

Supervised Classification of the Accuracy of the Time Delay Estimation in Ultrasound Elastography

Mohamad Ghasemi Amidabadi, M. Omair Ahmad, and Hassan Rivaz

Abstract—The accuracy of Time-Delay Estimation (TDE) in ultrasound elastography is usually measured by calculating the value of Normalized Cross Correlation (NCC) at the estimated displacement. NCC value, however, could be very high at a displacement estimate with large error, a well-known problem in TDE referred to as peak-hopping. Furthermore, NCC value could suffer from jitter error, which is due to electric noise and signal decorrelation. Herein, we propose a novel method to assess the accuracy of TDE by investigating the *NCC profile* around the estimated time-delay. We extract several features from the NCC profile, and utilize Support Vector Machine (SVM) to classify peak-hopping and jitter error. The results on simulation, phantom and *in-vivo* data show the significant improvement of the proposed algorithm compared to the state of the art techniques.

Index Terms—Ultrasound Elastography, Supervised Quality Assessment, Accuracy Assessment, Ultrasound Imaging

I. INTRODUCTION

Ultrasound elastography is an emerging medical imaging modality that involves measuring tissue deformation field caused by an external or internal force [1]. Several ultrasound-based techniques have been established in past years in the literature to find the deformation field [2], [3], [4], [5], [6] also referred to as Time-Delay Estimates (TDE). TDE is calculated between two frames of ultrasound Radio-Frequency (RF) data, and is used to infer tissue mechanical properties, in particular Young's modulus. On one hand, elastography has been applied in several clinical trials in breast, liver and prostate cancer, and is rapidly finding new clinical applications [7], [8], [9], [10], [11]. On the other hand, it has evolved into several different techniques such as shear-wave elastography [12], [13] and quasi-static elastography [14], [15]. In quasi-static elastography, either the spatial gradient of the displacement field (i.e. a strain image) is used, or an inverse problem is solved to calculate the Young's modulus [16], [17], [18], [19].

TDE is challenging due to signal decorrelation between the two ultrasound frames. Failure in TDE creates artifacts in elastography, which can adversely affect diagnosis or surgical operations. The importance of assessment of the accuracy of TDE is four-fold. First, it can be used to mask out erroneous areas of the elasticity image. Second, ultrasound frame rate is very high, and therefore, several TDEs can be calculated and utilized every second. Recent work has, in fact, focused on exploiting multiple images to improve the quality of elastography [15], [20], [21]. An accuracy map can be exploited

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) under grants RGPIN-2015-04136 and RGPIN-06623. Authors are with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, H3G 1M8, Canada. Email: m_hase, omair, hrivaz@ece.concordia.ca

to perform weighted averaging of these elasticity images, assigning smaller weights to uncertain TDEs. Third, quasi-static elastography methods are user-dependent, and rely highly on the skill of the sonographer. An accurate assessment method can be used to generate accuracy maps alongside strain images, which can help train sonographers to obtain elastography images of higher quality. And fourth, elastography algorithms often use the displacement estimation of previous samples to reduce the search range and computational complexity, which can lead to propagation of errors. The proposed method can be used to prevent such propagation of errors.

TDE methods are always subject to small and large error values. Small error is widely referred to as jitter and can be quantified by studying the fundamental limits on the performance evaluation of TDE [20]. Large errors in the displacement field, also called peak-hopping, create outlier data in the displacement field. Peak-hopping happens when maximum value of the cross-correlation function is found somewhere far from the actual displacement estimate within a predefined search region. Several algorithms have been proposed to reduce the occurrence of large errors and limit their effect in both quasi-static [21], [22], [23] and shear-wave elastography [24], [25]. Although accuracy assessment of TDE is essential in both quasi-static and shear-wave elastography, this work focuses on the former.

Widespread adoption of ultrasound elastography relies on reliable quality assessment of TDE. Early works focused on prediction of a lower bound for the variance of the displacement estimate errors [26], [27], [28], [29]. These contributions derive a closed-form expression for the TDE error variance, which is parameterized in terms of the TDE algorithm and ultrasound system configuration, and obtain the minimum achievable error by any unbiased TDE algorithm [26], [27]. Although mathematically elegant, this approach does not provide the accuracy of the TDE and instead produces a lower bound value for the error variance.

To overcome the shortcomings of the lower bound error analysis, general performance measurements such as signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) of the strain image are widely used [30], [31], [32], [33], [34]. These measures are often called strain filter since they behave similar to a bandpass filters when they are measured as a function of axial strain. For example in [30], SNR is defined as the ratio of the mean m and standard deviation σ of the estimated strain over a small window as following:

$$\text{SNR} = \frac{m}{\sigma}$$

The strain filter proposes a framework that allows a limited

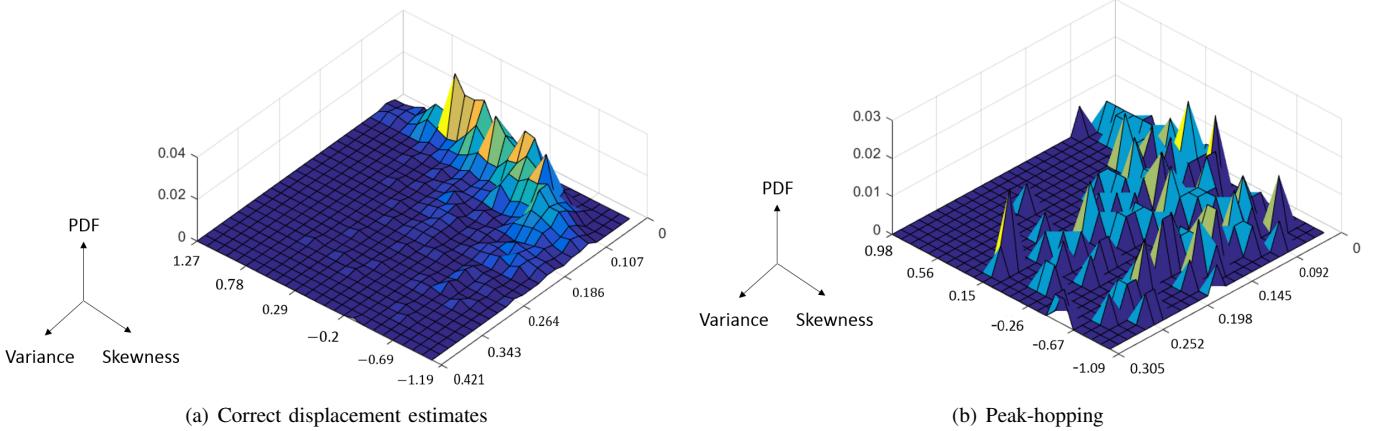


Fig. 1. Joint probability density function of the skewness and variance for 1000 correct TDEs and 200 peak-hopping TDEs .

