Classification of Aortic Stenosis Before and After Transcatheter Aortic Valve Replacement Using Cardio-mechanical Modalities*

Chenxi Yang, *Student Member, IEEE*, Banish Ojha, Nicole D. Aranoff, Philip Green, *M.D.*, and Negar Tavassolian, *Senior Member, IEEE*

***Abstract*—This paper reports our study on the impact of transcatheter aortic valve replacement (TAVR) on the classification of aortic stenosis (AS) patients using cardio-mechanical modalities. Machine learning algorithms such as decision tree, random forest, and neural network were applied to conduct two tasks. Firstly, the pre- and post-TAVR data are evaluated with the classifiers trained in the literature. Secondly, new classifiers are trained to classify between pre- and post-TAVR data. Using analysis of variance, the features that are significantly different between pre- and post-TAVR patients are selected and compared to the features used in the pre-trained classifiers. The results suggest that pre-TAVR subjects could be classified as AS patients but post-TAVR could not be classified as healthy subjects. The features which differentiate pre- and post-TAVR patients reveal different distributions compared to the features that classify AS patients and healthy subjects. These results could guide future work in the classification of AS as well as the evaluation of the recovery status of patients after TAVR treatment.**

I. Introduction

Cardiovascular disease (CVD) is the dominant global cause of death, accounting for more than 17.9 million deaths per year [1]. Among the different categories of heart disease, valvular heart disease (VHD) accounts for approximately 20% of all cardiac surgical procedures in the United States [1]. Aortic stenosis (AS) is the most prevalent of all VHDs, and is defined as the narrowing of the aortic valve which prevents an adequate outflow of blood [2], [3]. Conventional AS detection techniques such as echocardiography and magnetic resonance imaging are often constraining, expensive, and limited to use in the clinic [3]. Recently, wearable sensors using non-invasive modalities have been investigated as out-of-clinic ubiquitous monitors for the detection of CVDs. Specifically, cardio-mechanical modalities, i.e. seismo- and gyro-cardiography, have been evaluated for monitoring various kinds of cardiovascular diseases using inertial sensors attached to the chest walls of subjects [4]-[8]. Researchers have verified the feasibility of detecting atrial fibrillation [4], [6], heart arrhythmia [5], and heart failure [7]. Our group has previously studied the capability of detecting AS based on time-frequency features extracted from seismo- and gyro-cardiograms [8]. We collected seismo-cardiogram (SCG) signals using an accelerometer, and gyro-cardiograms (GCG) using a gyroscope. These inertial sensors detect heart-induced vibrations that are related to valve activities [8]. Our results indicate that we can effectively differentiate between AS patients before treatment and healthy subjects based on features from cardio-mechanical signals using conventional machine learning algorithms [8].

Transcatheter aortic valve replacement (TAVR) is the replacement of the aortic valve of the heart through the blood vessels without an open heart surgery, and is a major method of treatment for AS [9]-[12]. It is a feasible and less risky option for frail and elderly patients who need valve replacement or implementation compared to other treatments [11], [12]. However, valve replacement does not provide a definitive cure to the patient and might result in prosthesis-related complications such as prosthesis-patient mismatch and thromboembolic complications [11]. Many of the complications can be prevented through careful medical management and long-term follow-up after implantation [12]. A non-invasive home-based monitor that provides an evaluation of the wellness of aortic valves could hugely benefit the long-term healthcare service for AS patients after TAVR.

There are two concerns in the evaluation of pre- and post-TAVR subjects with the framework developed in [8]. Firstly, our framework is based on the characteristics of vibration signals produced by heart valves. It is unknown if artificial valves provide the same mechanical characteristics compared to natural valves. It has been shown that artificial valves may produce different sounds in the hearing frequency range compared to natural valves [10]. Since sound is also induced by vibrations, this observation raises our concern regarding

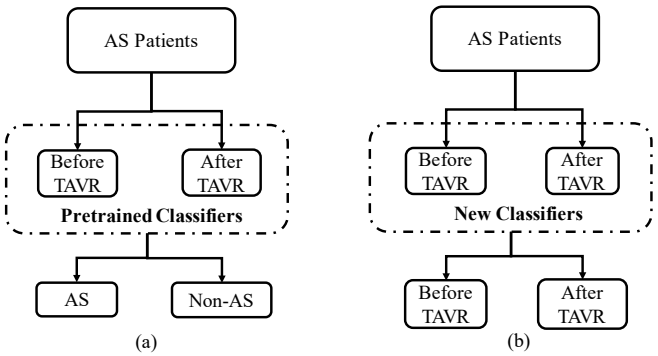

Fig. 1. The overall structure of the classification tasks studied in this work. (a) Testing with pre-trained classifiers for the classification of AS patients and healthy subjects. (b) Developing new classifiers for the classification of pre- and post-TAVR subjects.

