IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 15, 2019, accepted September 7, 2019, date of publication September 10, 2019,
date of current version September 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2940645

A Deep Siamese-Based Plantar Fasciitis Classification Method Using Shear Wave Elastography

JUNLING GAO[1], LEI XU[1,2], AYACHE BOUAKAZ[3], AND MINGXI WAN[1], (Member, IEEE)
[1]Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China
[2]Xi'an Hospital of Traditional Chinese Medicine, Xi'an 710021, China
[3]UMR 1253, iBrain, Université de Tours, Inserm, 37032 Tours, France

Corresponding author: Mingxi Wan (mxwan@mail.xjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 81827801, Grant 81771854, and Grant 11874049.

ABSTRACT Two-dimensional shear wave elastography (2D-SWE) is an effective and feasible method for plantar fasciitis (PF) evaluation. Until now, only experienced doctors have been able to give relatively accurate evaluation via ultrasound images, resulting in low efficiency and high cost. Therefore, designing automatic algorithms to recognize the pattern of these ultrasound images is urgently required. In recent years, deep learning (DL) has made considerable progress in computer-aided diagnosis (CAD). However, there have been no studies that apply DL to the diagnosis of PF. To achieve robust PF classification, this paper builds a deep Siamese framework with multitask learning and transfer learning (DS-MLTL), which learns discriminative visual features and effective recognition functions using 2D-SWE. The DS-MLTL model comprises two VGG-style branches and a multitask loss including a classification loss and a Siamese loss. The Siamese loss leverages the intrinsic structure (similarities) of different images and contains a contrastive constraint and a similar constraint. In our framework, visual features and the multitask loss are learned jointly, and they can benefit from each other. To train the DS-MLTL model effectively, the model transfers knowledge from the large-scale ImageNet dataset to the PF classification task. For model evaluation, an SWE dataset of plantar fascia, which contains 282 images of a PF pattern and 60 images of a healthy pattern, is collected. Experimental results show that the DS-MLTL method achieves favorable accuracy of $85.09 \pm 6.67\%$ and performs better than human-crafted features extracted from B-mode ultrasound and SWE. In addition, DS-MLTL also obtains the best performance compared with different DL models.

INDEX TERMS Plantar fasciitis, Siamese network, transfer learning, shear wave elastography.

I. INTRODUCTION

Plantar Fasciitis (PF), a commonly reported cause of plantar heel pain, is a chronic disease caused by aseptic inflammation of plantar fascia [1]. The etiology and pathogenesis of PF remain unclear. Early diagnosis and treatment of PF are essential for improving the patients' condition and quality of life, and ultrasound is the most common method for classifying PF patterns and assisting clinical diagnosis [2]. Ultrasound diagnosis of this disease contains edema, thickening, and hypoechogenicity of plantar fascia. Some studies [3], [4] also found calcification in tendons in some patients. Recently, plantar-fascia stiffness measurement (PFSM) based on noninvasive ultrasonic imaging has been strongly recommended by many studies [5], [6] because the stiffness of biological tissue can be used as an important basis for disease diagnosis and the essence of PF is the abnormal elastic stress of the arch of a foot.

Two-dimensional (2D) shear wave elastography (SWE), which integrates B-mode imaging and color-coded tissue stiffness maps, is an effective stiffness measurement technology with real-time and high stability. 2D-SWE has been widely applied for the clinical diagnosis of various diseases, such as breast cancer [7], prostate cancer [8], and malignancy in thyroid nodules [9]–[11]. 2D-SWE has also been increasingly used in the evaluation of muscle lesions, tendons, fascia, and neuropathy [12]. Although the PFSM of 2D-SWE has become one of the most common methods for the clinical diagnosis of PF [13], 2D-SWE is still subject to many limitations. The definitive standard for defining the optimal region

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

of interest (ROI) of PFSM, monitoring the overall image quality and determining reliable and unreliable measurements is still ambiguous. This results in the cut-off of SWE values for PF classification showing great variability. Some studies have shown that the cut-off of SWE values for PF is 36.8 ± 7.7 kPa [13]. Therefore, it is inadequate to distinguish PF only by SWE values. Moreover, diagnoses based on ultrasound image technology are limited by the doctor's experience. Even in the same case, different doctors may give different diagnostic results.

To overcome the above issues, radiomics, an emerging computer-aided diagnosis (CAD) technology, which can automatically quantify large amounts of medical image features to uncover disease that is unable to be recognized by doctors' naked eyes [14]–[16], has drawn considerable attention in recent years. Many image processing methods have been proposed [17]–[21] for radiomics. Most of these methods rely on traditional hand-crafted image features, which may not be robust enough to capture the discriminative features for disease diagnoses. In addition, these methods only learn a linear or shallow classification function, which is not comprehensive enough to represent the mapping from image features to classification results. Compared with these shallow models using hand-crafted features, the deep learning (DL) models can automatically extract robust features in an ''end-to-end'' way and learn effective nonlinear functions to classify disease. Many researchers have shown that DL models are valuable for CT, MR and ultrasound images [22]. For example, Shi et al. proposed a novel stacked deep polynomial network for tumor classification, which achieved favorable performance with a small ultrasound image dataset [23]. Furthermore, there is also some evidence showing that DL models can obtain more discriminative information from total 2D-SWE images than only 2D-SWE values from ROI [24] and achieve higher accuracy in disease assessment. Recently, various DL architectures for 2D-SWE image analysis have been proposed, such as the DL model for classifying breast tumors with 2D-SWE images [25], and convolutional neural networks (CNN) for liver fibrosis stage determination [24].

Developing a DL-based CAD framework for plantar fasciitis (PF) very significant to efficient diagnosis. To the best of our knowledge, no work has applied DL for the CAD of PF. As a result, it is a promising direction to develop a robust DL model for PF classification, which can improve the accuracy of diagnosis and reduce the burden of doctors. Although we can directly use traditional DL models (e.g., CNNs such as AlexNet and VGGNet) to perform PF classification, these models ignore the intrinsic structures between 2D-SWE images. In fact, a PF image should be more similar to another PF image than a healthy image, and vice versa. If a deep model can learn similar/dissimilar features for images with the same/different labels, learning the PF classification task will be facilitated. Therefore, adding comparisons of different images during the learning process will help improve the discriminative ability of the model.