range of strains to be included in the strain image. The reasons why the filter deviates from an all-pass characteristics in the strain domain are ultrasound system parameters, the finite value of the SNR, and the effect of signal decorrelation for high strain values. In fact, the strain filter approach provides valuable insight to design strain-imaging systems to generate high quality results. However, there are three disadvantages in using general performance measurements. First, they should be estimated in small windows that must be regions of constant strain [30]. Such regions are not necessarily available in real tissue that is largely heterogeneous. Second, these methods do not directly estimate the accuracy of TDE but rely solely on strain, the spatial derivative of TDE. As such, the derivative operator has a strong impact on these measures, such that higher SNR and CNR can be obtained by smoother derivative operators [35]. Third, strain filters only provide an upper bound for the quality of the estimated strain value, and do not produce the quality of the specific TDE. For example, these filters usually predict a high value of SNR and CNR at around 1% strain for most displacement estimation algorithms. However, an incorrect displacement may be estimated at 1% strain due to a large signal decorrelation.

The application of normalized cross-correlation (NCC) similarity metric in performance evaluation of TDE is an active field of research. NCC is generally calculated between two corresponding windows in the two RF frames to quantify the accuracy of the displacement field. In [36], a method has been proposed to mask out the incorrect areas of the strain image where NCC falls below a user-defined threshold. In [37], [38], [39], the value of NCC is linked to the precision of the displacement field, which is used later to blur out the areas of the strain image that are not accurate. In [33], standard deviation of the jitter error is evaluated by finding NCC and the sum of squared difference (SSD) values of the corresponding windows. In [40], a technique has been introduced that incorporates the consistency information of consecutive frames as a measure to evaluate the quality of strain images. Finally, a frame quality indicator has been presented in [41] that selects a few representative frames from a large pool of axial strain images based on the value of NCC

and performs weighted averaging of the strain images based on the NCC value. While these methods improve the quality of strain images, they utilize only NCC at the TDE. We will show that NCC profile around the TDE contains information that is otherwise not available from the peak NCC value.

In Fig. 1, the joint probability density functions of the skewness and variance for 1000 correct estimated samples and 200 peak-hopping samples are shown. For each sample, nine NCC values are calculated by shifting the center of the post-compressed window by ± 1 samples in the axial and lateral directions (i.e. nine values in a window of size 3×3). As it is obvious, the surface in Fig. 1(a) is concentrated in a specific area whereas it is spread randomly in Fig. 1(b). This difference in the behavior of the probability density functions enables us to use the skewness and variance of the neighboring samples around the estimated one as two invaluable features to train the classifier. The skewness of the NCC profile is calculated as following:

$$\text{skewness} = \frac{\sum_{i=1}^n (x_i - \mu)^3}{n}$$

where n and μ represent the number and average of the samples, respectively. Moreover, NCC curve by shifting ± 1 samples in the axial direction in the post-compressed image is shown in Fig. 2. It is expected to have a local maximum at the correct estimated displacement. In 2D, this curve becomes a 2D surface, and its shape can help in the assessment of the accuracy of the TDE.

Herein, we present a novel technique for assessing the accuracy of TDE that relies on multiple NCC measurements. Instead of relying on the value of NCC at the estimated displacement, we look at the NCC profile around the estimated displacement. Our technique identifies locations that contain errors larger than an accepted value, so that these regions can be marked out of the strain image. In particular, we look at the 4 NCC values obtained by varying the displacement by ± 1 sample in either axial or lateral directions in addition to the NCC value at the estimated displacement. Five aforementioned NCC values and skewness and variance of nine samples in the 3×3 neighborhood around the estimated TDE are concatenated into a feature-vector of size seven. We show that these seven

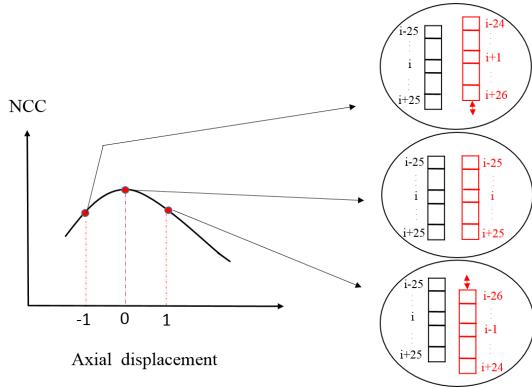


Fig. 2. NCC curve obtained by shifting the post-compressed window by ± 1 samples in both axial and lateral directions. NCC windows of pre- and post-compressed images are shown in black and red respectively.

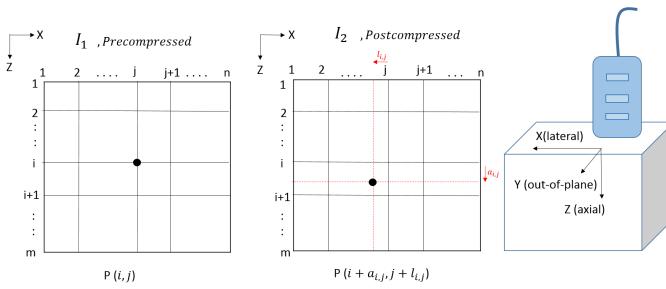


Fig. 3. Displacement between pre-and post compressed images. I_1 and I_2 are pre- and post-compressed images, respectively. Z, X, Y are axial, lateral and out-of-plane directions, respectively. The coordinate system is attached to the ultrasound probe. The sample P (i,j) in I_1 has been moved by $(a_{i,j}, l_{i,j})$ in I_2 .

features are significantly superior in discriminating the correct versus incorrect displacement estimates using simulation, phantom and *in-vivo* experiments.

II. METHODS

In elastography, pre- and post-compressed images correspond to each other via a displacement field. Some popular methods for estimation of the displacement field are discussed in detail in [14], [42]. Assume that for each sample (i,j) in the pre-compressed image, axial and lateral displacement values are a and l as illustrated in Fig. 3. Our goal is to quantify the accuracy of the displacement estimates.

NCC is a widely used similarity metric in TDE and it evaluates the degree of similarity between pre- and post-compressed images. The main advantage of NCC is its sensitivity to linear changes in the intensity of the two images compared to the ordinary cross correlation metric [43]. In addition, NCC is confined to be in the range between -1 and 1, which avoids dealing with very small or large values in training a classifier. We, therefore, use the NCC profile as our feature set.

