

# Classification of Aortic Stenosis Using Time-Frequency Features from Chest Cardio-mechanical Signals

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

**Abstract— Objectives:** This paper introduces a novel method for the detection and classification of aortic stenosis (AS) using the time-frequency features of chest cardio-mechanical signals collected from wearable sensors, namely seismo-cardiogram (SCG) and gyro-cardiogram (GCG) signals. Such a method could potentially monitor high-risk patients out of the clinic. **Methods:** Experimental measurements were collected from twenty patients with AS and twenty healthy subjects. Firstly, a digital signal processing framework is proposed to extract time-frequency features. The features are then selected via the analysis of variance test. Different combinations of features are evaluated using the decision tree, random forest, and artificial neural network methods. Two classification tasks are conducted. The first task is a binary classification between normal subjects and AS patients. The second task is a multi-class classification of AS patients with co-existing valvular heart diseases. **Results:** In the binary classification task, the average accuracies achieved are 96.25% from decision tree, 97.43% from random forest, and 95.56% from neural network. The best performance is from combined SCG and GCG features with random forest classifier. In the multi-class classification, the best performance is 92.99% using the random forest classifier and SCG features. **Conclusion:** The results suggest that the solution could be a feasible method for classifying aortic stenosis, both in the binary and multi-class tasks. It also indicates that most of the important time-frequency features are below 11 Hz. **Significance:** The proposed method shows great potential to provide continuous monitoring of valvular heart diseases to prevent patients from sudden critical cardiac situations.

**Index Terms—** Aortic stenosis, gyro-cardiography (GCG), machine-learning, MEMS accelerometer, MEMS gyroscope, seismo-cardiography (SCG), signal processing, time-frequency analysis.

## I. INTRODUCTION

Valvular heart diseases (VHDs) are abnormal conditions of the heart caused by damages of the heart valves. VHDs affect a huge population and have high mortality rates compared to other cardiovascular diseases (CVDs) [1]. There

are four heart valves in the heart: the aortic valve, the mitral valve, the pulmonary valve, and the tricuspid valve [2]. These valves can be influenced by two types of mechanical incidents: stenosis and insufficiency (also known as regurgitation). Stenosis is the narrowing of the valvular orifice that prevents an adequate outflow of blood, and insufficiency describes the inability of the valve to prevent the backflow of blood [3]. Hence, there are majorly eight types of VHDs from the combination of four valves and two incidents. Among all valvular heart diseases, aortic stenosis (AS) has the highest prevalence [3]. AS affects 2%-5% of the population, especially in the senior citizens [3]. According to the Euro Heart Survey on VHDs, 43.1% of VHD incidents are AS [4]. Although AS is a fairly common disease, between one-third and two-thirds of the AS patients go untreated [3]. One main reason is that treatment is fairly conservative before the detection of symptoms [3]. It is therefore critical to detect AS so that proper treatment could be performed. There are several tools for the detection of AS, such as echocardiography, computed tomography, magnetic resonance imaging, and cardiac catheterization [3]-[6]. These methods are often costly and constraining. Furthermore, they are often deployed inside the clinic. A prolonged, continuous monitoring of high-risk subjects outside of the clinic is not feasible with the modalities stated above.

Wearable sensors using non-invasive modalities have been considered as one of the most promising devices to be used in a home-based ubiquitous monitoring scenario for high-risk populations [7]-[9]. In our scope, a wearable device that is capable of detecting cardiovascular abnormalities such as AS could potentially prevent the progressive development of this disease and increase the ratio of successful treatments.

Chest cardio-mechanical sensing is a non-invasive wearable sensing modality that has been enthusiastically researched in recent years [10]. It can be categorized into two types, seismo-cardiography (SCG), which is the measurement of the linear acceleration components of the chest wall induced by the heartbeat [11], [12], and gyro-cardiography (GCG), which is the recording of heart-induced rotational vibrations of the chest wall in the form of angular speed [13]. SCG and GCG signals can be conveniently acquired by placing a micro-electromechanical system (MEMS) inertial measurement unit (IMU) on the chest wall, where the accelerometer picks up the SCG signal and the gyroscope measures the GCG waveform.

Seismo- and gyro-cardiogram signals have been effective for

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

Chenxi Yang is a Ph.D. candidate at the Department of Electrical and Computer Engineering at Stevens Institute of Technology, Hoboken, NJ, 07030 USA. (e-mail: cyang13@stevens.edu).

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

Philip Green is an Assistant Professor of Medicine and interventional cardiologist at the Columbia University Medical Center, New York, NY 10032 USA.

Negar Tavassolian is an Associate Professor at the Department of Electrical and Computer Engineering at Stevens Institute of Technology, Hoboken, NJ 07030 USA.

the evaluation or classification of multiple CVDs such as coronary artery disease [14], [15], myocardial infarction [16], [17], atrial fibrillation [18], [19] and heart failure [20]. Specifically, our research group has conducted a study on the binary classification of cardiovascular system abnormality based on SCG and GCG signals, which is presented in [21]. However, there are no studies in the literature that are related to the detection and classification of aortic stenosis (AS) based on cardio-mechanical recordings.

The development of AS is a complicated progress. Furthermore, AS might be accompanied by different co-existing cardiac diseases which happen concurrently with the major disease. For instance, in a study performed on 240 AS patients, 32.5% of the patients had at least moderate mitral regurgitation co-existing with AS [22]. In this work, we propose a classification framework of AS. We evaluate the topic in two aspects to address the complicated development of AS. The first aspect is the classification between healthy subjects and general AS patients. We aim to extract the features that can distinguish between normal subjects and AS patients regardless of the specific AS conditions. In other words, we propose a binary classification framework that can detect general AS. The second aspect is the classification of AS with different co-existing cardiac diseases. There are many kinds of co-existing cardiac diseases such as VHDs, coronary artery diseases, and pulmonary hypertension [22]. In this study, we focus on the classification of AS with three different co-existing VHDs as explained in the next section.

In the literature, features from the heartbeat and heartbeat segments are generally used for feature extraction. For example in [19], features from HR and three different orders of HRV were used for the training and testing of a binary classifier of atrial fibrillation (AF). However, the development of AF is closely correlated with HR and HRV since AF corresponds to heart arrhythmia. On the other hand, aortic stenosis (AS) is a structural disease of the heart. According to a long-term study, there are no correlations between changes in HR and the progression of AS [23]. The seismo-cardiogram (SCG) and gyro-cardiogram (GCG) signals used in this study are closely related to valve activities [11]-[13]. It has been reported that the time-frequency components of SCG and GCG can be used to assess left ventricular health [12], [13], [16]. The aortic valve, located in the human heart between the left ventricle and the aorta, is closely related to the health of the left ventricle. Therefore, the features extracted from the time-frequency components of SCG and GCG reveal great potential to classify between healthy and AS patients. Hence in this work we focus on these features statistically and morphologically.

From the SCG and GCG recordings, we extract features out of single-cycle and multi-cycle segments. The empirical mode decomposition (EMD) and continuous wavelet transform (CWT) techniques are applied to perform time-frequency analyses. The significance level of the features is assessed by the analysis of variance (ANOVA) method. Then the selected components are fed to machine learning classifiers for two classification tasks.

TABLE I SUMMARY OF DEMOGRAPHIC INFORMATION OF SUBJECTS PARTICIPATED IN EXPERIMENTS. (AVERAGE  $\pm$  STANDARD DEVIATION)

| Category | Age (years)      | Height (cm)        | Weight (kg)       |
|----------|------------------|--------------------|-------------------|
| AS       | 68.90 $\pm$ 8.43 | 162.47 $\pm$ 11.30 | 72.31 $\pm$ 13.71 |
| Healthy  | 24.75 $\pm$ 1.83 | 170.59 $\pm$ 8.62  | 64.95 $\pm$ 11.30 |

To the best of our knowledge, this is the first time that the feasibility of AS classification is evaluated using cardio-mechanical modalities. Our work also reveals the important features for the classification of AS. It is, however, worth mentioning that impedance cardiography (ICG) has been proposed as a non-invasive modality to perform classification of VHDs. In a study using the ICG signal, it was revealed that the use of cepstral features is effective for the detection of cardiovascular diseases including VHDs [24]. These features were extracted from the ICG signals of twenty-five CVD classes and used as an input to a classifier. The classification had an accuracy of 95.40%. In another study, five classes of VHDs were classified using a combination of temporal and time-frequency features from 75 patients [25]. These features were fed to the support vector machine classifier and a diagnostic accuracy of 98.94% was reported.