Motivated by the above observations, this study builds a deep Siamese framework via multitask learning and transfer learning (DS-MLTL), which learns visual features and effective nonlinear classification functions to identify PF in a unified framework using SWE. The method adopts a Siamese network to evaluate whether an image pair is similar. The key idea behind a Siamese network is to learn a matching function that constrains similar pairs (i.e., images with the same label) to be close to each other while dissimilar pairs (i.e., images with different labels) to be separated by a predefined margin. Compared with traditional Siamese-based networks, which only adopt similarity learning, the proposed DS-MLTL preforms multitask learning, where the classification task and the similarity learning task are learned jointly and can promote each other. The classification task can learn effective class-specific information, which is crucial for learning a robust similarity function. Similarity learning can build a matching function for guiding more robust feature learning, which facilitates the classification process. As shown in Figure 1, in the training phase, a pair of 2D-SWE images is taken as input to the Siamese network with two identical branches. This network then outputs both classification results and similarity scores of the two images. In the test phase, only the output classification results are used. For training data, compared to datasets in regular computer vision tasks that have millions of training samples, medical image datasets are generally small (e.g., hundreds/thousands of samples). Therefore, transfer learning provides an alternative for radiomics with small datasets. A simple and effective transfer learning methodology uses a pretrained network (typically on large-scale natural images) for model fine-tuning on new tasks. Here, the model uses the convolutional layers of VGG-16 net [26] as a feature extractor and fine-tunes the last 3 fully connected layers with the multitask loss. To evaluate the performance of the DS-MLTL for PF classification, a doctor-labeled high-quality 2D-SWE image dataset, which contains 282 images of a PF pattern and 60 images of a healthy pattern, is collected.

The main contributions of this paper are as follows:

- To the best of our knowledge, this is the first prospective study that aims to utilize a DL model for the intelligent diagnosis of PF using 2D-SWE. It is helpful for conducting an effective diagnosis of PF, reducing the rate of misdiagnosis and missed diagnosis, and decreasing the burden of doctors.
- Compared with a conventional single classification task with a single diagnostic variable as output (e.g., disease present or not), the DS-MLTL model adopts a multitask learning strategy to simultaneously perform a classification task and similarity learning task. Common hidden representations shared between both related tasks improve the performance of the model. We also use transfer learning to address the problem of lacking training data.
- Experimental results demonstrate that the deep Siamese-based method achieves favorable accuracy of 85.09 ±

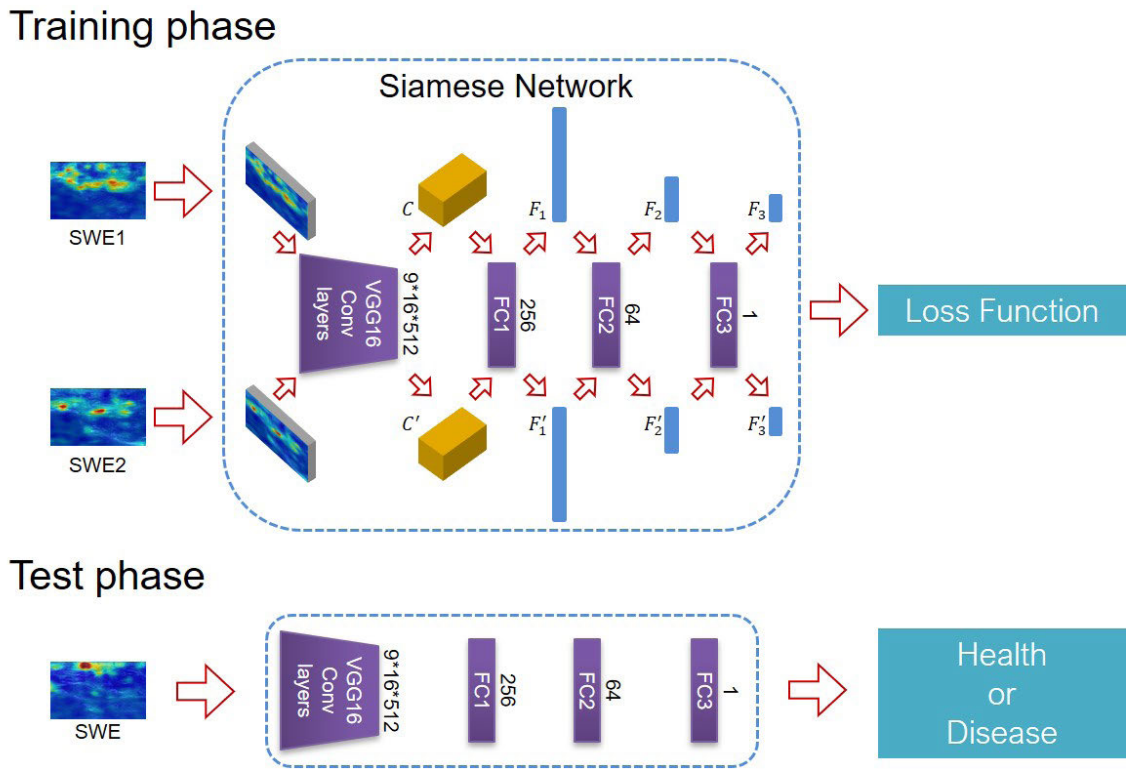

FIGURE 1. Architecture of the DS-MLTL model, which consists of 16 convolutional layers (transferred from VGG-16 Conv layers) and three fully connected layers. The number on the right side of the layer represents the dimension of the output layer. The total loss of our model is comprised of the classification losses for all samples, the Siamese losses for all sample pairs and the regularization item. In the test phase, only a single branch with the learned weights is used for PF classification.

6.67% and performs better than human-crafted features extracted from B-mode ultrasound and SWE.

II. PATIENTS AND METHODS

In this section, the design and overview of this study are illustrated first. Next, how to conduct data preprocessing is introduced in detail. Finally, the deep Siamese-based model architecture and the loss function are shown.

