in ultrasound b-scans,” IEEE Trans. Sonics & Ultrason., vol. 30, no. 3, pp. 156–163, 1983.
[5] M. L. Oelze and W. D. O’Brien Jr., “Method of improved scatterer size estimation and application to parametric imaging using ultrasound,” J. Acoust. Soc. Am., vol. 112, no. 6, pp. 3053–3063, 2002.
[6] E. Feleppa, F. Lizzi, D. Coleman, and M. Yaremko, “Diagnostic spectrum analysis in ophthalmology: a physical perspective,” Ultrasound Med. Biol., vol. 12, no. 8, pp. 623–631, 1986.
[7] J. Zagzebski, Z. Lu, and L. Yao, “Quantitative ultrasound imaging: in vivo results in normal livers,” Ultrasound. Imaging, vol. 15, no. 4, pp. 335–351, 1993.
[8] F. L. Lizzi, M. Ostromogilsky, E. J. Feleppa, M. C. Rorke, and M. M. Yaremko, “Relationship of ultrasonic spectral parameters to features of tissue microstructure,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 34, no. 3, pp. 319–329, 1987.
[9] W. Liu, J. A. Zagzebski, M. A. Kliuwer, T. Varghese, and T. J. Hall, “Ultrasound scatterer size estimation in liver tumor differentiation,” Med. Phys., vol. 34, pp. 2597, year=2007..
[10] M. F. Insana, J. G. Wood, and T. J. Hall, “Identifying acoustic scattering signatures in normal renal parenchyma in vitro by varying arterial and ureteral pressures,” Ultrasound Med. Biol., vol. 18, no. 6, pp. 587–599, 1992.
[11] M. F. Insana, T. J. Hall, J. G. Wood, and Z. Yan, “Renal ultrasound using parametric imaging techniques to detect changes in microstructure and function,” Invest. Radio., vol. 28, no. 8, pp. 720–725, 1993.
[12] T. J. Hall, M. F. Insana, L. A. Harrison, and G. G. Cox, “Ultrasound measurement of glomerular diameters in normal adult humans,” Ultrasound Med. Biol., vol. 22, no. 8, pp. 987–997, 1996.
[13] M. F. Insana, “Modeling acoustic backscatter from kidney microstructure using an anisotropic correlation function,” J. Acoust. Soc. Am., vol. 97, no. 1, pp. 649–655, 1995.
[14] E. J. Feleppa, T. Liu, A. Kalisz, M. C. Shao, N. Fleschner, V. Reuter, and W. R. Fair, “Ultrasonic spectral-parameter imaging of the prostate,” Int. J. Imaging Syst. Technol., vol. 8, no. 1, pp. 11–25, 1997.
[15] S. L. Bridal, P. Fornè, P. Bruneval, and G. Berger, “Parametric (integrated backscatter and attenuation) images constructed using backscattered radio frequency signals (25–56 mHz) from human aorta in vitro,” Ultrasound Med. Biol., vol. 23, no. 2, pp. 215–229, 1997.
[16] E. J. Feleppa, R. H. Silverman, R. M. Lofberg, H. C. Boldt, H. O. Lloyd, M. J. Rondeau, M. G. Mehaeff, F. L. Lizzi, and D. J. Coleman, “Correlation of ultrasound parameter imaging with microvascular patterns in uveal melanomas,” Ultrasound Med. Biol., vol. 23, no. 4, pp. 573–581, 1997.
[17] T. Liu, F. L. Lizzi, R. H. Silverman, and G. J. Kutcher, “Ultrasonic tissue characterization using 2-d spectrum analysis and its application in ocular tumor diagnosis,” Med. Phys., vol. 31, no. 5, pp. 1032–1039, 2004.
[18] E. Feleppa, J. Machi, T. Noritomi, T. Tateishi, R. Oishi, E. Yanagihara, and J. Jucha, “Differentiation of metastatic from benign lymph nodes by spectrum analysis in vitro,” in Proc. IEEE, Ultrason. Symp., vol. 2, 1997, pp. 1137–1142.
[19] J. Mamou, A. Coron, M. Hata, J. Machi, E. Yanagihara, P. Laguer, and E. J. Feleppa, “Three-dimensional high-frequency characterization of cancerous lymph nodes,” Ultrasound Med. Biol., vol. 36, no. 3, pp. 361–375, 2010.
[20] R. Golub, R. Parsons, B. Sigel, E. Feleppa, J. Justin, H. A. Zaren, M. Rorke, J. Sokil-Melgar, and H. Kimitaki, “Differentiation of breast tumors by ultrasonic tissue characterization,” J. Ultrasound Med., vol. 12, no. 10, pp. 601–608, 1993.
[21] H. Tadayyon, A. Sadeghi-Naini, L. Wirtzfeld, F. C. Wright, and G. Czarnota, “Quantitative ultrasound characterization of locally advanced breast cancer by estimation of its scatterer properties,” Med. Phys., vol. 41, no. 1, 2014.
[22] R.-F. Chang, W.-J. Wu, K. M. Moon, and D.-R. Chen, “Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors,” Breast cancer research and treatment, vol. 89, no. 2, p. 179, 2005.
[68] W. K. Moon, C.-S. Huang, W.-C. Shen, E. Takada, R.-F. Chang, J. Joe, M. Nakajima, and M. Kobayashi, “Analysis of elastographic and b-mode features at sonoelastography for breast tumor classification,” Ultrasound Med. Biol., vol. 35, no. 11, pp. 1794–1802, 2009.

[69] N. Hines, P. J. Slanetz, and R. L. Eisenberg, “Cystic masses of the breast,” American Journal of Roentgenology, vol. 194, no. 2, pp. W122–W133, 2010.

[70] W. Zhu, N. Zeng, N. Wang, et al., “Sensitivity, specificity, accuracy, associated confidence interval and roc analysis with practical sas implementations,” NESUG proceedings: health care and life sciences, Baltimore, Maryland, vol. 19, pp. 1–9.

[71] B. W. Matthews, “Comparison of the predicted and observed secondary structure of t4 phage lysozyme,” Biochimica et Biophysica Acta (BBA)-Protein Structure, vol. 405, no. 2, pp. 442–451, 1975.

[72] S. Zhou, J. Shi, J. Zhu, Y. Cai, and R. Wang, “Shearlet-based texture feature extraction for classification of breast tumor in ultrasound image,” Biomedical Signal Processing and Control, vol. 8, no. 6, pp. 688–696, 2013.

[73] R. A. Fisher, “On the interpretation of χ² from contingency tables, and the calculation of p,” Journal of the Royal Statistical Society, vol. 85, no. 1, pp. 87–94, 1922.

[74] T. Varghese and J. Ophir, “Estimating tissue strain from signal decorrelation using the correlation coefficient,” Ultrasound Med. Biol., vol. 22, no. 9, pp. 1249–1254, 1996.

[75] A. Nahiyan and M. K. Hasan, “Hybrid algorithm for elastography to visualize both solid and fluid-filled lesions,” Ultrasound Med. Biol., vol. 41, no. 4, pp. 1058–1078, 2015.

[76] M. S.-E. Rabbi and M. K. Hasan, “Speckle tracking and speckle content based composite strain imaging for solid and fluid filled lesions,” Ultrasonics, vol. 74, pp. 124–139, 2017.