Compared to other modalities such as ICG, there are two key advantages in using chest cardio-mechanical signals. The first advantage is that they do not need direct contact with the skin via electrodes, and are therefore more convenient for users to wear, especially for long-term monitoring at home. Furthermore, the wide availability of IMUs in daily electronics, such as smartphones and wearable gadgets, makes seismo- and gyro-cardiography promising solutions for a low-cost, ubiquitous, and mobile healthcare system. For instance, several pilot studies have validated the feasibility of extracting seismo- and gyro-cardiogram signals from IMUs in smartphones [17], [26], [27].

The layout of this paper is as follows. Section II presents the experimental protocol and data acquisition. The methodology of extracting, selecting, and training the features is introduced in Section III. Section IV introduces the experimental results and Section V provides discussions of the results. We conclude the paper and outline the future work in Section VI.

## II. EXPERIMENTAL PROTOCOL AND DATA ACQUISITION

### A. Datasets

Twenty inpatient subjects from the cardiac care units of

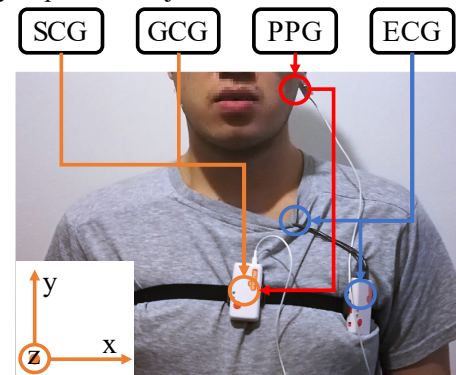


Fig. 1. Experimental setup with the sensor nodes which collect SCG, GCG, PPG, and ECG.

TABLE II  
SUMMARY OF CO-EXISTING VHDs OF PATIENTS WITH AORTIC STENOSIS

| Subject Index | Mitral Insufficiency | Mitral Stenosis | Tricuspid Regurgitation |
|---------------|----------------------|-----------------|-------------------------|
| 1             | No                   | No              | Yes                     |
| 2             | No                   | No              | No                      |
| 3             | No                   | No              | Yes                     |
| 4             | No                   | No              | No                      |
| 5             | No                   | No              | Yes                     |
| 6             | No                   | No              | Yes                     |
| 7             | No                   | No              | No                      |
| 8             | No                   | Yes             | Yes                     |
| 9             | No                   | No              | No                      |
| 10            | No                   | Yes             | No                      |
| 11            | No                   | No              | No                      |
| 12            | No                   | Yes             | No                      |
| 13            | Yes                  | No              | No                      |
| 14            | No                   | No              | No                      |
| 15            | No                   | No              | No                      |
| 16            | No                   | No              | Yes                     |
| 17            | No                   | No              | No                      |
| 18            | No                   | No              | No                      |
| 19            | Yes                  | No              | No                      |
| 20            | Yes                  | Yes             | Yes                     |

(Mitral insufficiency is also known as mitral regurgitation, which is used in the original records from CUMC).

Columbia University Medical Center (CUMC) participated in the collection of AS recordings. All the patients 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. The AS cohort includes ten male and ten female subjects whose ages, weights and heights varied as 69-97 years old, 44-91 kg, and 139-182 cm, respectively. The healthy subjects' recordings correspond to ten male and ten female subjects of 22-35 years old, 52-80 kg, and 154-180 cm. Table I summarizes the statistical demographic information of all the recruited subjects in average and stand deviation.

Table II shows the co-existing VHDs of the AS subjects. The VHD conditions of the AS group include aortic stenosis (AS) with or without concurrent mitral insufficiency (MI), AS with or without concurrent mitral stenosis (MS), and AS with or without concurrent tricuspid regurgitation (TR). As seen in Table II, three patients have a co-existing VHD of MI, four patients have a co-existing VHD of MR, and seven patients have AS with TR. Nine out of the twenty patients do not have any of these VHDs. It is to be noted that the co-existing VHDs are not mutually exclusive. A patient could possibly have multiple co-existing VHDs. For instance, subject #20 has MI, MS, and TR simultaneously. In this scenario, an ultimate solution to the classification problem is a multi-label classification, in which the co-existing VHDs are evaluated

individually. The subjects could then be classified into multiple co-existing VHD classes at the same time. However, the training of such a multi-label classifier requires a larger database with wider coverage of co-existing VHD combinations than the patient database in this study. Therefore, we select a more practical multi-class classification, where we classify among four different AS classes: AS without any co-existing disease, AS with MI, AS with MR, and AS with TR. Therefore, the data for the training and testing of the multi-class classification are from AS patients with only one co-existing VHD or without any co-existing VHDs. In other words, subjects #8 and #20 were excluded from the training and testing of the multi-class classification. On the other side, all of the patient data were used for binary classification since the classification task is to detect general AS regardless of the specific kind of the co-existing VHD.

### B. Experimental Protocol

The subjects were asked to sit at rest on a bed or chair for at least five minutes during each experiment. It is to be noted that the subjects can also be measured at other postures such as the supine position without any significant influence on the results. Subjects breathed naturally without controlling their breathing depths. The patient experimental protocol was approved by the Institutional Review Board of CUMC under protocol number AAAR4104. The experiments with healthy subjects were approved by the Committee for the Protection of Human Subjects at Stevens Institute of Technology under protocol number 2017-008 (N).

## III. METHODS

### A. The Hardware System

As shown in Fig. 1, a commercial wearable sensor node (Shimmer 3 from Shimmer Sensing, marked by the orange circle [28]) is attached to the center of the sternum along the third rib using a chest strap. A three-axis MEMS accelerometer (Kionix KXRB5-2042, Kionix, Inc.) measures the SCG signal, and a three-axis MEMS gyroscope (Invensense MPU9150, Invensense, Inc.) records the 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 also taken by a standard four-lead ECG system (the blue circle in Fig. 1), which is wire-connected to the center chest sensor. In addition, an ear-lobe photoplethysmography (PPG) sensor marked by the red circle on the right side of the figure is connected to a second shimmer sensor node attached to the left side of the

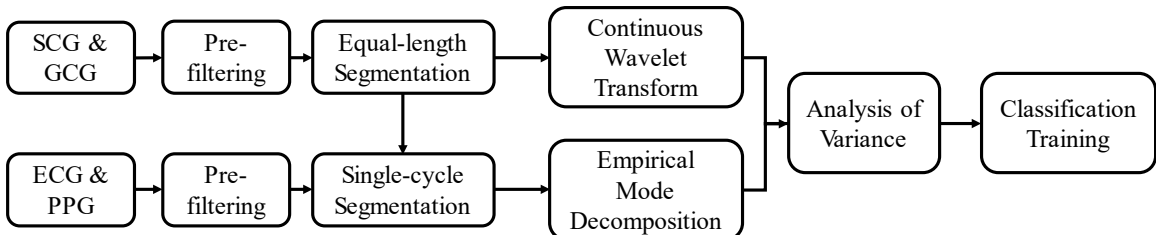


Fig. 2. Block diagram of the signal processing framework.

chest wall to provide additional cardiovascular information as illustrated in Fig. 1. Both the ECG and PPG signals work as references to provide segmentation information during the signal pre-processing, which will be introduced in Section II.B. The accelerometer and gyroscope ranges are  $\pm 2$  g and  $\pm 250$  degrees per second (DPS) respectively, and all the sensor recordings are sampled at a sampling rate of 256 Hz. Data are transmitted via Bluetooth to a computer for digital signal processing. The recorded data from the two sensor nodes are synchronized by the software provided by the manufacturer and imported into MATLAB (R2018) for further processing.

Fig. 2 illustrates the block diagram of the proposed signal processing framework. This framework is extended from our previous study in [21]. First, the raw signals are pre-processed and segmented into equal-length segments and cycle-to-cycle segments. Then, the features are extracted from processed cardio-mechanical signals using time-frequency analyses by CWT and EMD methods. This step is followed by a feature selection which selects the most significant features via analysis of variance (ANOVA). Finally, the features from labeled observations that will be fed into training algorithms. Details of the signal processing steps are introduced in the following sections.