* Research supported by National Science Foundation (NSF) under award number 1855394.

Chenxi Yang and Banish Ojha are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, 07030, USA (e-mail: negar.tavassolian@stevens.edu).

Nicole D. Aranoff is at the Yeshiva University, New York, NY 10032 USA.

Philip Green is with Columbia University Medical Center, New York, NY 10032 USA.

Negar Tavassolian is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology.

low-frequency vibrations induced by artificial valves as well. Secondly, from the perspective of the classifier, it is not clear whether post-TAVR patients could be classified in the same category as healthy subjects. In other words, patients with artificial valves might need to be categorized as a separate class due to the distinct characteristics of their cardio-mechanical signals.

In this work, for the first time, we evaluate cardio-mechanical recordings from AS patients before and after TAVR treatment. Fig. 1 illustrates the two classification tasks studied in this paper. As presented in Fig. 1 (a), we first evaluate the classification results of pre- and post-TAVR using the pre-trained classifiers developed in our previous work [8]. Next, using the same framework, we train new classifiers to classify between pre- and post-TAVR data as shown in Fig. 1 (b). Finally, we compare the feature sets used in Fig. 1 (a) and Fig. 1 (b).

The structure of the paper is as follows. Section II introduces the experimental materials and protocol. The hardware and software methods are explained in Section III, followed by the results in Section IV. Section V concludes the study and discusses future work.

II. Experimental Materials and Protocol

A. Datasets

1) Dataset of AS patients with TAVR treatments

Ten inpatient subjects from the Structural Heart & Valve Center of Columbia University Medical Center (CUMC) participated in the collection of recordings. All the patients were first measured before receiving the TAVR treatment. They were measured again on the same day after the procedure. The AS cohort includes four male and six female subjects whose ages, weights, and heights varied as 63-95 years old, 57-88 kg, and 155-180 cm, respectively.

2) Dataset of AS patients and healthy subjects

Data from 40 subjects were used in the training of the pre-trained classifier. Twenty AS subjects from the cardiac care units of Columbia University Medical Center (CUMC) participated in the data collection, all of whom were measured prior to receiving any treatments. A control group of twenty healthy subjects also participated in this study. Measurements from normal subjects were performed at Stevens Institute of Technology. More details of these two demographic groups can be found in our previous work [8].

B. Experimental Protocol

The subjects were asked to sit at rest on a bed for five minutes during both pre- and post-TAVR measurements. The subjects breathed naturally during the experiments. The Institutional Review Board of CUMC approved the patient experimental protocol under protocol number AAAR4104.

III. Methods

A. The Hardware System

We used the same hardware system as in [8] to ensure the consistency of data collection. A commercial wearable sensor node (Shimmer 3 from Shimmer Sensing) is attached to the center of the sternum along the third rib using a chest strap. A three-axis accelerometer measures the seismo-cardiogram (SCG) signal, and a three-axis gyroscope records the gyro-cardiogram (GCG) signal. Both sensors share the same axis definition, where the *z*-axis refers to the dorso-ventral direction of the body, the *y*-axis is along the head-to-foot direction, and the *x*-axis is along the shoulder-to-shoulder direction. Reference heartbeat measurements are taken by a standard four-lead electrocardiogram (ECG) system, which is wire-connected to the sensor. All the recordings are sampled at a sampling rate of 256 Hz. Data are stored in an SD card during measurements and then imported into MATLAB (R2018) for further processing.

B. The Software Methods

Fig. 2 illustrates the block diagram of the proposed signal processing framework, which is modified from our previous study in [8]. The signal processing framework consists of three main parts. The first part is feature generation based on SCG and GCG signals, as shown in Fig. 2 (a). The features are then used for the first and second classification tasks introduced in Section I, as presented in Fig. 2 (b) and Fig. 2 (c) respectively. The details are as below.

1) Feature generation

The feature generation is consistent with our previous study to ensure a fair comparison of results. A simple introduction of this step is presented in this paper. More details can be found in [8].