Let I_1 and I_2 be two matrices of dimension $(m \times n)$ representing the pre- and post-compressed ultrasound images, respectively. For each sample (i,j) of the pre-compressed image, a window centered at (i,j) is considered for the calculation of

the NCC. Suppose that $\mathbf{X}=(x,y)$ is a vector containing sample coordinates such that $(x, y) \in \{i-25, \dots, i, \dots, i+25\} \times \{j\}$ (Fig. 4a). Moreover, $\mathbf{Y}_1=(x,y)$ is a shifted and linearly interpolated window in the post-compressed image containing sample coordinates such that $(x, y) \in \{i+a_{i,j}-25, \dots, i+a_{i,j}, \dots, i+a_{i,j}+25\} \times \{j-l_{i,j}\}$. The corresponding NCC value of these windows is called NCC_1 (Fig. 4c). The goal of this scheme is to investigate the behavior of the similarity metric in the neighborhood of the estimated sample. After finding eight neighboring windows of \mathbf{Y}_1 which are called $\mathbf{Y}_2, \mathbf{Y}_3, \mathbf{Y}_4, \mathbf{Y}_5, \mathbf{Y}_6, \mathbf{Y}_7, \mathbf{Y}_8, \mathbf{Y}_9$ according to Fig. 4b, we calculate the corresponding NCC between each window and \mathbf{X} , separately. These nine NCC values of each sample (i,j) in image I_1 are called $NCC_{i,i=1,2,\dots,9}$ and are shown in Fig. 4c.

A. Feature Selection

Fig. 1 indicates that the probability density function (PDF) of the peak-hopping samples is randomly distributed, whereas the PDF of the correct estimated samples is compact. Thus, the variance and skewness of the nine neighboring samples around the peak-hopping and correct estimated samples are calculated to be considered as two features for recognizing the peak-hopping error.

Assuming that a correct TDE is available for the sample X at (i,j) , NCC_1 must be larger than the other four neighboring NCC values, which are NCC_2, NCC_3, NCC_4 , and NCC_5 (NCC_1 is the similarity calculated at the correct TDE). In other words, the NCC profile has a local maximum at the correct TDE as it is shown in Fig. 2. The steepness of the maximum and convexity of the NCC profile will be used for identification of the jitter error.

B. Supervised Learning

In this work, Support Vector Machine (SVM) [44] is used as a binary classifier to find the accuracy map of the elastography algorithm. SVM is a supervised classification method, and as such, requires training data. SVM performs non-linear classification and, therefore, often outperforms linear classification techniques. In addition, while training a SVM is computationally intensive, the testing stage is very fast. This is ideal in ultrasound elastography, wherein SVM can be trained offline and be used to test the TDE results in real-time. The flow chart of our approach for obtaining the training data is shown in Fig. 5. Preparation of the training data, shown as true and false classes in this figure, is described in the following section.

C. Training Sets

We require training data that corresponds to correct and incorrect displacement estimates. For simulation data, the ground truth TDE is available from our finite element simulation. For phantom experiments and *in-vivo* data, the ground truth TDE is not available. Therefore, we use a real-time elastography algorithm [45], [46] to find TDE between pre- and post-compression RF data to obtain a silver standard. This method is based on dynamic programming and analytic

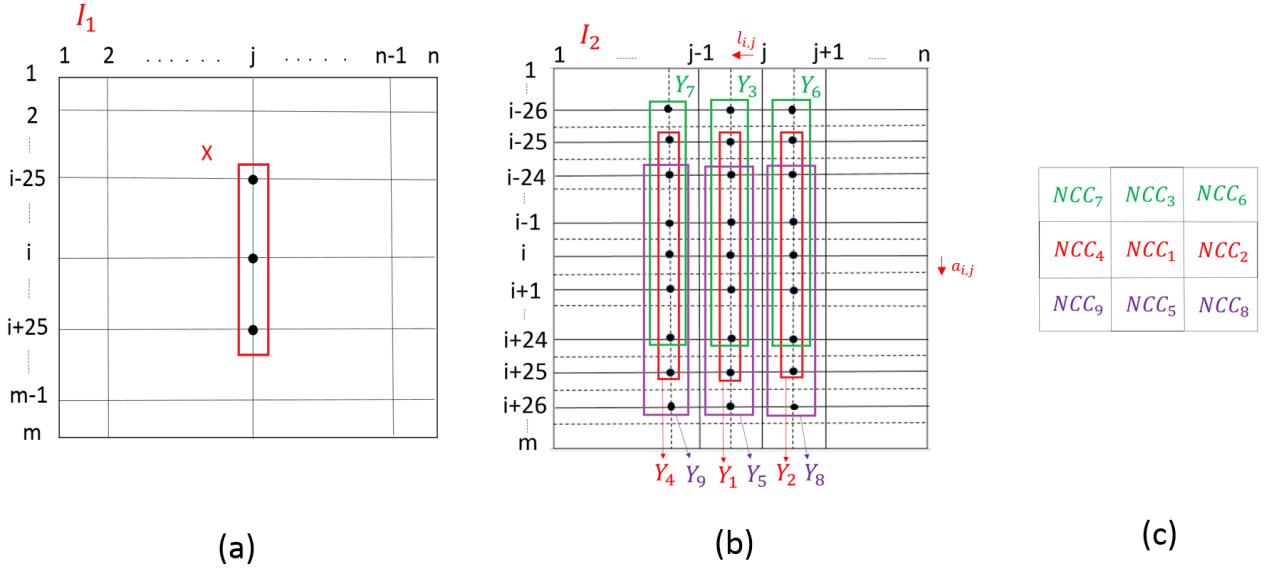


Fig. 4. (a), (b) I_1 and I_2 are pre- and post-compressed images, respectively. Windows X and Y_i , ($i = 1, 2, \dots, 9$) correspond to the sample (i, j) in I_1 . (c) Nine NCC values are illustrated, NCC_1 was the only value which was used to asses accuracy.

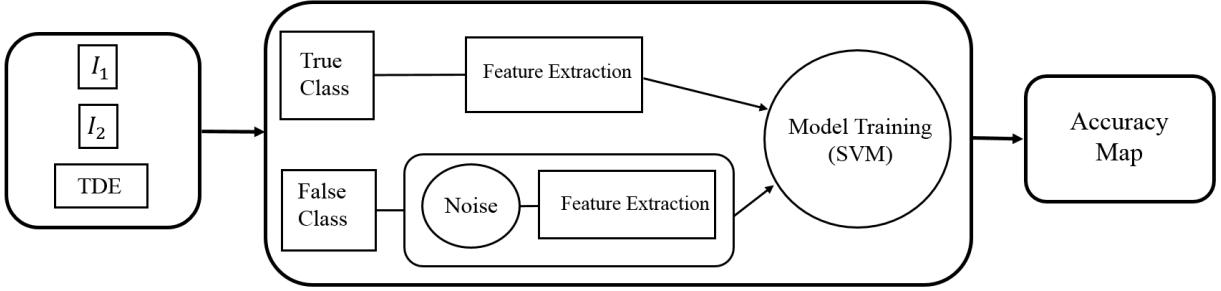


Fig. 5. Flow chart of the proposed method. I_1 and I_2 are pre- and post-compressed images, respectively. Displacement field is either available from the simulation data, or calculated from a displacement estimation algorithm.