### B. Signal Pre-processing and Segmentation

Although seismo- and gyro-cardiogram signals are recorded from all three axes of the devices, we focus on recordings from one axis in each modality. For the SCG signal, the  $z$ -axis in our system, i.e., the standard single-axis SCG signal, is chosen. For the GCG modality, the  $y$ -axis rotation signal is selected due to the higher quality for this axis reported in [12] and [29].

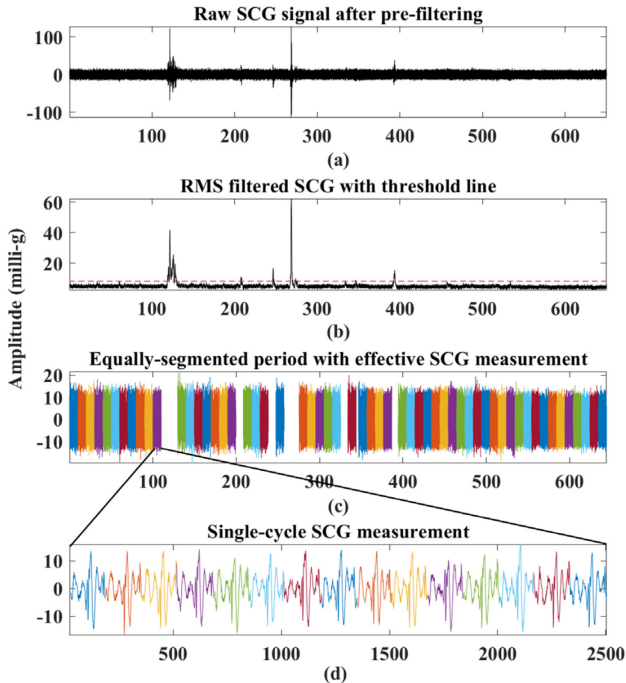


Fig. 3. (a) Raw seismo-cardiogram (SCG) signal after band-pass filtering. (b) RMS-filtered SCG signal with median product threshold line (red-dot line). (c) Equally-segmented SCG measurements based on (1)-(3). (d) Single-cycle segmentation of SCG from the tenth equally-segmented recording. (Unit of  $x$ -axis: seconds for (a) to (c), samples for (d)).

All signals are pre-filtered with zero-phase infinite impulse response (IIR) bandpass filters to remove baseline wandering and focus on the informative bandwidths. The frequency band for SCG and GCG is from 0.8 Hz to 25 Hz. The ECG signal is pre-filtered from 5 Hz to 30 Hz, and the PPG signal from 0.8 Hz to 60 Hz.

The filtered SCG and GCG signals are divided into low-noise equal-length segments by using a threshold-based exclusion method with a root-mean-square (RMS) filter modified from [19]. The purpose of this step is to reject noisy segments for the time-frequency analysis and feature generation steps. Features from equal-length segments provide the highest consistency of observations for the best results of classification training. In this study, we use a period length of 10 s, which is an experimental number based on our trials in [21] and references [19]. The step size of the RMS filter is set to 500 ms, and the threshold is set to 1.5 times of the median RMS value. More details of the equal-length segmentation could be found in [21].

After the segmentation of the 10-second recordings, the segments are further separated into single cardiac cycles. The segmentation of the cardiac cycles starts with the detection of the R-peaks in the simultaneous ECG recordings.

The R-peaks are marked as the local maxima in the ECG signal with a minimum value of 40% of the maximum peak and a minimum distance of 30 data samples between consecutive peaks. This is a modified heartbeat detection algorithm [30] constructed to apply with our sensing device. LL-LA ECG recordings were used for most of the patients while the LL-RA ECG recordings worked as a back-up. In other words, the highest-priority segmentation source was LL-LA ECG. In the cases where the LL-LA ECG signal was noisy, the LL-RA ECG signal was selected. The PPG recordings supplied additional information when both LL-LA and LL-RA ECG recordings were noisy. In such cases, the cardio-mechanical recordings were segmented based on the PPG peaks instead of ECG peaks. The systolic maximum peaks in PPG are considered as the segmentation markers and are detected with the method described in [31]. A fixed offset of 45 samples is applied to adjust the difference between ECG and PPG cycles.

Fig. 3 illustrates this step on a representative seismo-cardiogram graph. It is seen that the recordings with large motion artifacts have been excluded due to the RMS filtering step in Fig. 3 (b). After the equal-length segmentation in Fig. 3 (c), the temporal recording segments are further divided into cardiac cycles. As an example, the recordings are separated into 15 single-cycle segments as shown in Fig. 3 (d) from one ten-second segment.

### C. Feature Extraction

The training of the classification algorithms requires a set of labeled observations, i.e., many groups of features which are categorized into different classes (normal, AS, and AS with co-existing VHDs). In this study, one observation is defined as the features from one ten-second segment combined with the features from all the single-cycle segments that are related to

that ten-seconds segment. As shown in Fig. 2, the ten-second segment is processed with continuous wavelet transform while the single-cycle segments are processed via empirical mode decomposition. The detailed procedures are as follow.

### 1) Continuous Wavelet Transform

Our preliminary study in [21] indicates that the abnormality of cardio-mechanical signals may occur in intermediate cycles in between several regular cycles. In addition, the abnormal signal might have a dominant component at certain frequencies. Therefore, we expand the signal in the time-frequency domain by using the continuous wavelet transform (CWT) method [35]. The wavelet used in this study is a Morse wavelet. The equation for this wavelet is expressed as follows.

$$\Psi_{P,\gamma}(\omega) = U(\omega)a_{P,\gamma}\omega^{\frac{P^2}{\gamma}}e^{-\omega^\gamma} \quad (1).$$

In (1),  $P$  is the time-bandwidth product and  $\gamma$  is the symmetry parameter. In this study,  $\gamma$  is selected as 3 and the value of  $P$  is 120. In combination with the sampling frequency of 256 Hz, the configured CWT results in a high-resolution region between 0.8 Hz and 25 Hz (the passing band of our pre-filter). To focus on this band, we then select from 0.79 Hz to 25.39 Hz precisely. More details about our CWT method can be found in [21].

Several statistical features are extracted from the full-wavelet array. The first feature is the maximum power which is defined in detail in [21]. Also, additional statistical features are extracted in each frequency band including mean, standard deviation, and median. These features are designed to detect abnormal activities in the scope of a ten-second segment. Next, we perform an empirical mode decomposition (EMD) of each single-cycle cardio-mechanical beat.

### 2) Empirical Mode Decomposition

The empirical mode decomposition (EMD) is the method used in Hilbert-Huang transform as a time-frequency analysis technique [36]. It decomposes the signal into components called intrinsic mode functions (IMFs) which are determined using a technique named the sifting process [36], [37]. This technique analyzes the envelopes of the signal based on the local maxima and local minima, denoted as  $H_{\max}(n)$  and  $H_{\min}(n)$ , respectively. For a given signal  $s(n)$ , the mean  $m(n)$  of the two envelopes is calculated as follows:

$$m_1(n) = \frac{H_{\max}(n) + H_{\min}(n)}{2} \quad (2)$$

The first IMF is determined as:

$$IMF_1(n) = s(n) - m_1(n) \quad (3)$$

The subsequent IMFs are then obtained by considering the first IMF as new input and repeating the procedure iteratively. The iterations of finding new IMFs are stopped when the standard deviation of the differences between two adjacent IMFs is smaller than 0.2. The maximum number of iterations is set to 100.