A. DESIGN AND OVERVIEW

The aim of this study is to design a high-performance algorithm to realize CAD of PF effciently and accurately. Although traditional DL models are capable of performing image classification tasks, this study uses a deep Siamese-based method to further improve discriminative ability. All the data for our study are provided by Xi'an Hospital of Traditional Chinese Medicine and the study was approved by the ethics committee of the principal investigator's hospital. The clinical results of PF are used as the ground truth. To prove the superiority of the DS-MLTL model, the performance comparisons between four DL models (CNN, transfer learning model, DS-ML and DS-MLTL) are performed first, and then the performance comparisons between the DS-MLTL model and various features of sono and elastograms (thickness, hypoechogenicity, elasticity value.) are implemented. Moreover, the DS-MLTL model is trained on different size datasets to test the performance of the model.

B. DATA COLLECTION AND PREPROCESSING

In this work, an SWE dataset is obtained by a Mindray R7 scanner with an L14-6 MHz transducer. During an ultrasound session, a B-mode ultrasound scan is first performed, and then 2D-SWE is implemented near the heal of the participants. The thickness of the plantar fascia and corresponding B-mode ultrasound images are obtained, and then the 2D-SWE value and corresponding 2D-SWE images are obtained from each patient. Two doctors with 10 years of ultrasound operating experience are employed as quality controllers for reviewing all 2D-SWE images and excluding unqualified images. Finally, a total of 342 2D-SWE images of plantar-fascia, including 282 PF and 60 healthy images, is collected. However, the 2D-SWE dataset is still too small to achieve good performance by using DL models. To avoid overfitting, data augmentation is conducted for the training set and test set by a cutoff of 5 or 10 pixels from four directions of the raw images (See Figure 2A). Next, the images are normalized to a fixed size of $313 * 526$, which is the largest height and width in the dataset. These images have three channels (RGB). Twenty healthy individuals and 20 patients with PF were selected randomly as the test set while maintaining the others as the training set. It is noteworthy that

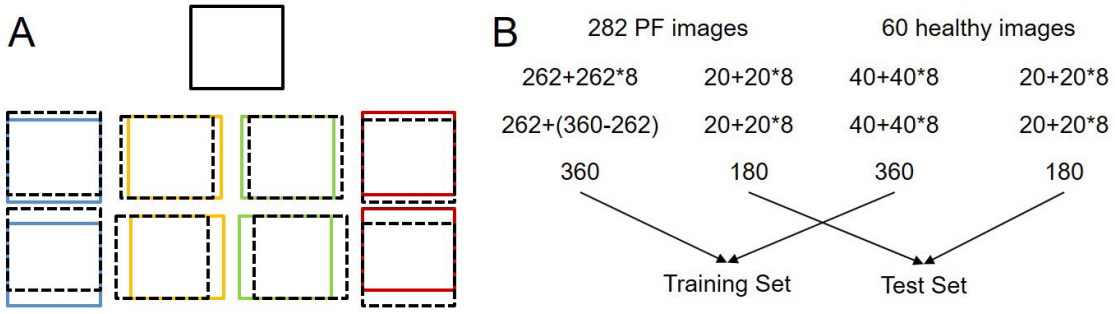

FIGURE 2. Illustrations of data augmentation and split. (A) Data augmentation for the training and test sets. The image is translated by 5 or 10 pixels from four directions. (B) Raw images and augmentation images are divided into the training and test set.

the constructed dataset is balanced: (1) For the training data, as shown in Figure 2B, we conduct data augmentation for the training samples (262 PF images and 40 healthy images). After data augmentation, we have $262 + 262 * 8 = 2358$ PF images and 360 healthy images. We then choose 360 PF images (including the initial 262 PF images and randomly selected 98 augmented PF images) and all the healthy images as training data. As a result, the ratio of positive to negative training samples is 1:1, which is balanced. (2) For the test data, as shown in Figure 2B, 20 PF images and 20 healthy images (180 PF images and 180 healthy images after data augmentation) are adopted, which is also balanced.

C. OUR DEEP MODEL ARCHITECTURE

In this paper, the CAD of PF is formulated as a multitask learning framework, which includes a classification task and a similarity learning task. The similarity learning task can guide more effective visual feature learning, which is crucial for improving the classification task. The details are as follows. For classification, given an image $\mathbf{x}$, the model learns a classification function $f(\mathbf{x})$ to determine whether this image belongs to PF. For similarity learning, the model aims to learn a similarity metric function $g(\mathbf{x}, \mathbf{y})$ to determine whether the input image pairs belong to set P or Q. Here, P is used to denote a set of similar image pairs and Q is used to denote a set of dissimilar image pairs. If the image pair $(\mathbf{x}, \mathbf{y}) \in P$, it means that the images $\mathbf{x}$ and $\mathbf{y}$ have the same classification label. If the image pair $(\mathbf{x}, \mathbf{y}) \in Q$, $\mathbf{x}$ and $\mathbf{y}$ have different labels. To achieve the above goals, a novel DS-MLTL learning framework is built, as shown in Figure 1. Here, the training and test phases of DS-MLTL are shown. The DS-MLTL model contains two identical branches with shared weights. Each branch includes 16 convolutional layers (transferred from VGG-16 [26] Conv layers) for feature extraction and 3 fully connected layers (FC1, FC2, FC3) for feature fine-tuning. The output sizes of the FC layers are shown in Figure 1. In the training phase, an image pair ($\mathbf{x}$ and $\mathbf{y}$) is simultaneously input to the DS-MLTL model. In the forward propagation, different features of the image pair are calculated by the shared weights in the convolutional and fully connected layers. Although the same branch is adopted to learn a nonlinear classification function, the two images are separately propagated forward through each layer. The Euclidean distance of the two output vectors of the FC1 layer $\left(F_1 \text{ and } F_1'\right)$ is calculated to denote the similarity of the two images. The output of the FC3 layer $\left(F_3 \text{ and } F_3'\right)$ denotes the classification results of the input images. The weights of the convolutional layers are fixed to avoid overfitting. In the fully connected layers, the dropout operator is carried out after FC1 and FC2 with a dropout rate of 0.5, which is not shown in Figure 1. The rectified linear units (RELU) [27], as the activation function, is used after FC1 and FC2. The sigmoid activation function is used after FC3 for classification. In the test phase, only the single branch with learned weights is used for PF classification.