minimization (DPAM) of a regularized cost function. We visually inspect the results to assure that the algorithm has successfully calculated the displacement field. We also vary the regularization parameter of DPAM to show that the results have small variance (Discussion Section). Therefore, for all the samples in image I_1 seven features are extracted by the correct displacement field that is available from either FEM or DPAM (Fig. 5) to form the class of correct displacements. The training set of incorrect displacements is composed of peak-hopping and jitter samples. For peak-hopping, among all the samples in I_1 , we find samples in I_2 which have larger NCC values somewhere far from the correct estimated displacement. For jitter, uniform noise is added to the rest of samples of the correct displacement field. The amplitude of this noise is between 0.4 and 0.6 samples. After creating the incorrect displacement field (which includes jitter and peak-hopping samples), we calculate seven features for all samples to form the class of incorrect displacements (Fig. 5).

D. Classification

The main idea in the proposed scheme is to employ information of the neighboring pixels to evaluate the accuracy of TDE. Therefore, instead of using one NCC value that is called NCC_1 as the only feature, four neighboring NCC values in addition to skewness and variance of nine NCC values are used according to Fig. 4(c). Therefore, for each sample (i, j) in I_1 , seven features are calculated for training and validation procedure.

III. RESULTS

The proposed scheme is implemented in MATLAB and is evaluated employing simulated, phantom and clinical data. In this study, all the samples of the pre-compressed image are considered as the training and testing data set. For each sample (i, j) in I_1 , windows of size 51 by 1 are taken in calculating the NCC values. As discussed earlier, the true class is constructed by the available ground truth or by utilizing the DPAM method. False class samples are formed by finding the

peak-hopping samples or adding a uniform random noise to the rest of samples in the axial and lateral displacements. The uniform noise is in the range of [0.4 0.6] sample in the axial and [-0.1 0.1] sample in the lateral direction. Therefore, there is a minimum of 0.4 sample error in the axial direction in the false class.

In order to find the accuracy of the classifier for each data set, 10-fold cross validation is performed for using one NCC value and seven proposed features. In 10-fold cross validation, the original data set is partitioned randomly into 10 equal sized subsets. Each time, one subset is used as the validation set to test the accuracy of the classifier, and the remaining 9 subsets are utilized as the training set. This will ensure that the algorithm is not trained and tested on the same data. In the final step, the averaged accuracy of all 10 experiments is calculated as the final accuracy of the classifier.

A. Running time

A critical and computationally expensive step in SVM classification is finding the support vectors, i.e. the training samples that are close to the decision boundary. As a result, training the proposed method on 100,000 samples takes 381.02 sec on a single core of an i7 3.4 GHz Intel CPU. This training can be performed offline. The testing step of SVM is usually very fast, since the decision boundary is determined in the training step. In our implementation, evaluating the accuracy of an image of size 100×100 takes 0.78 sec on the same CPU, which means that the method can show an accuracy map in real-time.

B. Simulation Data

Ultrasound data has been created by Field II [47], and the digital phantoms are deformed using the Finite Element Method (FEM) by ABAQUS Software (Providence, RI). The parameters of the ultrasound probe are as follows: The sampling rate is 40 MHz and probe frequency is 7.27 MHz, and the fractional bandwidth is 60%. A Hanning window is used for apodization, the single transmit focus is at 22.5 mm, equidistance receive foci are from 5 mm to 45 mm at each 5 mm, the transmit is sequential, and the number of active elements is 64. This phantom is uniform and is assumed to be isotropic and homogeneous.

The results show that accuracy of the classifier is substantially increased from about 94% to 98% (Table 1). Receiver Operating Characteristic (ROC) [48] curves for the simulation data set are also shown in Fig. 6. It shows that the area under the curve for using one NCC value and the proposed method are respectively 0.9606 and 0.9826, a substantial improvement achieved by using the NCC profile.

C. Phantom Data

An Antares Siemens system (Issaquah, WA) at the center frequency of 6.67 MHz with a VF10-5 linear array at a sampling rate of 40 MHz is used to acquire RF data. An RF data is collected from freehand palpation of a CIRS breast elastography phantom (CIRS, Norfolk, VA) with a lesion three times stiffer than the surrounding tissue.

TABLE I
CLASSIFICATION ACCURACY USING ONE NCC VALUE AND THE PROPOSED METHOD FOR THE SIMULATION DATA.

Data set	1-NCC	7-Features
Simulation data	93.9	98.4

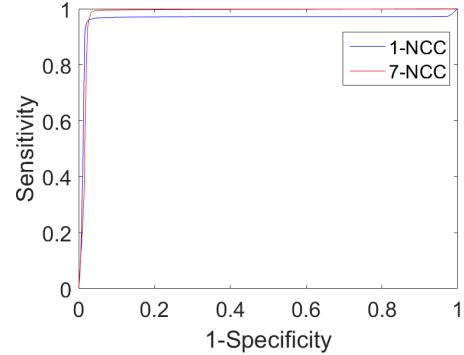


Fig. 6. ROC curve for the simulation data.

The DPAM method [45] is utilized to calculate the displacement field. In Table 2, the overall accuracy of the proposed classifier by using one NCC value and seven NCC values is shown. The substantial improvement from 82.2% to 96.7% is achieved by using the seven features. The ROC curves are shown in Fig. 7. The areas under curve for the one NCC method and the proposed method are, respectively, 0.8994 and 0.9846, showing substantial improvement.

D. In-Vivo Data

The RF data was collected by ablation therapy of three patients with liver cancer using an Antares Siemens (Issaquah, WA) ultrasound machine and A VF10-5 linear array at the center frequency of 6.67 MHz with a sampling rate of 40 MHz for the RF data acquisition at John Hopkins Hospital. The displacement matrix for all the three patients is calculated by the DPAM method [45].

The overall classification accuracy using one NCC value and seven NCC values are depicted in Table 3 for all the three patients. The accuracy is improved by more than 13% in the average by using the proposed method. Also, for patient 1 with the worst available data set in terms of quality of RF data, it shows a very high improvement of about 21%. Finally, the ROC curves for all the three patients are depicted in Fig. 8. The area under the ROC curve is increased from 0.6906 to 0.9255 for patient 1, from 0.8624 to 0.9598 for patient 2 and from 0.9610 to 0.9926 for patient 3. The significant

TABLE II
CLASSIFICATION ACCURACY USING ONE NCC VALUE AND THE PROPOSED METHOD FOR THE CIRS PHANTOM DATA.