Fig. 4 illustrates the EMD results of a representative SCG single-cycle recordings. All 4 IMFs are plotted to show the different signal components. The order of the IMFs from low

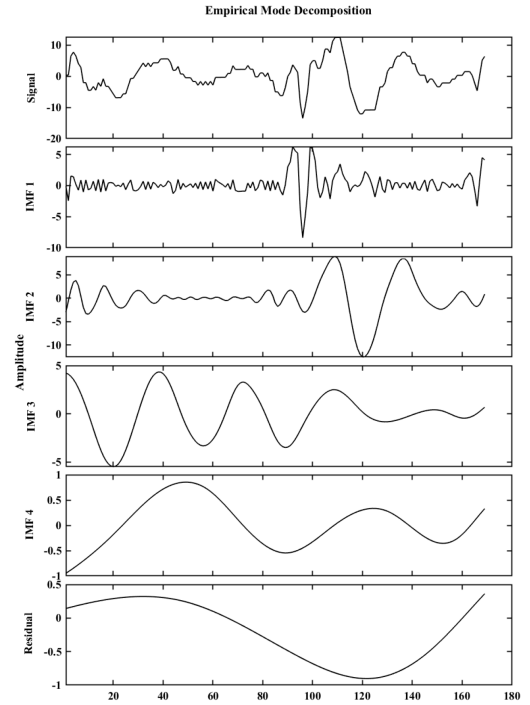


Fig. 4. A representative EMD of a single-cardiac-cycle seismo-cardiogram recording showing the ensemble-averaged signal, all 4 of the IMFs, and the residual signal.

to high represents the density of the frequency components, i.e., the first IMF contains the highest frequency components.

For each ten-second measurement, the IMFs are extracted from all the related signal-cycle recordings and then ensemble-averaged. Then the features are extracted from each of the four ensemble-averaged IMFs, including the mean, the standard deviation, the skewness, and the entropy.

### 3) Feature Selection

Table III summarizes the features from CWT and EMD outputs. It is shown that 4 features are generated from 55 frequency bins based on CWT, which summarizes as 220 features in total. Moreover, 4 features are extracted from 4 IMFs of the EMD. As a result, 236 features are generated from each of the SCG and GCG signals.

In this study, we conduct two separate one-way ANOVA tests for the classification training tasks. The first test is for the binary classification of AS in general. The PCs extracted from the time-frequency analyses are collected into a table with observation marks of either 'AS' or 'normal'. Then the ANOVA test evaluates the significance between the two classes. The second test is for the multi-class classification of AS with co-existing VHDs. AS patients with different co-existing VHDs are labeled as 'aortic stenosis (AS)', 'AS + mitral insufficiency (MI)', 'AS + mitral stenosis (MS)', and, 'AS + tricuspid regurgitation (TR)'. Then the ANOVA test is applied to all these four classes.

The selected features after the ANOVA are then fed into the machine learning algorithm as introduced in the following section.

TABLE III  
SUMMARY OF FEATURES FROM CWT AND EMD

| Source             | Feature Type           | Range (number of features)            |
|--------------------|------------------------|---------------------------------------|
| <b>CWT</b>         |                        |                                       |
| <b>SCG and GCG</b> | Maximum                | 0.79-25.39 Hz (55 from each modality) |
|                    | Mean                   | 0.79-25.39 Hz (55 from each modality) |
|                    | Standard Deviation     | 0.79-25.39 Hz (55 from each modality) |
|                    | Median                 | 0.79-25.39 Hz (55 from each modality) |
| <b>EMD</b>         |                        |                                       |
| <b>SCG and GCG</b> | Mean                   | IMF 1-4 (4 from each modality)        |
|                    | Standard deviation     | IMF 1-4 (4 from each modality)        |
|                    | Skewness               | IMF 1-4 (4 from each modality)        |
|                    | Entropy                | IMF 1-4 (4 from each modality)        |
| <b>Total</b>       | 236 from each modality |                                       |

#### D. Machine-learning Classifier Training

In this study, we have used three types of classifiers: The decision tree (DT), the random forest (RF), and the neural network (NN).

##### 1) Decision Tree

The decision tree (DT) is a decision support tool representing a set of choices in a tree-like graph [40], [41]. In this study, a medium tree with a maximum of 20 splits is applied. The split criterion is based on entropy. The maximum depth is set to 7. We optimize the minimum number of leaf size by using the automatic optimization function in MATLAB, which searches among integers in the range from 1 to half of the observations at log-scale.

##### 2) Random Forest

Random forest (RF) is an ensemble learning method that constructs multiple decision trees. [41]. The number of trees is 30, and the maximum number of splits is 798. The minimum leaf size is set to 1, and the number of variables for each decision split is selected by the square root rule, which takes the square root of the number of variables.

##### 3) Neural Network

We implemented a simple neural network with 30 input, 20 hidden, and 10 output neurons. The initial learning rate is 0.001.

##### 4) Evaluation Metrics and Validation Methods

To evaluate the performance of the classification algorithm, several metrics are calculated. The class-specific metrics are sensitivity (SE), specificity (SP), and accuracy (AC), which are defined as follows:

$$SE = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

$$SP = \frac{TN}{TN+FP} \times 100\% \quad (5)$$

$$AC = \frac{TP+TN}{TP+FN+TN+FP} \times 100\% \quad (6)$$

where TP, FP, TN, and FN are true positives, false positives, true negatives, and false negatives accordingly.

In addition to the class-specific metrics, the overall accuracy is also calculated to evaluate the general performance of the multi-class classification results. The overall accuracy is defined as the percentage of the true positives from all classes divided by all the classification datapoints.

The average recording length of all the subjects is 5.7 minutes. A total of 1157 segments are extracted from 40 subjects. For the binary classification, a total of 960 seismocardiogram (SCG) segments are collected as the balanced training set from 16 AS patients (480 abnormal segments) and 16 healthy subjects (480 normal segments). Similarly, a total of 960 gyro-cardiogram (GCG) segments are extracted from the same group of patients and healthy subjects. The trained models are validated using the classic ten-fold cross-validation method. Data from the remaining 4 patients (88 segments) and 4 healthy subjects (88 segments) are left out for testing.

Only the data from AS patients are used in the multi-class classification. As mentioned in Section II, eighteen out of the twenty AS patients are selected for the classification of co-existing VHDs. A total of 540 segments are extracted from the SCG and GCG recordings. There are 56 segments from subjects with AS + MI, 55 segments from subjects with AS + MS, and 158 segments from subjects with AS + TR. The subjects with no co-existing CVD provide 271 segments. The evaluation method for the multi-class classifier model is ten-fold cross-validation.

## IV. EXPERIMENTAL RESULTS

### A. Feature Selection Results

All seismo- and gyro-cardiogram segments are extracted into 55 frequency bins with four features per bin, and four IMFs with four features per IMF.

#### 1) Feature Selection for Binary AS Classification

Table IV shows the ANOVA results for the binary classification based on seismo- and gyro-cardiogram features. The selected features are based on the threshold of  $p < 0.05$ .

It is reported that 24 features are extracted from SCG. Out of the 24 features, 22 are from CWT. There are 13 maximum features, which range from 1.07 Hz to 9.20 Hz. 3 features from 2.30 Hz to 4.60 Hz are extracted based on mean values. Standard deviations of the CWT outputs provide 4 features from 1.00 Hz to 1.23 Hz while the median values give 2 features at 1.32 Hz and 2.14 Hz. Only two of the features are from EMD, which are the mean value from IMF1 and the skewness from IMF3. Overall, the frequency range of the selected features is from 1.00 Hz to 9.20 Hz.

The gyro-cardiogram (GCG) reports significant results from 30 features. Specifically, 27 of the features are from CWT. Particularly, 18 features ranging from 3.03 Hz to 10.58 Hz are extracted from the maximum features. 4 features are extracted based on mean values from 0.87 Hz to 1.07 Hz. Furthermore, 3 features are from standard deviations ranging from 4.29 Hz to 4.94 Hz. It is also seen that 2 median features are selected at 4.00 Hz and 4.29 Hz. The EMD provides 3 features which are mean features from IMF1 and IMF3 and the skewness feature

TABLE IV  
ANOVA TEST RESULTS FOR BINARY CLASSIFICATION

| Source       | Feature Type       | Range and number of features |
|--------------|--------------------|------------------------------|
| <b>SCG</b>   |                    |                              |
| <b>CWT</b>   | Maximum            | 1.07 - 9.20 Hz, 13 features  |
|              | Mean               | 2.30 - 4.60 Hz, 3 features   |
|              | Standard Deviation | 1.00 - 1.23 Hz, 4 features   |
|              | Median             | 1.32 and 2.14 Hz, 2 features |
| <b>EMD</b>   | Mean               | IMF1                         |
|              | Skewness           | IMF3                         |
| <b>Total</b> |                    | 24                           |
| <b>GCG</b>   |                    |                              |
| <b>CWT</b>   | Maximum            | 3.03 - 10.58 Hz, 18 features |
|              | Mean               | 0.87 - 1.07 Hz, 4 features   |
|              | Standard Deviation | 4.29-4.94 Hz, 3 features     |
|              | Median             | 4.00 and 4.29 Hz 2 features  |
| <b>EMD</b>   | Mean               | IMF 1, IMF 3                 |
|              | Skewness           | IMF 3                        |
| <b>Total</b> |                    | 30                           |

from IMF3. In summary, the features from GCG range from 0.87 Hz to 10.58 Hz.