D. LOSS FUNCTION

As a multitask learning model, the total loss of the DS-MLTL model is comprised of the classification losses for all samples, the Siamese losses for all sample pairs and the regularization item. The stochastic gradient descent (SGD) is adopted for the optimization of the DS-MLTL. For an image $\mathbf{x}$, the model utilizes a binary cross-entropy to optimize the classification loss as follows:

$$\mathcal{L}_c(\mathbf{x}) = l_c \log f(\mathbf{x}) + (1 - l_c) \log(1 - f(\mathbf{x})) \tag{1}$$

where $l_c \in \{0, 1\}$ is the classification label. $l_c = 0$ indicates that $\mathbf{x}$ belongs to the healthy pattern, while $l_c = 1$ means $\mathbf{x}$ belongs to the PF pattern. $f(\mathbf{x})$ is the learned classification score for $\mathbf{x}$, indicating the PF probability of $\mathbf{x}$. For an image pair ($\mathbf{x}$ and $\mathbf{y}$), the model optimizes the Siamese loss to learn the similarity between the pair of images. The Siamese loss consists of two constraints: a contrastive constraint and a similarity constraint of the image pair ($\mathbf{x}$ and $\mathbf{y}$). The two constraints and the Siamese loss are described as follows:

$$\mathcal{L}_p(\mathbf{x}, \mathbf{y}) = s\left(\tau - D^2\right) \tag{2}$$

$$\mathcal{L}_q(\mathbf{x}, \mathbf{y}) = D^2 \tag{3}$$

$$\mathcal{L}_s(\mathbf{x}, \mathbf{y}) = l_d \mathcal{L}_p(\mathbf{x}, \mathbf{y}) + (1 - l_d)\, \mathcal{L}_q(\mathbf{x}, \mathbf{y}) \tag{4}$$

Here, τ is a threshold that enforces the similarity margin of image $\mathbf{x}$ and $\mathbf{y}$. $\mathrm{D} = \left\|F_1 - F_1'\right\|_2$ is the Euclidean distance

of two learned features, and $\left(F_1 \text{ and } F_1'\right)$ are two vectors generated by the first fully connected layer. $s(x) = \max(0, x)$ is the nonsaturating nonlinearity function. $l_d \in \{0, 1\}$ is an indicator scalar. $l_d = 0$ indicates that image $\mathbf{x}$ is similar to image $\mathbf{y}$, while $l_d = 1$ indicates that image $\mathbf{x}$ and image $\mathbf{y}$ are dissimilar. Finally, the total loss is defined as (5).

$$\mathcal{L} = \frac{1}{2|G|}\left[\sum_{\mathbf{x}\in\mathbf{X}} \mathcal{L}_c(\mathbf{x}) + \sum_{(\mathbf{x},\mathbf{y})\in\mathcal{G}} \mathcal{L}_s(\mathbf{x}, \mathbf{y})\right] + \lambda_1 \|\Theta\|_F^2 \quad (5)$$

Here, $\mathbf{X}$ denotes all the training images. $\mathcal{G} = \mathcal{P} \cup Q$ contains all the similar and dissimilar image pairs. Θ denotes all weights of the DS-MLTL model. We use the Frobenius norm of all the learnable weights as the model regularization, which can limit the model complexity and avoid the overfitting problem. In the DS-MLTL model, the PF classification loss and Siamese loss are learned jointly in a unified framework, where they can enhance and complement each other. The classification loss can learn effective class-specific information, which is crucial for similarity metric learning. Additionally, the Siamese loss can learn a better matching function to guide more robust visual feature learning, which facilitates the classification process. For the test phase, the input is a single image, and the DS-MLTL model uses one branch with the learned weights to predict whether the input image is healthy.

III. EXPERIMENTS AND RESULTS

A. TRAINING AND TEST SETS

The proposed DS-MLTL algorithm is evaluated on the collected 2D-SWE dataset, which contains 282 images of PF pattern and 60 images of healthy pattern. After data augmentation, 720 images (360 images with healthy labels and 360 images with PF labels) are selected as training images and 360 images are selected as test images (180 images with healthy label and 180 images with PF label). See Figure 2 for details. Each training image is paired with other training images with either the same category label or a different category label.

B. IMPLEMENTATION DETAILS

The DS-MLTL model is implemented using the public deep learning library Keras. The parameter τ in the Siamese loss is set to 1.0. The weight decay λ_1 is set to 0.001. The weights of the three FC layers are initialized with Gaussian filter with a standard deviation of 0.005 and constant bias of 0. In the training phase, we use 10% of the training data as a validation set. Three-fold cross-validation is employed to select appropriate hyperparameters. The learning rate of the model is set to $1E-5$. The early stopping function is used to monitor the training process. The training model stops training when it has not improved after five epochs. In the test phase, we randomly split the dataset three times for model evaluation. The following experiments were run using a PC with an Intel I7-4790 3.60GHz CPU, 32GB RAM, and an Nvidia Titan X GPU card.

C. EVALUATED ALGORITHMS

To evaluate the effectiveness of the DS-MLTL algorithm, performance comparisons between the following DL methods are performed: (1) CNN model. CNN is a classic model that has been successfully applied in the computer vision field. The CNN model in this study is comprised of 3 convolutional layers, 2 fully connected layers, and a binary cross-entropy loss. The convolutional neural layers are adopted to extract features for category representation, and the fully connected layers are adopted to learn a nonlinear function for the diagnosis of PF. (2) TL model. Instead of training a model from scratch, the TL model uses the convolution layers of VGG-16 to extract features, and 3 fully connected layers, similar to the DS-MLTL model for classification. The loss of this model is also the binary cross-entropy. (3) DS-ML model without transfer learning. Compared with the CNN model, DS-ML takes two images as input to a two-branch CNN network at the same time. Each branch in DS-ML is the same as the CNN model. The output vectors of the first fully connected layer of the two images are used to measure the Siamese loss. This is a multitask learning model that uses a binary cross-entropy loss for classification and a Siamese loss for similarity learning. Moreover, performance comparisons between the DS-MLTL model and various features such as sono and elastograms are also conducted.