Data set	1-NCC	7-Features
Phantom data	82.2	96.7

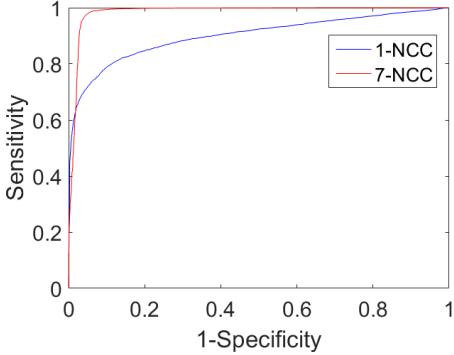


Fig. 7. ROC curve for the CIRS phantom data.

TABLE III

CLASSIFICATION ACCURACY USING ONE NCC VALUE AND PROPOSED SEVEN FEATURES FOR THREE PATIENT DATA.

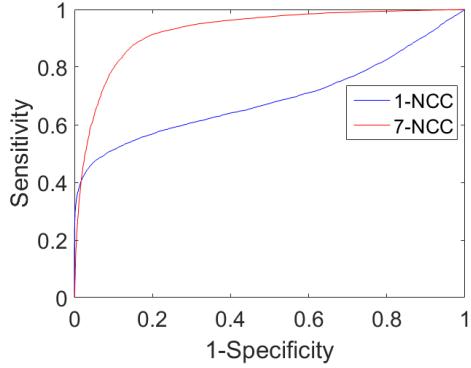
Data set	1-NCC	7-Features
Patient data 1	67.6	88.6
Patient data 2	79.3	92.2
Patient data 3	90.6	98.2
Average	79.2	93.0

improvement in the results is obtained by exploiting the NCC profile around the estimated displacement field.

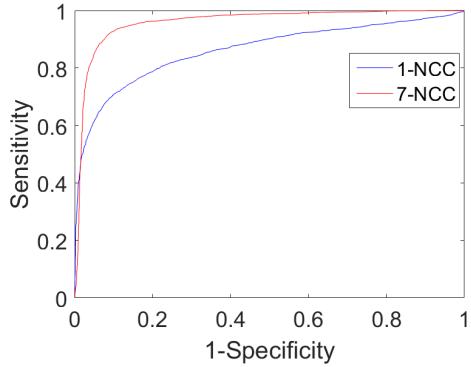
E. Accuracy Map in Region of Interest (ROI)

The proposed method has been quantitatively validated in the previous sections. To visualize the performance of the proposed method, we show the map of the accuracy assessment. In Figs. 9 (a) and (b), the red boxes in the figures indicate the tumor region in the B-mode ultrasound strain images for patient 3. The displacement field is calculated using DPAM and is visually checked. Therefore, all samples belong to the true set and the classifier should label those pixels as the true displacement estimate. Figs. 9 (c) and (d) show the results of the proposed method using one NCC and seven NCC values. Blue samples denote successfully classified regions and yellow samples show the unsuccessfully classified samples. In a second experiment, all the samples are either peak-hopping samples or are corrupted by the jitter error. Therefore, the algorithm should classify all the samples as incorrect displacement. Figs. 9 (e) and (f) show the result of the proposed method for using one NCC and seven NCC values, respectively. Again, blue and yellow respectively represent correct and incorrect classification. The results of this figure clearly demonstrate that the proposed method substantially improves the results.

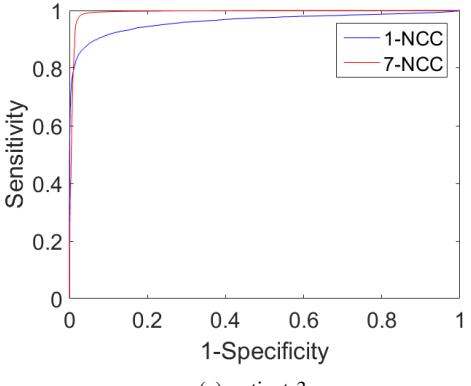
The accuracy maps are created for the True Positive (TP) and False Positive (FP) cases around the tumor. The tumor region is chosen as region of interest to locally show the performance of the classifiers using one NCC value or the seven proposed features. It is important to note that the samples in this region (i.e. the vicinity of the tumor) are not used for training and are only used in the testing stage. In the TP case, all the samples in the tumor region have correct



(a) patient 1



(b) patient 2



(c) patient 3

Fig. 8. ROC curves for three patients.

displacement estimates, which are obtained using DPAM. For the FP case, all the samples of the validation set have incorrect displacement estimates in the form of either peak-hopping or jitter error. These incorrect displacements are generated by the procedure described in Section II.C.

IV. DISCUSSION

The information around TDE is always informative in discriminating the correct versus incorrect displacement estimates. Since estimation of the displacement is generally more