We can observe that most of the significant features are based on CWT maximum values. Also, there are more features from CWT than from EMD. It is also shown that the features are mostly below 11 Hz.

## 2) Feature Selection for the Multi-class Classification of AS with Co-existing VHDs

The features for multi-class classification are also analyzed with a one-way ANOVA test. Table V summarizes the ANOVA test results of seismo- and gyro-cardiogram features with multi-class labels. The selected features are based on the threshold of  $p < 0.05$ . The SCG results provide 21 features in total. It is seen that 18 out of the 21 features are from CWT and 3 are from EMD. The features have a frequency range of 0.87 Hz to 6.51 Hz. Also, 10 features from 1.07 Hz to 6.51 Hz are from the maximum features of CWT, which are the majority of the features.

Moreover, there are 29 features selected from GCG. Among them, 27 are from CWT and 2 from EMD. The features range from 0.87 Hz to 9.87 Hz.

Comparing the feature selection results of binary and multi-class classifications, it is observed that the selected features are different. For instance, SCG-based features have a larger frequency range for binary classification than for multi-class classification (1.00 – 9.20 Hz vs. 0.87 – 6.51 Hz respectively). This difference suggests that the best features for the binary classification of AS are not the same as the best features for the classification of AS with co-existing VHDs. In other words, the features that represent the difference between normal and general AS patients may not represent the difference among AS patients with different co-existing VHDs.

## B. Classification Results

As mentioned, we use three different classifiers, decision tree (DT), random forest (RF), and the neural network (NN). The SCG and GCG features can be used independently for

TABLE V  
ANOVA TEST RESULTS FOR MULTI-CLASS CLASSIFICATION

| Source       | Feature Type       | Range and number of features |
|--------------|--------------------|------------------------------|
| <b>SCG</b>   |                    |                              |
| <b>CWT</b>   | Maximum            | 1.07 – 6.51 Hz, 10 features  |
|              | Mean               | 2.30 - 4.29 Hz, 3 features   |
|              | Standard Deviation | 0.87 - 1.23 Hz, 3 features   |
|              | Median             | 1.32 and 1.15 Hz, 2 features |
| <b>EMD</b>   | Mean               | IMF1                         |
|              | Skewness           | IMF 2, IMF 3                 |
| <b>Total</b> |                    | 21                           |
| <b>GCG</b>   |                    |                              |
| <b>CWT</b>   | Maximum            | 4.29 – 9.87 Hz, 15 features  |
|              | Mean               | 0.87 - 1.41 Hz, 5 features   |
|              | Standard Deviation | 4.29-4.94 Hz, 4 features     |
|              | Median             | 3.03 – 6.51 Hz 3 features    |
| <b>EMD</b>   | Mean               | IMF 3                        |
|              | Skewness           | IMF 2                        |
| <b>Total</b> |                    | 29                           |

training or can be combined as one set of features. Therefore, three combinations of extracted features (SCG only, GCG only, and SCG + GCG) are used as inputs to these two classifiers. Hence, we analyze the performance with nine methods, i.e., i) SCG + DT, ii) SCG + RF, iii) SCG + NN, iv) GCG + DT, v) GCG + RF, vi) GCG + NN, vii) SCG + GCG + DT, viii) SCG + GCG + RF, and ix) SCG + GCG + NN. These nine methodologies are evaluated and compared.

## 1) Binary Classification of General AS

Table VI summarizes the binary classification metrics of the trained classification model with seismo-cardiogram (SCG), gyro-cardiogram (GCG), and combined SCG + GCG features respectively. The means of the classification results during the 10-fold cross-validation are presented. The test results based on the leave-out data were also summarized. It is observed that the metrics from leave-out data analysis are lower than the corresponding values from the cross-validation test. This result is expected since the cross-validation doesn't rule out subject-dependent factors. The test with the leave-out data is closer to future application scenarios. Therefore, the results from the leave-out data test are evaluated as the final performance metrics of this binary classification task.

As shown in Table V, the RF classifier outperforms the DT classifier. The results from SCG + RF are better than the results from SCG + DT. The SE, SP and AC are 95.83%, 93.75%, and 94.79% from DT and 96.46%, 95.42%, and 95.94% from RF. The sensitivity (SE), specificity (SP) and accuracy (AC) are higher from GCG + RF (97.29%, 97.50%, and 97.40%, respectively) compared to GCG + DT (96.88%, 97.08%, and 96.98%, respectively). Similarly, all the three metrics from SCG + GCG + RF are higher than the results from SCG + GCG + DT. The reports mention SE of 98.33% vs. 96.25%, SP of 99.58% vs. 97.71%, and AC of 98.96% vs. 96.98%, respectively.

The performance of SCG + NN is lower than that of SCG + DT in SE (94.38% vs. 95.83%), SP (92.71% vs. 93.75%), and AC (93.54% vs. 94.79%). Similarly, GCG + DT also

TABLE VI  
RESULT METRICS OF THE BINARY CLASSIFICATION

| Features  | Classifier | 10-Fold Cross Validation |             |          | Leave-out Data Test |             |          |
|-----------|------------|--------------------------|-------------|----------|---------------------|-------------|----------|
|           |            | Sensitivity              | Specificity | Accuracy | Sensitivity         | Specificity | Accuracy |
| SCG       | DT         | 95.83%                   | 93.75%      | 94.79%   | 93.64%              | 91.82%      | 92.73%   |
|           | RF         | 96.46%                   | 95.42%      | 95.94%   | 94.55%              | 94.55%      | 94.55%   |
|           | NN         | 94.38%                   | 92.71%      | 93.54%   | 91.82%              | 92.73%      | 92.27%   |
| GCG       | DT         | 96.88%                   | 97.08%      | 96.98%   | 95.45%              | 93.64%      | 94.55%   |
|           | RF         | 97.29%                   | 97.50%      | 97.40%   | 96.36%              | 93.64%      | 95.00%   |
|           | NN         | 96.25%                   | 95.83%      | 96.04%   | 94.55%              | 91.82%      | 93.18%   |
| SCG + GCG | DT         | 96.25%                   | 97.71%      | 96.98%   | 96.36%              | 97.27%      | 96.82%   |
|           | RF         | 98.33%                   | 99.58%      | 98.96%   | 97.27%              | 98.18%      | 97.73%   |
|           | NN         | 96.04%                   | 98.13%      | 97.08%   | 92.73%              | 96.36%      | 94.55%   |

(DT: decision tree, RF: random forest, NN: neural network).

outperforms GCG + NN. The results give 96.25%, 95.83%, and 96.04% from the NN classifier and 96.88%, 97.08%, and 96.98% from the DT classifier in SE, SP, and AC respectively. However, the SE and AC are higher from SCG + GCG + NN than from SCG + GCG + DT. The SE from DT is slightly lower than the SE from NN.

Based on these obtained results from Table V, we noticed that the RF classifier fed with combined SCG + GCG features provides the highest performance results compared to the other five methodologies. Also, the results from GCG are higher than results from SCG in all classifiers. This observation is similar to the observation in our earlier work in [21], where GCG reported better results in the classification of general cardiovascular abnormalities than SCG.

Fig. 5 illustrates the classification accuracy values using the three classifiers with different feature combinations. It shows that the RF classifier has the highest accuracy among all three feature combinations.