D. LEARNED CLASSIFICATION RESULTS

1) COMPARISON WITH DEEP LEARNING-BASED METHODS

The sensitivity, specificity and accuracy of the compared models are enumerated in Table 1, where the mean value and standard deviation are reported. In this table, n (P) denotes the number and percentage of data sets. We compared the performance of DL-based models with 720 training images and 360 test images. Our DS-MLTL model achieves the best performance in test set with classification sensitivity, specificity and accuracy of $76.85 \pm 7.40\%$, $93.33 \pm 6.19\%$, $85.09 \pm 6.67\%$, respectively. The TL model achieves the best performance in the training set. Figure 3 shows the comparison of AUC curves of four models from a random data split. It can also be find that the TL model achieves best AUC score of 0.97 in the training phase (Figure 3A) and the DS-MLTL model has the best AUC score of 0.90 in the test phase (Figure 3B). This verifies the advantages of transfer learning in the small medical dataset. It also shows the effectiveness of the joint learning of visual features and the multitask loss. In fact, our model has better ablility to avoid overfitting. In all test datasets in Table 1, the augmented images only come from the translation of the original images. To evaluate the performance of these models more comprehensively, we reconstruct a test dataset without data augmentation for model evaluation, which only includes 40 raw images (20 healthy images vs 20 PF images). Table 2 shows the classification results on the test dataset without data augmentation. The DS-MLTL model achieves the best sensitivity, specificity and accuracy of $78.33 \pm 5.77\%$, $91.67 \pm 5.78\%$, and $85.0 \pm 5.0\%$, respectively.

TABLE 1. Diagnostic performance of CNN, TL, DS-ML, DS-MLTL for the evaluation of PF on the training and test sets.

	Training Set				Test Set			
Method	n (P)	Sensitivity (%)	Specificity (%)	Accuracy (%)	n (P)	Sensitivity (%)	Specificity (%)	Accuracy (%)
CNN	720 (67%)	84.72 ± 2.54	81.94 ± 1.73	83.33 ± 2.08	360 (33%)	76.48 ± 9.03	86.48 ± 7.34	81.48 ± 8.17
TL	720 (67%)	94.07 ± 2.64	90.09 ± 2.79	92.08 ± 1.19	360 (33%)	71.11 ± 5.88	88.89 ± 7.35	80.0 ± 5.02
DS-ML	720 (67%)	72.78 ± 3.06	74.72 ± 2.65	73.75 ± 2.53	360 (33%)	76.85 ± 9.0	87.04 ± 8.49	81.94 ± 6.91
DS-MLTL	720 (67%)	88.43 ± 4.88	87.96 ± 5.78	88.19 ± 4.33	**360 (33%)**	**76.85 ± 7.40**	**93.33 ± 6.19**	**85.09 ± 6.67**
DS-MLTL	540 (50%)	90.62 ± 4.57	88.77 ± 3.19	89.69 ± 3.31	540 (50%)	76.91 ± 4.63	89.38 ± 7.46	83.15 ± 5.47
DS-MLTL	900 (83%)	94.30 ± 3.24	88.89 ± 4.44	91.59 ± 3.43	180 (17%)	77.41 ± 6.51	92.96 ± 5.13	85.19 ± 4.79
DS-MLTL	990 (92%)	93.54 ± 3.16	88.82 ± 4.48	91.18 ± 3.72	90 (8%)	77.04 ± 5.13	92.59 ± 5.59	84.81 ± 5.25

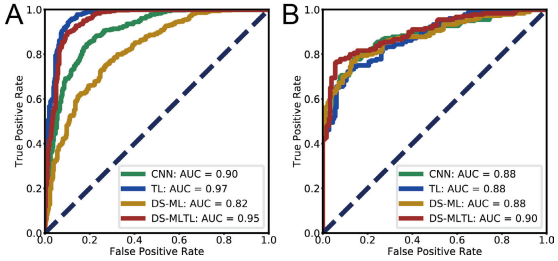

FIGURE 3. Comparison of ROC curves between CNN, TL, DS-ML, DS-MLTL for the assessment of PF. (A) ROC curves for the training set. (B) ROC curves for the test set.

TABLE 2. Comparison results on the test dataset without data augmentation.

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)
CNN	75.0 ± 8.66	86.67 ± 7.64	80.83 ± 8.04
TL	71.67 ± 7.64	88.33 ± 2.89	80.0 ± 4.33
DSML	76.67 ± 7.64	86.67 ± 5.78	81.67 ± 5.20
DS-MLTL	**78.33 ± 5.77**	**91.67 ± 5.78**	**85.0 ± 5.0**

2) PERFORMANCE COMPARISON WITH DIFFERENT QUANTITIES OF TRAINING DATA

To evaluate the DS-MLTL model, the model is also trained on different size datasets (540 training images vs 540 test images, 900 training images vs 180 test images, 990 training images vs 90 test images). The deep Siamese-based method achieves consistently high performance on these datasets with different sizes, which again demonstrates the effectiveness of the DS-MLTL model (Table 1).

3) COMPARISON WITH HUMAN-CRAFTED FEATURES

There are varieties of hand-crafted features extracted from B-mode ultrasound and SWE available to doctors for PF evaluation. (1) Thickness, which is the plantar fascia thickening at its calcaneal insertion; (2) Hypoechogenicity, which indicates echo on the plantar fascia in B-mode ultrasound images; (3) Young's modulus, which defines the relationship between stress and strain in the ROI of the plantar fascia. The mean and the maximum value in the ROI are measured for plantar fascia elasticity assessment; (4) SWE value, which is deduced by shear wave propagation; SWE value also measures the maximum and mean in ROI of plantar fascia for plantar fascia elasticity assessment; (5) Calcification, which is a condition of calcification on the plantar fascia determined by a doctor. (6) Blood flow, in which the doctor determines whether there is blood flow on the plantar fascia. Table 3 shows the performance comparison between human-crafted features and the DS-MLTL model. From this table we can find that our proposed DS-MLTL achieves the best accuracy. For all human-crafted features, the thickness has the best sensitivity of 80.0 ± 5.0%. For elasticity evaluation (both Young's modulus and the SWE value), the mean value is more robust than the maximum value. Figure 4 shows the comparison of AUC curves between our model and human-crafted features from a random data split. The AUC of thickness, SWE value (mean) and hypoechogenicity are 0.81, 0.77 and 0.63, respectively. Our DS-MLTL model achieves favorable performance with AUC in both the training and test dataset. The result suggests the superiority of the DS-MLTL model over other human-crafted features.