First, the raw signals are pre-processed. We use the *z*-axis of the SCG signals and the *y*-axis of the GCG signals for feature generation. Two zero-phase infinite impulse response (IIR) bandpass filters are used to prefilter the SCG, GCG, and ECG signals with a passing band of 0.8-25 Hz for SCG/GCG, and a passing band of 0.8-30 Hz for ECG signals. The filtered signals are divided into equal-length segments using a threshold-based exclusion method with a root-mean-square (RMS) filter modified from [4]. The method is first applied to the SCG waveform to generate timestamps for the segmentation of data with a length of 10 seconds. The timestamps are then applied to all SCG, GCG, and ECG waveforms to ensure the alignment of recordings. All equal-length SCG and GCG segments are then further divided into single-cardiac-cycle segments based on the R-R intervals from the corresponding ECG waveforms.

The equal-length 10-second segments of SCG and GCG are processed using continuous wavelet transform (CWT), with a wavelet selection of Morse wavelet. We focus on 55 frequency bins within a frequency range of 0.79-25.39 Hz and

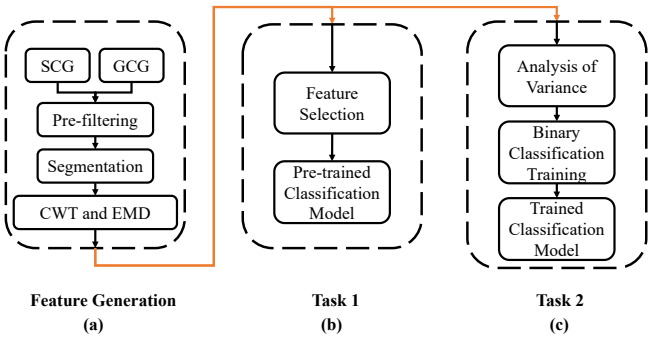

Fig. 2. Diagram of the signal processing framework. (a) Pre-processing and feature generation. (b) Feature selection based on previous results and test on the pre-trained classifiers. (c)Perform new feature analysis and conduct new training/validation tasks.

extract four statistical features from each frequency bin. The statistical features are the maximum power, the mean, the standard deviation, and the median values.

In addition, we apply empirical mode decomposition (EMD) to the single-cardiac-cycle SCG/GCG segments and generate intrinsic mode functions (IMFs). Four IMFs are generated from each segment and the segments from the same 10-second waveform are ensemble-averaged. Then the mean, the standard deviation, the skewness, and the entropy of the averaged IMFs are extracted as features. Consequently, 236 features are extracted from each of the SCG and GCG signals respectively.

2) Feature selection and pre-trained classification

The pre-trained classifiers are based on a selection of features from all the features generated in Step 1). Table I summarizes the feature selection results of the pre-trained classification task. In summary, 24 features from SCG and 30 features from GCG are selected for the classification task. In our previous study, it was shown that the combination of SCG and GCG features produces the best results [8]. Therefore, the new data is evaluated using the classifier trained by combined SCG and GCG features. More details of the feature selection could be found in [8]. The types and configurations of the classifiers are introduced in the following Section 3.

3) Feature analysis and new training task.

We conduct a new feature analysis with a one-way analysis of variance (ANOVA) test for the classification of pre- and post-TAVR measurements. The ANOVA test results are used to train classifiers via decision tree (DT), random forest (RF), and neural network (NN) algorithms. The types and configurations of the algorithms are the same as in [8]. The maximum split for DT is 20, and the maximum depth is 7. The number of trees in RF is 30 with a maximum number of 798 splits. The NN has 30 input, 20 hidden and 10 output neurons with an initial learning rate of 0.001.

4) Evaluation and Validation

The standard sensitivity (SE), specificity (SP), and accuracy (ACC) metrics are evaluated in this study. We implemented ten-fold cross-validation, as also implemented in [8]. A total of 610 segments are extracted from 20 measurements. To balance the training sets from pre- and post-TAVR measurements, 480 segments were selected, with 240 from pre- and 240 from post-TAVR measurements.

TABLE I.
FEATURE SELECTION FOR THE PRE-TRAINED CLASSIFIERS

Source	Feature Type	Frequency Bin (Hz) or IMF Number			
SCG					
CWT	Maximum	1.07	1.15	1.23	1.32
		1.42	1.52	2.83	3.04
		3.26	3.49	6.08	6.51
		9.21			
	Mean	2.30	4.30	4.60	
	Standard Deviation	1.00	1.07	1.15	
		1.23			
	Median	1.32	2.14		
EMD	Mean	IMF1			
	Skewness	IMF3			
Total	24				
GCG					
CWT	Maximum	3.04	3.26	3.49	3.74
		4.01	4.01	4.3	4.6
		4.94	5.29	5.67	6.51
		7.48	8.02	8.59	9.21
		9.87	10.58		
	Mean	0.87	0.94	1.00	
		1.07			
	Standard Deviation	4.30	4.60	4.94	
	Median	4.00	4.29		
EMD	Mean	IMF1	IMF3		
	Skewness	IMF3			
Total	30				