It can be summarized that using the DT classifier provides an average AC of 96.25%, an average SE of 96.32%, and an average SP of 96.18% regardless of the type of features used. Also, the RF classifier provided similar results of average accuracy, average sensitivity, and average specificity of 97.43%, 97.50%, and 97.36% respectively. The results from NN report average AC, SE, and SP of 95.55%, 95.56%, and

95.56% respectively. Compared to the metrics from the DT and NN classifier, the average metrics from the RF classifier are the highest. In conclusion, the experimental results suggest that the RF method slightly outperforms the DT and NN method in binary classification. The features from combined SCG and GCG provide the highest sensitivity and specificity in both classifiers. As a result, the combination with the best performance is SCG + GCG + RF, which reports SE of 98.33%, SP of 99.58%, and AC of 98.96%. The corresponding metrics are highlighted in green in Table V.

## 2) Multi-class Classification

Table VII shows the class-specific performance results for the four classes of diseases (AS, AS + MI, AS + MS, AS + TR) using the DT and RF classifiers fed with SCG, GCG, and combined SCG and GCG features. The mean values during the 10-fold cross-validation are presented for each metric.

On average, the SCG + RF combination reports the highest values of SE = 92.25% and AC = 94.26% for the class of AS. The highest SP is reported from SCG + GCG + DT, which is 96.65%. The corresponding table cells are highlighted in green. The highest AC for the class of AS + MI is 97.22% from SCG + GCG + RF results, which also reports the highest SE of 94.64% and the highest SP of 97.52%. These table cells are highlighted in red. For the class of AS + MS, the best results come from SCG + GCG + DT with SE of 96.36% and AC of 97.41%. Moreover, the best SP is 97.94% which is from SCG + GCG + RF. These cells are highlighted in blue. The SCG + RF combination produces the best classification of AS + TR with SE of 90.51%, SP of 97.64%, and AC of 95.56%. The corresponding table cells are highlighted in orange.

The class-specific accuracy metrics for the average of the three classifiers are 92.84%, 96.23%, 96.71%, and 94.38% for AS, AS + MI, AS + MS, and AS + TR respectively. It can be observed that the AS reports the lowest accuracy among all the classes. The results from the co-existing VHD of MS has the highest accuracy of 96.71%.

We also compared the overall performance among the three classifiers by averaging the metrics from all classes and all features combinations. The average accuracy from the DT

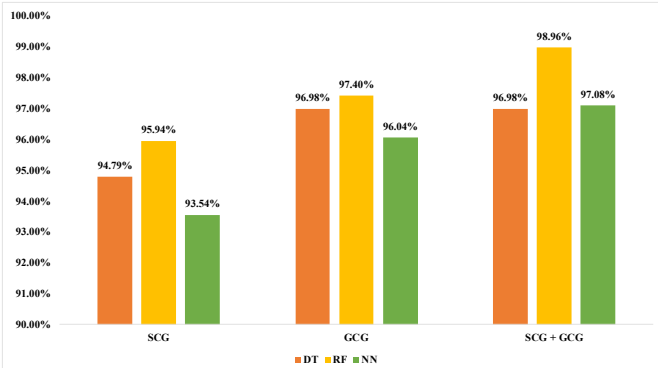


Fig. 5. A comparison of the accuracy in binary classification with different sets of features and methods (DT: Decision tree, RF: Random forest, NN: Neural network).

classifier is 94.85%, which is slightly lower than the value from the RF classifier which reports as 95.56%. However, it is higher than the result from NN, which is 94.72%. Similarly, the average SE from DT is 90.24%, slightly lower than the value of 91.70% from RF but higher than the value of 89.35% from NN. The average SP of all the classes from DT is 96.43%, which is marginally lower than the average SP of 96.88% from RF and higher than the average SP of 96.33% from NN.

The overall accuracy metrics from all the methods are illustrated in Fig. 6. It can be observed from this figure that the classification performances based on SCG features are higher than the results based on GCG and SCG + GCG features for both types of classifiers. The overall sensitivity of all classes from SCG is 88.28% for DT, 92.99% for RF, and 90.25% for NN methods. In comparison, the results from GCG are 90.42%, 90.58%, and 89.37% for DT, RF, and NN classifiers respectively. The SCG results report lower values in DT and

NN classifiers but higher values in the RF method. The outputs from SCG + GCG features are slightly higher than the results from GCG-only features, with a sensitivity of 92.02% compared to 90.42% via the DT classifier, and an accuracy of 91.54% compared to 90.58% via the RF classifier. However, the result from the NN classifier is the lowest among the three methods. In conclusion, the highest overall sensitivity is from SCG + RF.

Compared to the results of the binary classification in the previous section, the results in Table VI are lower in SE, SP, and AC. The highest accuracy from the binary classification is 98.33% (SCG + GCG + RF), while the highest class-specific accuracy from the multi-class classification is 96.36% (AS + MS with SCG + GCG + DT method). The best overall performance in accuracy comes from SCG + GCG + RF for binary classification (AC = 98.96%), while the best overall performance in overall performance for multi-class classification is from SCG + RF (92.99%). Furthermore, the results from GCG is generally better than SCG in binary classification, but worse than SCG in multi-class classification. Among the three classifiers, the performance of RF is the best in most cases. DT is comparable with slightly inferior metrics than RF. The results of NN are higher than those from DT and lower than those from RF in most tests.

## V. DISCUSSION

Table VIII summarizes the comparison of performance metrics between our work and other studies. Compared to other binary classification algorithms that deal with other types of CVD using SCG and GCG features [17], [19], and [26], our results show a slightly better performance in accuracy (AC). In [17], a binary classification of acute myocardial infarction was conducted with 17 patients and 23 healthy subjects. The highest results reported were 95.7% in specificity (SP) and 82.4% in sensitivity (SE). In our work in comparison, the highest SP is 99.58% and the highest SE is 98.96%, which are obtained from SCG + GCG + RF in binary classification. In [19], atrial fibrillation (AF) was classified using a linear least-square classifier with data from 13 patients. The specificity and sensitivity results were 96.4% and 99.9% respectively. The SP is lower than the SP of 99.58% in our work, and the SE is higher than the SE of 98.33% in our work. In [26], AF was detected

TABLE VII  
CLASS-SPECIFIC RESULTS OF THE MULTI-CLASS CLASSIFICATION

| Features  | Classifier                           | Detection Classes | SE     | SP     | AC     |
|-----------|--------------------------------------|-------------------|--------|--------|--------|
| SCG       | DT                                   | AS                | 88.56% | 94.80% | 91.67% |
|           |                                      | AS + MI           | 87.50% | 96.90% | 95.93% |
|           |                                      | AS + MS           | 89.09% | 97.53% | 96.67% |
|           |                                      | AS + TR           | 87.97% | 94.24% | 92.41% |
|           | RF                                   | AS                | 92.25% | 96.28% | 94.26% |
|           |                                      | AS + MI           | 94.64% | 97.52% | 97.22% |
|           |                                      | AS + MS           | 94.55% | 97.73% | 97.41% |
|           |                                      | AS + TR           | 90.51% | 97.64% | 95.56% |
|           | NN                                   | AS                | 90.41% | 95.54% | 92.96% |
|           |                                      | AS + MI           | 91.07% | 96.49% | 95.93% |
|           |                                      | AS + MS           | 90.91% | 97.11% | 96.48% |
|           |                                      | AS + TR           | 88.61% | 97.12% | 94.63% |
| GCG       | DT                                   | AS                | 89.30% | 95.91% | 92.59% |
|           |                                      | AS + MI           | 92.86% | 96.49% | 96.11% |
|           |                                      | AS + MS           | 90.91% | 97.32% | 96.67% |
|           |                                      | AS + TR           | 88.61% | 96.07% | 93.89% |
|           | RF                                   | AS                | 90.04% | 94.80% | 92.41% |
|           |                                      | AS + MI           | 92.86% | 96.69% | 96.30% |
|           |                                      | AS + MS           | 92.73% | 96.70% | 96.30% |
|           |                                      | AS + TR           | 86.71% | 97.38% | 94.26% |
|           | NN                                   | AS                | 89.30% | 94.80% | 92.04% |
|           |                                      | AS + MI           | 89.29% | 96.90% | 96.11% |
|           |                                      | AS + MS           | 90.91% | 96.49% | 95.93% |
|           |                                      | AS + TR           | 87.97% | 96.60% | 94.07% |
| SCG + GCG | DT                                   | AS                | 90.77% | 96.65% | 93.70% |
|           |                                      | AS + MI           | 91.07% | 96.90% | 96.30% |
|           |                                      | AS + MS           | 96.36% | 97.53% | 97.41% |
|           |                                      | AS + TR           | 89.87% | 96.86% | 94.81% |
|           | RF                                   | AS                | 91.88% | 95.17% | 93.52% |
|           |                                      | AS + MI           | 92.86% | 97.31% | 96.85% |
|           |                                      | AS + MS           | 90.91% | 97.94% | 97.22% |
|           |                                      | AS + TR           | 90.51% | 97.38% | 95.37% |
|           | NN                                   | AS                | 90.41% | 94.42% | 92.41% |
|           |                                      | AS + MI           | 87.50% | 96.28% | 95.37% |
|           |                                      | AS + MS           | 87.27% | 97.32% | 96.30% |
|           |                                      | AS + TR           | 88.61% | 96.86% | 94.44% |
| Green     | Best results to the class of AS      |                   |        |        |        |
| Red       | Best results to the class of AS + MI |                   |        |        |        |
| Blue      | Best results to the class of AS + MS |                   |        |        |        |
| Orange    | Best results to the class of AS + TR |                   |        |        |        |