TABLE 3. Comparison results between human-crafted features and the DS-MLTL model.

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)
thickness	80.0 ± 5.0	73.33 ± 2.89	76.67 ± 2.89
hypoechogenicity	10.0 ± 5.0	100.0 ± 0.0	55.0 ± 2.5
Young's modulus (max)	36.67 ± 7.64	88.33 ± 2.89	62.50 ± 2.50
Young's modulus (mean)	66.67 ± 7.64	81.67 ± 5.77	74.17 ± 6.29
SWE value (max)	40.0 ± 8.66	88.33 ± 5.77	64.17 ± 3.82
SWE value (mean)	63.33 ± 2.89	85.0 ± 5.0	74.17 ± 3.82
calcification	5.0 ± 5.0	100.0 ± 0.0	52.50 ± 2.50
blood flow	8.33 ± 2.89	100.0 ± 0.0	54.17 ± 1.44
DS-MLTL	**78.33 ± 5.77**	**91.67 ± 5.78**	**85.0 ± 5.0**

E. T-SNE VISUALIZATION OF THE LEARNED FEATURES

DL is a new method for learning representation from data, which emphasizes learning from continuous layers. These layers correspond to increasingly more meaningful representations of specific tasks. To demonstrate the effectiveness of the learned features, t-SNE [28] dimensionality reduction is utilized for the raw 2D-SWE images and the learned features. Figure 5 shows the t-SNE results of a random data split in two-dimensional space. The results show that the features learned by the model can perform PF classification better than the raw images, which proves the effectiveness of the DS-MLTL model again.

IV. DISCUSSION

Ultrasound is a common diagnostic method for PF. However, only experienced doctors can obtain relatively accurate evaluation via ultrasound images, resulting in low efficiency and

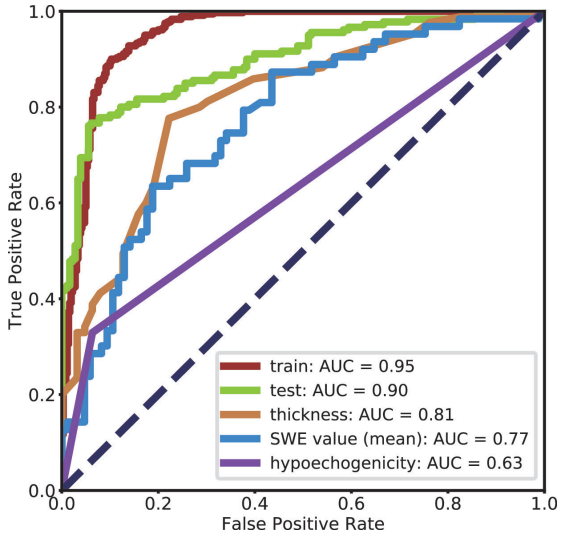

FIGURE 4. Comparison of ROC curves between DS-MLTL and sonographic characteristics.

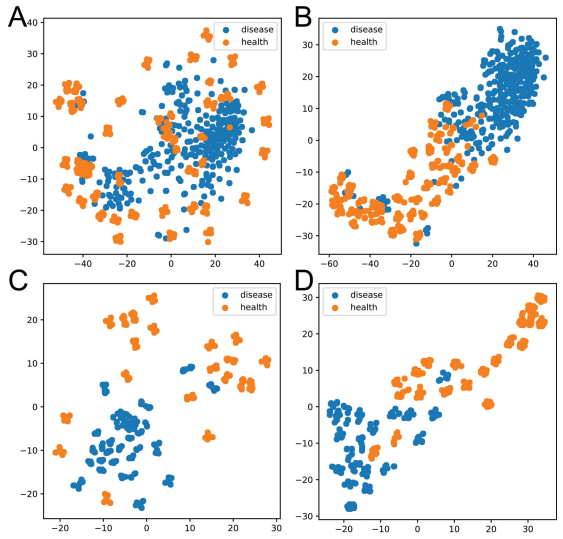

FIGURE 5. T-SNE visualization. (A, B) t-SNE for training data. (C, D) t-SNE for test data. (A, C) t-SNE for the raw images. (B, D) t-SNE for the FC2 output of the DMML-TL model.

high cost. Therefore, designing CAD algorithms to recognize the pattern of these ultrasound images is urgently required. In this study, various hand-crafted features extracted from B-mode ultrasound and SWE for PF are evaluated by statistical analysis of collected clinical cases. To realize an effective CAD algorithm, a deep Siamese-based PF classification model is proposed. This is the first prospective study that uses a DL model for intelligent diagnosis of PF based on 2D-SWE images. It is of great significance for improving diagnostic accuracy and reducing the burden of doctors. Compared with traditional deep learning methods (e.g., CNNs), the deep Siamese framework via multitask learning and transfer learning can further improve discriminative ability.

Doctors evaluate PF by synthesizing human-crafted features extracted from ultrasound images [2]. The features collected in this study include plantar fascia thickening at its calcaneal insertion, hypoechogenicity, calcification, blood flow, Young's modulus and SWE value. Here, the statistical analysis results show that the thickness of plantar fascia has a favorable performance with both sensitivity and accuracy. Young's modulus and SWE value are two metrics for elasticity assessment. Young's modulus defines the relationship between stress and strain. The SWE value is deduced by shear wave propagation. These two metrics have favorable specificity. In addition, we also find that the mean of both Young's modulus and SWE value in the ROI is better than the maximum for PF diagnosis.