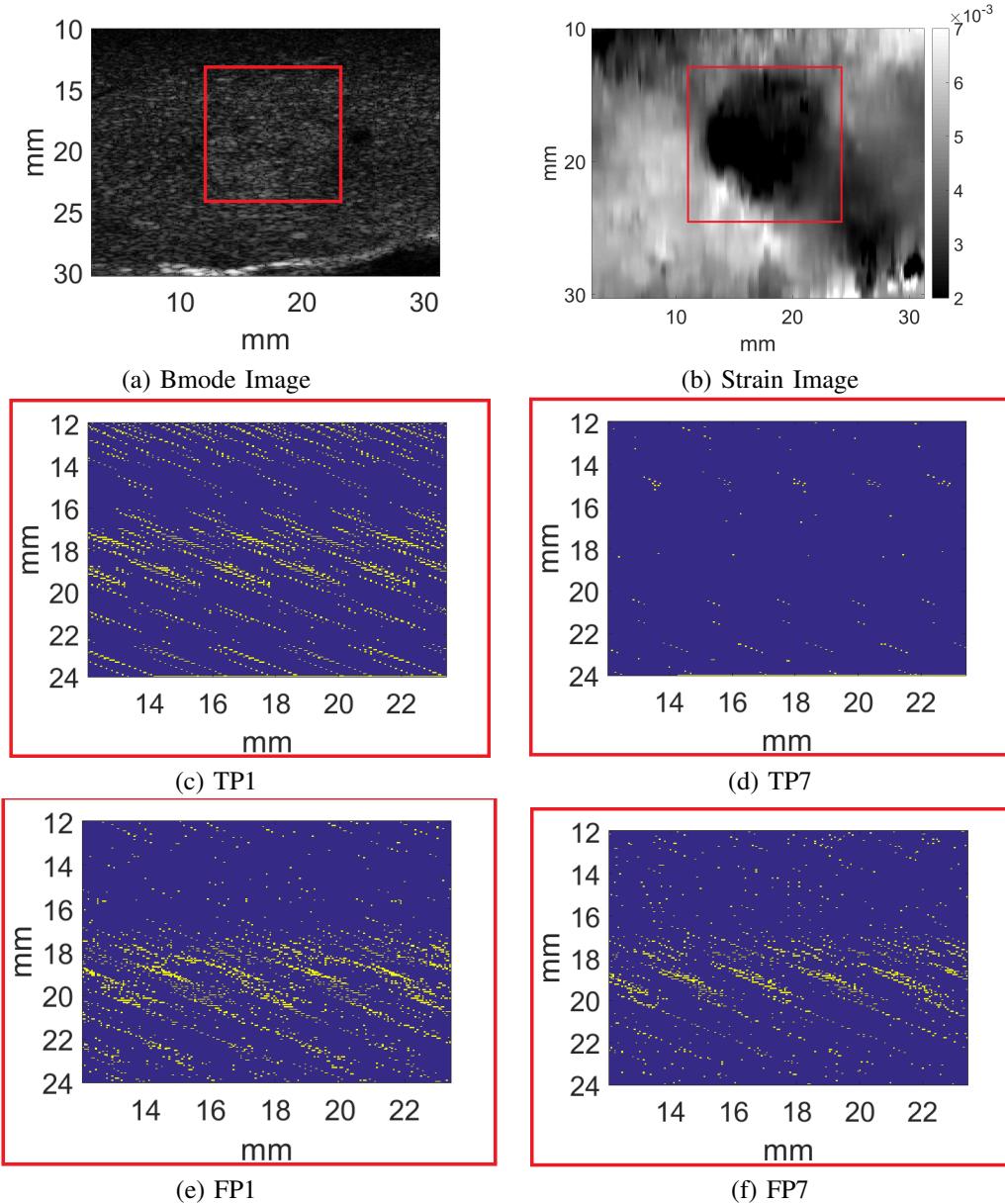


Fig. 9. Accuracy map in tumor region for patient 3 is shown in red boxes for true and false positive cases.

difficult close to boundaries between different organs or tissue types, it is more likely to find an incorrect displacement in these regions. Therefore, our method is more invaluable in these regions wherein incorrect displacement estimates may lead to incorrect diagnosis or surgical planning.

Although there is no gold standard for displacement fields for phantom and *in-vivo* data set, time-delay estimates are inspected visually to be accurate enough. In this paper, TDEs are obtained by DPAM, which is a promising method to obtain accurate displacement fields. In order to show that the obtained TDE is accurate, we changed the axial regularization weight by 20% in five equal steps (20 to 24) and ran the DPAM code. The mean squared error (MSE) for those displacement values is on average less than 0.02 sample (0.0004 mm) for all data sets (Table IV). The small MSE of the displacement fields quantitatively shows that the DPAM displacement estimates

TABLE IV
THE MEAN SQUARED ERROR (MSE) OF DISPLACEMENT VALUES OBTAINED USING DPAM FOR ALL DATA SETS USING DIFFERENT REGULARIZATION WEIGHTS.

Phantom	Patient1	Patient2	Patient3
0.0204	0.0172	0.0054	0.0039

have a low variance. Note that the 0.02 sample variance is substantially smaller than the 0.4 to 0.6 samples that are added to the correct displacement field to generate jitter.

In order to improve the accuracy of the classifier, the proposed seven features are utilized instead of using one NCC values at the estimated time-delay estimate. The results show that using the seven features improves the performance of the classifier substantially. However, one might consider additional

features that further improve the performance of the classifier. Future work will utilize a Random Forest [49] classifier to determine the most informative features [50].

The proposed automatic quality assessment of TDE is a novel approach which can play an important role in industrial and academic applications of ultrasound imaging specifically in quasi-static elastography. This method can be applied as a promising approach to quantitatively compare the performance of TDE algorithms, since it is more reliable than solely using the value of NCC at the estimated displacement. Moreover, this method is suitable for real-time applications, and therefore, can be used to train sonographers to obtain higher quality strain images by displaying the accuracy map instantaneously. Unsupervised classification methods, such as k-nearest neighbors (KNN) [51] have an advantage over the proposed method in that they do not require training data. However, they need to search for nearest neighbors in high-dimensional feature spaces and are not usually suitable for real-time applications. In contrast, the proposed supervised technique based on SVM runs in real-time, and therefore, provides clinically a more relevant solution.

V. CONCLUSIONS

We have presented a novel method for accuracy assessment of TDE using NCC profile around the estimated displacement value, which can be used to mask out the erroneous regions of the strain image. It can further reduce the user-dependence of strain imaging and help train the sonographer. Our technique is based on SVM classification, a nonlinear classifier that often substantially outperforms linear classifiers. Training our SVM classifier is computationally expensive, but it can be performed offline. Once trained, our classifier is computationally efficient and can classify the accuracy of TDE in real-time. The performance of the proposed method is validated through simulation, phantom and *in-vivo* data.

ACKNOWLEDGEMENTS

The liver data was collected at Johns Hopkins Hospital. We thank the principal investigators Dr. Emad M. Boctor, Dr. Michael A. Choti and Dr. Gregory D. Hager for sharing this data with us. We would like to also thank the anonymous reviewers for constructive comments. This work was funded by NSERC Discovery grants RGPIN-2015-04136 and RGPIN-06623.