(DT: decision tree, RF: random forest, NN: neural network, AS: aortic stenosis, MI: mitral insufficiency, MS: mitral stenosis, TR: tricuspid regurgitation).

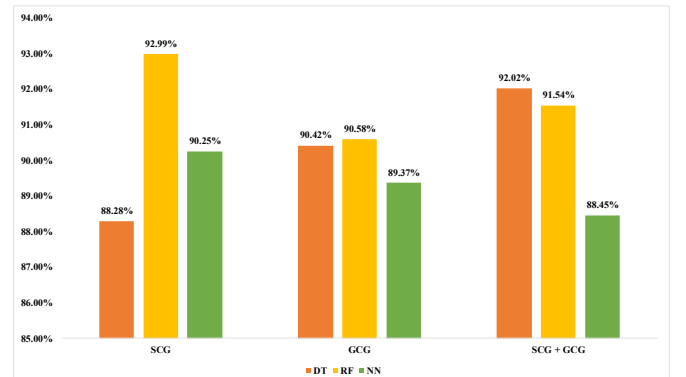


Fig. 6. A comparison of the overall sensitivity of multi-class classification with different sets of features and methods (DT: Decision Tree, RF: Random Forest, NN: Neural Network).

using a binary classifier trained with data from 16 patients and 23 healthy subjects. The best final accuracy, sensitivity, and specificity results are 97.4%, 93.8%, and 100% respectively. The AC and SE from [26] are lower than the AC of 98.96% and SE of 98.33% from our results with SCG + GCG + RF in binary classification. However, the SP from [26] is higher than the SP of 99.58% from our work.

In our previous study in [21], we conducted a binary classification of general cardiovascular abnormalities using SCG and GCG features. The classifiers were trained with data collected from 12 patients and 12 healthy subjects. The best performance of leave-out validation in AC, SE, and SP were 97.7%, 96.9%, and 97.0% respectively. In this work, the best performance in AC, SE, and SP are 98.96%, 99.58%, and 98.33% respectively from SCG + GCG features with RF classifier, which are all higher than the results in [21].

The results from ICG-based methods can also be compared with this work. In [24], features were extracted from the ICG signals of twenty-five CVD classes and used as an input to a classifier. The classification had an AC of 95.40%. In another study [25], five classes of VHDs from 75 patients were classified using a combination of temporal and time-frequency features. These gave a diagnosis accuracy of 98.94%. Our binary classification results report the highest AC of 98.96%, which is higher than the results in [24] and [25]. Our multi-class classification results are however lower than the values from ICG-based methods in [25].

In summary, our binary classification results report higher AC and SE values than the binary classification results with the same modalities that detect other cardiovascular abnormalities. The reported SP from our work is lower than the best result (100%) from [26]. Also, the results from the binary classification are comparable with results from ICG-based multi-class classifications. However, the multi-class classification results are less satisfactory compared to the results from other modalities and methods.

The classification of AS with co-existing VHDs is a more challenging task than regular CVD classifications since the

differences among AS with co-existing VHDs are less significant than the differences among different categories of CVD. This result is therefore expected. There are also other factors related to the classification results. For instance, the number of subjects included in the database influence the performance of a classifier. The segments extracted for the four co-existing CVDs are not well-balanced. The population bias might also influence the results of the classifiers. The severity and stages of the VHDs also bring differences to the characteristics of the subjects [42], [43].

## VI. CONCLUSIONS AND FUTURE WORK

This paper introduces a novel approach to detect and classify aortic stenosis (AS) using seismo-cardiogram (SCG) and gyro-cardiogram (GCG) signals. Different combinations of features are evaluated using DT, RF and NN methods. Feature analysis results suggest that the major features are maximum statistical features from CWT below 11 Hz. Our results suggest that the proposed solution could be used to classify and detect general aortic stenosis as well as to classify different co-existing VHDs. The accuracy of multi-class classification is lower than the binary classification, which remains to be further improved. There is a difference between the feature-method combinations that achieve the highest accuracy in the two different classification tasks. This indicates that binary and specific aortic stenosis classifications should be treated individually for optimized performances.

It is also worth mentioning that the segmentation of cardiac cycles relies on the modalities of ECG and PPG in the current framework. In future setups, a standalone segmentation algorithm could be used such as the methods in [44]-[47]. Without the integration of ECG and PPG, the form factor and cost of the device could be minimized. Moreover, the device will be more convenient to the users since no electrodes or clips are needed.

To achieve increased robustness of the system and better performance among specific AS classes, future research would encompass the use of other classification methods, such as the K-nearest neighbors (KNN) method [48]. The tuning of the

TABLE VIII  
PERFORMANCE COMPARISON WITH OTHER RESEARCH METHODS

| Methods                           | Target CVD(s)               | Classifier(s)         | AC     | SP     | SE     | Reference |
|-----------------------------------|-----------------------------|-----------------------|--------|--------|--------|-----------|
| <b>Binary Classification</b>      |                             |                       |        |        |        |           |
| SCG + GCG                         | General Abnormality         | SVM                   | 97.7%  | 97.0%  | 96.9%  | [21]      |
| SCG + GCG                         | Acute Myocardial Infarction | Kernel SVM            | N/A    | 95.7%  | 82.4%  | [17]      |
| SCG + GCG                         | Atrial Fibrillation         | RF                    | 97.4%  | 100%   | 93.8%  | [26]      |
| SCG                               | Atrial Fibrillation         | Linear Least-square   | N/A    | 96.4%  | 99.9%  | [19]      |
| Proposed Method                   | Aortic Stenosis (AS)        | RF                    | 98.96% | 99.58% | 98.33% |           |
| <b>Multi-class Classification</b> |                             |                       |        |        |        |           |
| ICG                               | 25 CVD classes              | Discriminant analysis | 95.40% | N/A    | N/A    | [24]      |
| SVM with ICG                      | 5 VHD classes               | SVM and KNN           | 98.94% | 100%   | 97.85% | [25]      |
| Proposed Method                   | 4 AS classes                | RF                    | 96.36% | 97.94% | 97.41% |           |

(DT: decision tree, RF: random forest, SVM: support vector machine, KNN: K-nearest neighbor)

parameters could further optimize the performance of the models [49], [50]. The importance of the features could also be analyzed using other methods such as the elastic net [51]. We could also increase the number of features to improve the learning of the classifiers. For example, we can add temporal and spectral features from and combine them with the time-frequency features. A larger database of subjects with a more extensive variety of co-existing VHDs should also be collected to improve the demographic significance and increase the coverage of co-existing VHDs. In addition, the full wavelet array could be used as a 2-D feature for a deep-learning task based on graphics [52]. Although this classifier would need more computing power, it could potentially reveal many new applications.

In the future, a two-stage classification framework could be envisioned in which the recordings are fed into a binary classification of AS as the first stage. The observations which are classified as AS will then be sent to the next stage where a multi-class classification algorithm determines the specific AS conditions. Furthermore, the multi-class classification could be replaced by a multi-label classification algorithm [53], [54]. The framework could then classify AS with multiple co-existing VHDs more effectively. The method proposed in this study shows promising potential for use in the monitoring of VHDs to prevent patients from sudden critical cardiac situations.