IV. EXPERIMENTAL RESULTS

A. Classification Results Using Pre-Trained Classifiers

The pre-TAVR measurements are from a similar cohort to the AS cohort in [8]. Therefore, we expect the pre-TAVR cohort in the new data to be classified as the true-negative of the pre-trained classifiers, i.e., the AS cohort of the previous research. Subsequently, we correlate the post-TAVR cohort to the label of healthy subjects. As a result, the specificity metrics represent the percentages of pre-TAVR observations that are classified as AS. The sensitivity values show the percentages of post-TAVR observations that have been estimated to be healthy.

Fig. 3 summarizes the average classification results from the ten-fold cross-validations via the three types of pre-trained classifiers. As shown in Fig. 3, the sensitivity metrics report low values, which are 0.33 from DT, 0.45 from RF, and 0.46 from NN. The standard deviations of the results are 0.22 from DT, 0.21 from RF, and 0.15 from NN, suggesting a significant variation in the classification performance. These values suggest that the post-TAVR cohort could not be classified either as AS patients or healthy subjects by the pre-trained classifiers.

On the other hand, the specificity values report 0.92 from DT, 0.95 from RF, and 0.89 from NN, showing a high agreement between pre-TAVR patients and the previously collected AS patients. The standard deviations are 0.04, 0.02, and 0.05 from DT, RF, and NN respectively. The results show that the pre-TAVR cohort could be effectively classified as AS patients. The best result from our previous work in [8] is 0.98 from the RF classifier, which is better than the best result of 0.95 from the RF classifier in this test. The performance difference might be caused by the demographic difference between the two cohorts since not all AS patients have conditions to be treated by TAVR. The average accuracy values are 0.62 from DT, 0.70 from RF, and 0.68 from NN, suggesting that the pre/post-TAVR classification should not be generally considered the same as AS/Healthy classification.

B. Feature Analysis and Comparison

Table II summarizes the new feature analysis and classification results based on the pre- and post-TAVR measurements. There are 23 significant features from SCG and 33 features from GCG, most of which are from the maximum features of CWT. Compared to the previous feature set in Table I, there are fewer SCG-based features and more GCG-based features. There are five features in common from SCG, which are the mean of IMF1 and the maximum features at 1.07, 1.15, 1.23, and 3.04 Hz. In comparison, nine features are found in common from GCG, which are the mean of IMF1,

TABLE II.
FEATURE ANALYSIS OF THE NEW DATA

Source	Feature Type	Frequency Bin (Hz) or IMF Number			
SCG					
CWT	Maximum	0.76	0.81	0.87	1.00
		1.07*	1.15	1.23	1.74
		2.64	3.04	4.94	5.29
		7.48	8.59	9.87	10.58
		14.96	19.74		
	Mean	10.58			
	Median	9.87	10.58		11.34
EMD	Mean	IMF1			
Total	23				
GCG					
CWT	Maximum	1.00	1.07	1.15	1.23
		1.42	1.52	1.74	1.87
		2.00	2.15	2.47	2.64
		2.83	3.04	3.26	3.49
		3.74	4.01	4.94	6.08
		6.51	6.98	7.48	13.96
		14.96	19.74	22.68	
	Mean	8.02	10.58		
	Median	7.48	8.02		10.58
EMD	Mean	IMF1			
Total	33				

Classification Results			
	DT	**RF**	**NN**
SE	0.990	0.995	0.988
SP	0.991	0.993	0.990
AC	0.991	0.994	0.989

*Values with orange shading are overlapping with Table I. *SE: sensitivity, SP: specificity, AC: accuracy, DT: decision tree, RF: random forest, NN: neural network.*

and the maximum features at 3.04, 3.26, 3.49, 3.74, 4.01, 4.94, 6.51, and 7.48 Hz. The features that overlap with the previous feature set are highlighted in orange shades. It is observed that there most features are different between the new and previous feature sets.

The classification results summarized in Table II report the best sensitivity, specificity, and accuracy from the RF classifier as 0.995, 0.993, and 0.994 respectively. These results are comparable with the classification results of 0.996 in SE, 0.995 in SP, and 0.995 in AC in [6]. Although these results are biased due to the small data size, they represent the effectiveness of the new feature set for the classification of pre- and post-TAVR observations.