2D-SWE is a new noninvasive stiffness measurement technology with great advantage. It has been applied to evaluate many diseases, such as cancer and diseases of the musculoskeletal system. In recent years, the deep learning method has made great progress in 2D-SWE image classification tasks in many studies [24], [25]. All of these DL models have achieved good performance. Undoubtedly, a deep learning model can improve the accuracy of PF diagnosis based on 2D-SWE images. In this work, four DL models are conducted for PF classification based on 2D-SWE images. Compared with the method based on SWE values, the DL strategy has the following advantages: (1) SWE values measure only a small area of plantar fascia stiffness, while the DL method considers the entire 2D-SWE image. More effective information can be obtained through a larger area of the ROI. (2) DL models can automatically extract image features in an "end-to-end" way, and more effective high-dimensional features can be obtained. (3) As convolutional neural networks have an apparent advantage and wide application in computer vision, it is a promising method for promoting the intelligent diagnosis of medical images.

Although traditional deep learning methods can perform ultrasound image classification tasks, many studies have improved the traditional model to achieve better performance. For example, Shi et al. proposed a novel stacked deep polynomial network for tumor classification that achieves favorable performance with a small ultrasound image dataset [23]. To improve our model performance, we proposed a deep Siamese-based network for PF classification.

In the DS-MLTL model, the classification loss and the Siamese loss are learned jointly in a unified framework to benefit each other. Here, the Siamese loss can constrain the distance of the input image pair, which makes similar image pairs close to each other while dissimilar image pairs are separated by a predefined margin. For these two related tasks, the classification loss can learn useful class-specific information, which facilitates the similarity learning process. The Siamese loss can guide more robust feature learning, which is crucial for improving the performance of classification. In our framework, the multitask strategy has better performance than the traditional classification model, which makes a single diagnostic result output. Table 1 and Table 2

show that the DS-MLTL model achieves the best performance compared with the CNN model and TL model in the test set, although the TL model achieves the best performance in the training set, which suggests that learning the classification loss and the Siamese loss jointly in a unified framework is robust and avoids model overfitting.

Moreover, transfer learning is widely used in medical image processing because medical image datasets are usually small [29]. Here, the DL model based on TL can effectively improve the performance of feature extraction and classification. It can be observed that the DS-MLTL model achieves better performance than the DS-ML model (Table 1, Table 2). Because there are large-scale weights to be fitted in the DL model, deep learning is highly dependent on the quantity and quality of data. We believe that the performance of the deep Siamese-based model will be further improved as more diverse and high-quality 2D-SWE images of plantar fascia (healthy and PF) are collected in the future.

V. CONCLUSION

This paper presents a deep Siamese-based architecture for plantar fascia classification based on 2D-SWE images. This architecture can effectively improve the discriminative ability of the model by adding comparisons of image pairs. The TL method avoids overfitting the training phase in the case of a small dataset. Compared with various features of sono and elastograms (thickness, hypoechogenicity, elasticity value, etc.), the DS-MLTL model achieves the highest accuracy of $85.09 \pm 6.67\%$. The model is also superior to the traditional CNN model. It is valuable and practical for the accurate diagnosis of PF. Motivated by the favorable performance of our framework, we intend to apply this approach to other medical image analysis tasks, such as breast cancer and thyroid nodules.

ACKNOWLEDGMENT

(Junling Gao and Lei Xu contributed equally to this work.)

REFERENCES

[1] P. Beeson, "Plantar fasciopathy: Revisiting the risk factors," *Foot Ankle Surg.*, vol. 20, no. 3, pp. 160–165, Sep. 2014.

[2] E. G. McNally and S. Shetty, "Plantar fascia: Imaging diagnosis and guided treatment," in *Seminars in Musculoskeletal Radiology*, vol. 14. Stuttgart, Germany: Thieme Medical Publishers, 2010, no. 3, pp. 334–343.

[3] M. Akfirat, C. Sen, and T. Günes, "Ultrasonographic appearance of the plantar fasciitis," *Clin. Imag.*, vol. 27, no. 5, pp. 353–357, Sep./Oct. 2003.

[4] S. C. Wearing, J. E. Smeathers, S. R. Urry, E. M. Hennig, and A. P. Hills, "The pathomechanics of plantar fasciitis," *Sports Med.*, vol. 36, no. 7, pp. 585–611, 2006.

[5] L. Zhang, W. Wan, L. Zhang, H. Xiao, Y. Luo, X. Fei, Z. Zheng, and P. Tang, "Assessment of plantar fasciitis using shear wave elastography," *Nan Fang Yi Ke Da Xue Xue Bao= J. Southern Med. Univ.*, vol. 34, no. 2, pp. 206–209, Feb. 2014.

[6] M.-Y. Hsiao, Y.-C. Chen, C.-Y. Lin, W.-S. Chen, and T.-G. Wang, "Reduced patellar tendon elasticity with aging: *In Vivo* assessment by shear wave elastography," *Ultrasound Med. Biol.*, vol. 41, no. 11, pp. 2899–2905, Nov. 2015.

[7] W. A. Berg, D. O. Cosgrove, and C. J. Doré, F. K. Schäfer, W. E. Svensson, R. J. Hooley, R. Ohlinger, E. B. Mendelson, C. Balu-Maestro, M. Locatelli, C. Tourasse, B. C. Cavanaugh, V. Juhan, A. T. Stavros, A. Tardivon, J. Gay, J.-P. Henry, and C. Cohen-Bacrie, "Shear-wave elastography improves the specificity of breast us: The be1 multinational study of 939 masses," *Radiology*, vol. 262, no. 2, pp. 435–449, Feb. 2012.

[8] R. G. Barr, R. Memo, and C. R. Schaub, "Shear wave ultrasound elastography of the prostate: Initial results," *Ultrasound Quart.*, vol. 28, no. 1, pp. 13–20, Mar. 2012.

[9] J.-B. Veyrieres, F. Albarel, J. V. Lombard, J. Berbis, F. Sebag, C. Oliver, and P. Petit, "A threshold value in Shear Wave elastography to rule out malignant thyroid nodules: A reality," *Eur. J. Radiol.*, vol. 81, no. 12, pp. 3965–3972, Dec. 2012.