REFERENCES

- [1] J. Ophir, I. Cesar, H. Ponnekanti, Y. Yazdi, and X. Li, "Elastography: a quantitative method for imaging the elasticity of biological tissues," *Ultrasonic Imaging*, vol. 13, no. 2, pp. 111–134, 1991.
- [2] M. O'Donnell, A. R. Skovoroda, B. M. Shapo, S. Y. Emelianov *et al.*, "Internal displacement and strain imaging using ultrasonic speckle tracking," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 41, no. 3, pp. 314–325, 1994.
- [3] S. K. Alam, J. Ophir, and E. E. Konofagou, "An adaptive strain estimator for elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 45, no. 2, pp. 461–472, March 1998.
- [4] A. Pesavento, C. Perrey, M. Krueger, and H. Ermert, "A time-efficient and accurate strain estimation concept for ultrasonic elastography using iterative phase zero estimation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 46, no. 5, pp. 1057–1067, Sept 1999.
- [5] C. Pellet-Barakat, F. Frouin, M. F. Insana, and A. Herment, "Ultrasound elastography based on multiscale estimations of regularized displacement fields," *IEEE transactions on medical imaging*, vol. 23, no. 2, pp. 153–163, 2004.
- [6] E. Brusseau, J. Kybic, J.-F. Déprez, and O. Bassett, "2-d locally regularized tissue strain estimation from radio-frequency ultrasound images: Theoretical developments and results on experimental data," *IEEE Transactions on Medical Imaging*, vol. 27, no. 2, pp. 145–160, 2008.
- [7] T. Shiina, K. R. Nightingale, M. L. Palmeri, T. J. Hall, J. C. Bamber, R. G. Barr, L. Castera, B. I. Choi, Y.-H. Chou, D. Cosgrove *et al.*, "Wfumb guidelines and recommendations for clinical use of ultrasound elastography: Part 1: basic principles and terminology," *Ultrasound in Medicine & Biology*, vol. 41, no. 5, pp. 1126–1147, 2015.
- [8] R. G. Barr, K. Nakashima, D. Amy, D. Cosgrove, A. Farrokhi, F. Schafer, J. C. Bamber, L. Castera, B. I. Choi, Y.-H. Chou *et al.*, "Wfumb guidelines and recommendations for clinical use of ultrasound elastography: Part 2: breast," *Ultrasound in Medicine & Biology*, vol. 41, no. 5, pp. 1148–1160, 2015.
- [9] G. Ferraioli, C. Filice, L. Castera, B. I. Choi, I. Sporea, S. R. Wilson, D. Cosgrove, C. F. Dietrich, D. Amy, J. C. Bamber *et al.*, "Wfumb guidelines and recommendations for clinical use of ultrasound elastography: Part 3: liver," *Ultrasound in Medicine & Biology*, vol. 41, no. 5, pp. 1161–1179, 2015.
- [10] D. Cosgrove, R. Barr, J. Bojunga, V. Cantisani, M. C. Chammas, M. Dighe, S. Vinayak, J.-M. Xu, and C. F. Dietrich, "Wfumb guidelines and recommendations on the clinical use of ultrasound elastography: Part 4: thyroid," *Ultrasound in Medicine & Biology*, vol. 43, no. 1, pp. 4–26, 2017.
- [11] R. G. Barr, D. Cosgrove, M. Brock, V. Cantisani, J. M. Correas, A. W. Postema, G. Salomon, M. Tsutsumi, H.-X. Xu, and C. F. Dietrich, "Wfumb guidelines and recommendations on the clinical use of ultrasound elastography: Part 5: prostate," *Ultrasound in Medicine & Biology*, vol. 43, no. 1, pp. 27–48, 2017.
- [12] M. Tanter, J. Bercoff, A. Athanasiou, T. Deffieux, J.-L. Gennisson, G. Montaldo, M. Muller, A. Tardivon, and M. Fink, "Quantitative assessment of breast lesion viscoelasticity: initial clinical results using supersonic shear imaging," *Ultrasound in Medicine & Biology*, vol. 34, no. 9, pp. 1373–1386, 2008.
- [13] M. L. Palmeri and K. R. Nightingale, "Acoustic radiation force-based elasticity imaging methods," *Interface Focus*, p. rsfs20110023, 2011.
- [14] T. J. Hall, P. Barbone, A. A. Oberai, J. Jiang, J. F. Dord, S. Goenezen, and T. G. Fisher, "Recent results in nonlinear strain and modulus imaging," *Current Medical Imaging Reviews*, vol. 7, no. 4, p. 313, 2011.
- [15] G. Treece, J. Lindop, L. Chen, J. Housden, R. Prager, and A. Gee, "Real-time quasi-static ultrasound elastography," *Interface Focus*, vol. 1, no. 4, pp. 540–552, 2011.
- [16] M. Doyley, "Model-based elastography: a survey of approaches to the inverse elasticity problem," *Physics in Medicine and Biology*, vol. 57, no. 3, p. R35, 2012.
- [17] S. R. Mousavi, A. Sadeghi-Naini, G. J. Czarnota, and A. Samani, "Towards clinical prostate ultrasound elastography using full inversion approach," *Medical Physics*, vol. 41, no. 3, 2014.
- [18] O. A. Babaniyi, A. A. Oberai, and P. E. Barbone, "Recovering vector displacement estimates in quasistatic elastography using sparse relaxation of the momentum equation," *Inverse Problems in Science and Engineering*, vol. 25, no. 3, pp. 326–362, 2017.
- [19] C. Hoerig, J. Ghoboussi, and M. F. Insana, "An information-based machine learning approach to elasticity imaging," *Biomechanics and modeling in mechanobiology*, vol. 16, no. 3, pp. 805–822, 2017.
- [20] W. F. Walker and G. E. Trahey, "A fundamental limit on delay estimation using partially correlated speckle signals," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 42, no. 2, pp. 301–308, March 1995.
- [21] L. Chen, G. M. Treece, J. E. Lindop, A. H. Gee, and R. W. Prager, "A quality-guided displacement tracking algorithm for ultrasonic elasticity imaging," *Medical Image Analysis*, vol. 13, no. 2, pp. 286–296, 2009.
- [22] R. Zahiri-Azar and S. E. Salcudean, "Motion estimation in ultrasound images using time domain cross correlation with prior estimates," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 10, pp. 1990–2000, 2006.
- [23] Y. Petrank, L. Huang, and M. O'Donnell, "Reduced peak-hopping artifacts in ultrasonic strain estimation using the viterbi algorithm," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 56, no. 7, pp. 1359–1367, 2009.
- [24] M. H. Wang, M. L. Palmeri, V. M. Rotemberg, N. C. Rouze, and K. R. Nightingale, "Improving the robustness of time-of-flight based shear