In summary, the features that differentiate pre- and post-TAVR patients do not have significant overlapping with the features that differentiate AS and healthy subjects. In other words, AS patients after TAVR treatment can not be simply

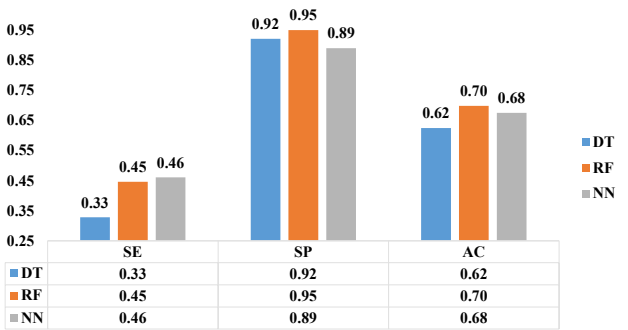

Fig. 3. Classification results from the pre-trained classifier. SE: sensitivity, SP: specificity, AC: accuracy, DT: decision tree, RF: random forest, NN: neural network.

classified as healthy even though their valves are replaced and considered as functional. Two hypotheses can explain this difference. The first is that the artificial valve structure produces different heart-induced vibrations compared to a natural valve. Secondly, the recovery status of the subjects after TAVR may influence the behavior of the valves. The post-TAVR patients may have weak hearts because they were measured the same day after the surgery.

V. CONCLUSION

This paper presents our evaluation of cardio-mechanical signals collected from AS patients before and after TAVR treatments. Our analyses can be summarized as three major conclusions. Firstly, pre-TAVR measurements can be classified as AS, although post-TAVR measurements can not be classified as healthy. Secondly, the features from cardio-mechanical signals are significantly different between pre- and post-TAVR measurements. Lastly, the significant features between pre- and post-TAVR measurements are mostly different from those between AS patients and healthy subjects.

Future work in this area includes the verification of our preliminary observations with a larger demographic group. The results of this work reveal potential in several applications. For instance, subjects with artificial valves can be labeled as a separate class from subjects with natural valves so that the algorithms used for AS analysis can be more accurate and robust. The feature differences could also be used to indicate the recovery status of AS patients after TAVR treatments. The existence of prosthetic structures or devices should also be noticed as a factor that might influence the training or classification tasks related to other types of CVDs.

REFERENCES

[1] WRITING GROUP MEMBERS, "Heart Disease and Stroke Statistics—2018 At-a-glance: A Report from the American Heart Association." *Circulation, 2018*

[2] R Klabunde, Cardiovascular Physiology Concepts. Lippincott Williams & Wilkins, 2011.

[3] Carabello, A. P. Blasé, J. Walter, "Aortic Stenosis," *Lancet* vol. 373, pp 956 – 966.,2009

[4] O. Lahdenoja, *et al.*, "A Smartphone-Only Solution for Detecting Indications of Acute Myocardial Infarction," *IEEE Int. Conf. on Bio. & Health Info. (BHI)*, pp. 197-200, 2017.

[5] Z. Iftikhar, *et al.* "Multiclass Classifier based Cardiovascular Condition Detection Using Smartphone Mechanocardiography," *Scientific Reports*, vol. 8, no.1, pp. 9344, 2018.

[6] T. Hurnanen, *et al.*, "Automated Detection of Atrial Fibrillation Based on Time–Frequency Analysis of Seismocardiograms," *IEEE J. of Biom. and Health Info.*, vol. 21, no. 5, pp. 1233-1241, Sept. 2017.

[7] O. T. Inan, et al., "Novel Wearable Seismocardiography and Machine Learning Algorithms Can Assess Clinical Status of Heart Failure Patients," *Circulation: Heart Failure*, vol. 11, no. 1, pp. e004313, 2016.

[8] C. Yang, *et al.*, "Classification of Aortic Stenosis Using Time-Frequency Features from Chest Cardio-mechanical Signals," *IEEE Trans. on Biomed. Eng.*.

[9] P. C. Cremer, Prosthetic heart valves. Manual of Cardiovascular Medicine, 2018.

[10] T. A. Pedersen, *et al.*, "Are Sounds from Mechanical Heart Valves Equal for Different Valve Types?" *J. Heart Valve Dis.*, vol.17, no.5, pp.579-82, 2008.

[11] S. Arora, "Transcatheter Aortic Valve Replacement: Comprehensive Review and Present Status," *Texas Heart Inst. J.*, vol.44, no.1, pp.29-38, 2017.

[12] P. Pibarot, and J.G. Dumesnil. "Prosthetic Heart Valves: Selection of the Optimal Prosthesis and Long-term Management," *Circulation*, vol. 119, no. 7, pp.1034-1048, 2009.