[10] F. Sebag, J. Vaillant-Lombard, J. Berbis, V. Griset, J. F. Henry, P. Petit, and C. Oliver, "Shear wave elastography: A new ultrasound imaging mode for the differential diagnosis of benign and malignant thyroid nodules," *J. Clin. Endocrinol. Metabolism*, vol. 95, no. 12, pp. 5281–5288, 2010.

[11] A. Y. Park, E. J. Son, K. Han, J. H. Youk, J.-A. Kim, and C. S. Park, "Shear wave elastography of thyroid nodules for the prediction of malignancy in a large scale study," *Eur. J. Radiol.*, vol. 84, no. 3, pp. 407–412, Mar. 2015.

[12] E. E. Drakonaki, G. M. Allen, and D. J. Wilson, "Ultrasound elastography for musculoskeletal applications," *Brit. J. Radiol.*, vol. 85, no. 1019, pp. 1435–1445, Nov. 2012.

[13] C.-Y. Lin, C.-C. Lin, Y.-C. Chou, P.-Y. Chen, and C.-L. Wang, "Heel pad stiffness in plantar heel pain by shear wave elastography," *Ultrasound Med. Biol.*, vol. 41, no. 11, pp. 2890–2898, Nov. 2015.

[14] R. Gillies, P. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.

[15] L. E. Court, X. Fave, D. Mackin, J. Lee, J. Yang, and L. Zhang, "Computational resources for radiomics," *Transl. Cancer Res.*, vol. 5, no. 4, pp. 340–348, Aug. 2016.

[16] R. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. P. M. van Stiphout, P. Granton, C. M. L. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. W. L. Aerts, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, Mar. 2012.

[17] I. El Naqa, P. W. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M. Schmitt, R. Laforest, W. L. Thorstad, and J. O. Deasy, "Exploring feature-based approaches in PET images for predicting cancer treatment outcomes," *Pattern Recognit.*, vol. 42, no. 6, pp. 1162–1171, Jun. 2009.

[18] R. M. Haralick and K. Shanmugam, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[19] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, "Texture analysis of medical images," *Clin. Radiol.*, vol. 59, no. 12, pp. 1061–1069, 2004.

[20] D. Zinovev, D. Raicu, J. Furst, and S. Armato III, "Predicting radiological panel opinions using a panel of machine learning classifiers," *Algorithms*, vol. 2, no. 4, pp. 1473–1502, Nov. 2009.

[21] F. Song, Z. Guo, and D. Mei, "Feature selection using principal component analysis," in *Proc. Int. Conf. Syst. Sci., Eng. Design Manuf. Inform. (ICSEM)*, vol. 1. Nov. 2010, pp. 27–30.

[22] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[23] J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, and T. Wang, "Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset," *Neurocomputing*, vol. 194, pp. 87–94, Jun. 2016.

[24] K. Wang, X. Lu, H. Zhou, Y. Gao, J. Zheng, M. Tong, C. Wu, C. Liu, L. Huang, F. Meng, Y. Lu, H. Ai, X. Y. Xie, L. P. Yin, P. Liang, J. Tian, and R. Zheng, "Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis b: A prospective multicentre study," *Gut*, vol. 68, no. 4, pp. 729–741, Apr. 2019.

[25] Q. Zhang, Y. Xiao, W. Dai, J. Suo, C. Wang, J. Shi, and H. Zheng, "Deep learning based classification of breast tumors with shear-wave elastography," *Ultrasonics*, vol. 72, pp. 150–157, Dec. 2016.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[28] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[29] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

JUNLING GAO received the M.S. degree from Southwest University, Chongqing, China, in 2017. She is currently pursuing the Ph.D. degree with the Key Laboratory of Biomedical Information Engineering of Ministry of Education, Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China. Her research interests include ultrasound imaging, image processing, pattern recognition, and deep learning.

LEI XU received the bachelor's degree in medicine and the master's degree in imaging and nuclear medicine from Forth Military Medical University. She is currently pursuing the Ph.D. degree with the Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University. She is also a Radiologist with the Department of Medical Ultrasonic, Xi'an Hospital of Traditional Chinese Medicine. Her research interests include musculoskeletal ultrasound diagnosis, ultrasound elastography, and ultrasonic intervention treatment.

AYACHE BOUAKAZ received the M.S. degree in acoustics and the Ph.D. degree from the Department of Electrical Engineering, Institut National des Sciences Appliquées de Lyon (INSA Lyon), France, in 1992 and 1996, respectively. In 1998, he joined the Department of Bioengineering, Pennsylvania State University, State College, PA, USA, as a Postdoctoral, for two years. From December 1999 to November 2004, he was employed as an Associate Professor with the Erasmus University Medical Center, Rotterdam, The Netherlands. Since 2009, he has been holds a permanent position as the Director of research, the Head of the Ultrasound and Imaging Laboratory, and the Deputy Director of the iBrain Institute. His research interests include imaging, ultrasound contrast agents, transducer design, and imaging and therapeutic applications of ultrasound. He was the General Chair of the 2016 International Conference IEEE IUS. He is the Vice President of the IEEE UFFC in Charge of Symposia.

MINGXI WAN (M'01) was born in Hubei, China, in 1962. He received the B.S. degree in geophysical prospecting from the Jianghan Petroleum Institute, Jingzhou, China, in 1982, and the M.S. and Ph.D. degrees in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 1985 and 1989, respectively, where he was the Chairman of the Department of Biomedical Engineering and the Dean of the School of Life Science and Technology, from 1994 to 2006 and from 2000 to 2010, respectively. From 1995 to 1996, he was a Visiting Professor with Drexel University, Philadelphia, PA, USA, and an Adjunct Professor with Pennsylvania State University, University Park, PA, USA. From 2001 to 2002, he was a Visiting Scholar with the Department of Biomedical Engineering, University of California at Davis, Davis, CA, USA. He has been a Professor and the Director of the Institute of Biomedical Engineering and Instrumentation, Xi'an Jiaotong University. He has authored or coauthored more than 120 peer-reviewed publications in international journals and six books about medical ultrasound. His current research interests include ultrasonic imaging, especially in contrast and tissue perfusion evaluation, tissue elasticity imaging, therapeutic ultrasound, theranostics, and voice science. He was a recipient of several important awards from the Chinese Government and university.