- wave speed reconstruction methods using ransac in human liver in vivo," *Ultrasound in Medicine & Biology*, vol. 36, no. 5, pp. 802–813, 2010.
- [25] C. Pellet-Barakat, M. Lefort, L. Chami, M. Labit, F. Frouin, and O. Lucidarme, "Automatic assessment of shear wave elastography quality and measurement reliability in the liver," *Ultrasound in medicine & biology*, vol. 41, no. 4, pp. 936–943, 2015.
- [26] G. C. Carter, *Coherence and time delay estimation: an applied tutorial for research, development, test, and evaluation engineers*. IEEE, 1993.
- [27] W. F. Walker and G. E. Trahey, "A fundamental limit on delay estimation using partially correlated speckle signals," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 42, no. 2, pp. 301–308, 1995.
- [28] A. Weiss and E. Weinstein, "Fundamental limitations in passive time delay estimation—part i: Narrow-band systems," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 2, pp. 472–486, 1983.
- [29] I. Cespedes, M. Insana, and J. Ophir, "Theoretical bounds on strain estimation in elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 42, no. 5, pp. 969–972, 1995.
- [30] T. Varghese and J. Ophir, "A theoretical framework for performance characterization of elastography: The strain filter," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 44, no. 1, pp. 164–172, 1997.
- [31] Y.-C. Chang, M.-C. Yang, C.-S. Huang, S.-C. Chang, G.-Y. Huang, W. K. Moon, and R.-F. Chang, "Automatic selection of representative slice from cine-loops of real-time sonoelastography for classifying solid breast masses," *Ultrasound in Medicine & Biology*, vol. 37, no. 5, pp. 709–718, 2011.
- [32] S.-C. Chang, Y.-W. Lee, Y.-C. Lai, C.-M. Tiu, H.-K. Wang, H.-J. Chiou, Y.-W. Hsu, Y.-H. Chou, and R.-F. Chang, "Automatic slice selection and diagnosis of breast strain elastography," *Medical Physics*, vol. 41, no. 10, p. 102902, 2014.
- [33] E. Shaswary, Y. Xu, and J. Tavakkoli, "Performance study of a new time-delay estimation algorithm in ultrasonic echo signals and ultrasound elastography," *Ultrasonics*, vol. 69, pp. 11–18, 2016.
- [34] R. Xia, G. Tao, and A. K. Thittai, "Dynamic frame pairing in real-time freehand elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 6, pp. 979–985, 2014.
- [35] F. Kallel and J. Ophir, "A least-squares strain estimator for elastography," *Ultrasonic Imaging*, vol. 19, no. 3, pp. 195–208, 1997.
- [36] A. Gee, J. Lindop, G. Treece, R. Prager, and S. Freeman, "Stable, intelligible ultrasonic strain imaging," *Ultrasonics*, vol. 16, no. 4, pp. 187–192, 2008.
- [37] G. M. Treece, J. E. Lindop, A. H. Gee, and R. W. Prager, "Uniform precision ultrasound strain imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 11, pp. 2420–2436, 2009.
- [38] J. E. Lindop, G. M. Treece, A. H. Gee, and R. W. Prager, "An intelligent interface for freehand strain imaging," *Ultrasound in Medicine & Biology*, vol. 34, no. 7, pp. 1117–1128, 2008.
- [39] H. Khodadadi, A. G. Aghdam, and H. Rivaz, "Edge-preserving ultrasonic strain imaging with uniform precision," in *Proc. 37th IEEE International Conference on Engineering, Medicine and Biology (EMBC)*, 2015, pp. 3835–3838.
- [40] J. Jiang, T. J. Hall, and A. M. Sommer, "A novel performance descriptor for ultrasonic strain imaging: a preliminary study," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 53, no. 6, pp. 1088–1102, 2006.
- [41] B. R. Chintada, A. V. Subramani, B. Raghavan, and A. K. Thittai, "A novel elastographic frame quality indicator and its use in automatic representative-frame selection from a cine loop," *Ultrasound in Medicine & Biology*, 2016.
- [42] H. Rivaz, E. M. Boctor, M. A. Choti, and G. D. Hager, "Ultrasound elastography using multiple images," *Medical Image Analysis*, vol. 18, no. 2, pp. 314–329, 2014.
- [43] Y. R. Rao, N. Prathapani, and E. Nagabhooshanam, "Application of normalized cross correlation to image registration," *International Journal of Research in Engineering and Technology*, vol. 3, no. 05, pp. 12–16, 2014.
- [44] V. Vapnik, S. E. Golowich, A. Smola *et al.*, "Support vector method for function approximation, regression estimation, and signal processing," *Advances in Neural Information Processing Systems*, pp. 281–287, 1997.
- [45] H. Rivaz, E. M. Boctor, M. A. Choti, and G. D. Hager, "Real-time regularized ultrasound elastography," *IEEE Transactions on Medical Imaging*, vol. 30, no. 4, pp. 928–945, 2011.
- [46] H. Rivaz, E. Boctor, P. Foroughi, R. Zellars, G. Fichtinger, and G. Hager, "Ultrasound elastography: a dynamic programming approach," *IEEE Transactions on Medical Imaging*, vol. 27, no. 10, pp. 1373–1377, 2008.
- [47] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10th Nordicbaltic conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1: 351–353*. Citeseer, 1996.
- [48] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [49] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [50] Z. Karimaghloo, H. Rivaz, D. L. Arnold, D. L. Collins, and T. Arbel, "Adaptive voxel, texture and temporal conditional random fields for detection of gad-enhancing multiple sclerosis lesions in brain mri," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 543–550.
- [51] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.



Mohamad Ghasemi Amidabadi received his B.Sc. degree from the Amirkabir University of Technology (Tehran Polytechnic) in electrical engineering double major with biomedical engineering, Tehran, Iran. He is currently continuing his M.A.Sc. degree in electrical and computer engineering at Concordia University, Montreal, Canada. His current research interests include data science, machine learning, medical imaging, ultrasound elastography, and signal processing.



M. Omair Ahmad received the B.Eng. degree from Sir George Williams University, and the Ph.D. degree from Concordia University, both in electrical engineering. He holds the Concordia University Research Chair (Tier I) in Multimedia Signal Processing. He was a Founding Researcher at Micronet from its inception in 1990 as a Canadian Network of Centers of Excellence until its expiration in 2004. He was an Associate Editor of the IEEE Trans. on Circuits and Systems Part I: Fundamental Theory and Application from June 1999 to December 2001. He was the Local Arrangements Chairman of the 1984 IEEE International Symposium on Circuits and Systems. In 1988, he was a member of the Admission and Advancement Committee of the IEEE. He has served as the Program Co-Chair for the 1995 IEEE International Conference on Neural Networks and Signal Processing, the 2003 IEEE International Conference on Neural Networks and Signal Processing, and the 2004 IEEE International Midwest Symposium on Circuits and Systems. He was a General Co-Chair for the 2008 IEEE International Conference on Neural Networks and Signal Processing. He is the Chair of the Montreal Chapter IEEE Circuits and Systems Society. He is a recipient of numerous honors and awards, including the Wighton Fellowship from the Sandford Fleming Foundation, an induction to Provosts Circle of Distinction for Career Achievements, and the Award of Excellence in Doctoral Supervision from the Faculty of Engineering and Computer Science of Concordia University.



Hassan Rivaz received his B.Sc. from Sharif University of Technology, his M.A.Sc. from University of British Columbia, and his Ph.D. from Johns Hopkins University. He did a post-doctoral training at McGill University. He is now an Assistant Professor and a Petro-Canada Young Innovator at the Department of Electrical and Computer Engineering and the PERFORM Centre at Concordia University, where he is directing the IMage Processing And Characterization of Tissue (IMPACT) Lab. He is an Associate Editor of IEEE Trans. on Medical Imaging and an Area Chair of MICCAI 2017. His main research interests are medical image processing, ultrasound elastography, ultrasound imaging, registration, segmentation and machine learning. Clinical applications of his work are in neurosurgery, liver surgery, breast and prostate radiotherapy and musculoskeletal imaging.