#### REFERENCES

- [1] WRITING GROUP MEMBERS, Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Turner, M. B. (2013). "Heart Disease and Stroke Statistics—2013 Update: A Report From the American Heart Association." *Circulation*, vol. 127, no. 1, e6–e245.
- [2] Klabunde, Richard. Cardiovascular Physiology Concepts. Lippincott Williams & Wilkins, 2011.
- [3] L. G. Svensson, "Aortic Valve Stenosis And Regurgitation: An Overview Of Management." *Journal of Cardiovascular Surgery*, vol. 49, no.2 pp.297, 2008.
- [4] Iung B, Baron G, Butchart EG, *et al.* "A Prospective Survey of Patients with Valvular Heart Disease in Europe: The Euro Heart Survey on Valvular Heart Disease." *European Heart Journal*, vol. 24, no. 13, pp. 1231-1243, 2003.
- [5] D. Mozaffarian, *et al.*, "Executive Summary: Heart Disease and Stroke Statistics-2016 Update: a Report from the American Heart Association," *Circulation*, vol. 133, no. 4, p. 447, 2016.
- [6] S. Yusuf, *et al.*, "The World Heart Federation's Vision for Worldwide Cardiovascular Disease Prevention," *The Lancet*, vol. 386, no. 9991, pp.399-402, 2015.
- [7] M. S. Kohn, *et al.*, "Implementation of a Home Monitoring System for Heart Failure Patients: A Feasibility Study." *JMIR Research Protocols* vol. 6, no.3, 2017.
- [8] J. Andreu-Perez *et al.*, "From Wearable Sensors to Smart Implants—Toward Pervasive and Personalized Healthcare," *IEEE Transaction on Biomedical Engineering*, vol. 62, no. 12, pp. 2750-2762, Dec. 2015.
- [9] J. M. Reh, S. Murphy, S. Kumar, *Mobile Health: Sensors, Analytic Methods, and Applications*. Springer. 2017
- [10] O. T. Inan, *et al.*, "Ballistocardiography and Seismocardiography: A Review of Recent Advances", *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1414-1427, 2015.
- [11] M. Etemadi and O. T. Inan, "Wearable Ballistocardiogram and Seismocardiogram Systems for Health and Performance," *Journal of Applied Physiology*, v. 124, no. 2, pp. 452-461, 2018.
- [12] C. Yang, N. Tavassolian, "Combined Seismo- and Gyro-cardiography: A More Comprehensive Evaluation of Heart-Induced Chest Vibrations," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1466-1475, September 2018.
- [13] M. J. Tadi, *et al.*, "Gyrocardiography: A New Non-invasive Monitoring Method for the Assessment of Cardiac Mechanics and the Estimation of Hemodynamic Variables." *Scientific Reports*, vol.7, 2017.
- [14] I. Kubacka, *et al.*, "Usefulness of Seismocardiography for The Diagnosis of Ischemia in Patients with Coronary Artery Disease", *Annual of Noninvasive Electrocardiology*, vol. 10, no. 3, pp. 281-287, 2005.
- [15] K. Tavakolian, *et al.*, "Estimating Cardiac Stroke Volume from The Seismocardiogram Signal", *Canadian Medical and Biological Engineering Society*, vol. 33, no. 1, Mar. 2018.
- [16] O. Inan, "Wearable Sensing of Left Ventricular Function," *Mobile Health*, vol. 362, pp. 265–287, July 2017.
- [17] O. Lahdenoja, *et al.*, "A Smartphone-Only Solution for Detecting Indications of Acute Myocardial Infarction," *2017 IEEE International Conference on Biomedical & Health Informatics (BHI)*, pp. 197-200.
- [18] Z. Ifikhar, *et al.* "Multiclass Classifier based Cardiovascular Condition Detection Using Smartphone Mechanocardiography." *Scientific Reports* vol. 8, no.1, pp. 9344, 2018.
- [19] T. Humanen *et al.*, "Automated Detection of Atrial Fibrillation Based on Time-Frequency Analysis of Seismocardiograms," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1233-1241, Sept. 2017.
- [20] 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.
- [21] C. Yang, *et al.*, "A Binary Classification of Cardiovascular Abnormality Using Time-Frequency Features of Cardio-mechanical Signals," *2018 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*.
- [22] P. Faggiano, *et al.*, "Prevalence of Comorbidities And Associated Cardiac Diseases in Patients with Valve Aortic Stenosis. Potential Implications for The Decision-Making Process." *International journal of cardiology*, vol. 159, no. 2, pp.94-99, 2012.
- [23] S. W. Davies, A. H. Gershlick, and R. Balcon. "Progression of Valvar Aortic Stenosis: A Long-Term Retrospective Study." *European Heart Journal*, vol. 12, no.1, pp.10-14, 1991.
- [24] R. Ben Salah, T. Alhadidi, S. Mansouri, M. Naouar, "A New Method for Cardiac Diseases Diagnosis", *Advances in bioscience and biotechnology*, vol. 6, no. 4, pp. 311–319, 2015.
- [25] S. Chabchoub, S. Mansouri, R. Ben Salah, "Detection of valvular heart diseases using impedance cardiography ICG," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 2, pp. 251–261, 2018.
- [26] O. Lahdenoja *et al.*, "Atrial Fibrillation Detection via Accelerometer and Gyroscope of a Smartphone," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 108-118, Jan. 2018.
- [27] C. Yang, N. Tavassolian, "A Feasibility Study on A Low-Cost, Smartphone-Based Solution of Pulse Transit Time Measurement Using Cardio-Mechanical Signals," *2017 IEEE Healthcare Innovations and Point of Care Technology (HI-POCT)*, pp. 93-96.
- [28] Shimmer Sensing. (2016). [Online]. Available: [www.shimmersensing.com](http://www.shimmersensing.com)
- [29] M. J. Tadi, *et al.*, "Automatic Identification of Signal Quality for Heart Beat Detection in Cardiac MEMS Signals," *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 137-140, 2017.
- [30] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *in IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230-236, March 1985.
- [31] C. Yang and N. Tavassolian, "Pulse Transit Time Measurement Using Seismocardiogram, Photoplethysmogram, and Acoustic Recordings: Evaluation and Comparison," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 733-740, May 2018.
- [32] J. Wahlström, I. Skog, P. Händel, F. Khosrow-Khavar, K. Tavakolian, P. K. Stein, and A. Nehorai, "A Hidden Markov Model for Seismocardiography," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 10, pp. 2361-2372, Oct. 2017.
- [33] F. Khosrow-Khavar, K. Tavakolian, A. Blaber and C. Menon, "Automatic And Robust Delineation of The Fiducial Points of The Seismocardiogram Signal for Noninvasive Estimation of Cardiac Time Intervals," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1701-1710, Aug. 2017.
- [34] V. Zakeri, A. Akhbardeh, N. Alamdari, R. Fazel-Rezai, M. Paukkunen and K. Tavakolian, "Analyzing Seismocardiogram Cycles to Identify The Respiratory Phases," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1786-1792, Aug. 2017.

- [35] A. Taebi, H. A. Mansy, "Time-Frequency Distribution of Seismocardiographic Signals: A Comparative Study." *Bioengineering*, vol. 4, no. 2, pp. 32, 2017.
- [36] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N-C. Yen, C.C. Tung, H.H. Liu, "The Empirical Mode Decomposition and The Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis", *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, vol. 454, no. 1974, pp. 903–995, 1998.
- [37] R.B. Pachori, P. Avinash, K. Shashank, R. Sharma, U.R. Acharya, "Application of Empirical Mode Decomposition for Analysis of Normal and Diabetic RR-Interval Signals", *Expert Systems with Applications*, vol. 42, no. 9, 4567–4581, 2015.
- [38] H. Abdi, and L. J. Williams. "Principal Component Analysis." *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no.4, pp. 433-459, 2010.
- [39] C. M. Judd, G. H. McClelland, & C. S. Ryan, *Data Analysis: A Model Comparison Approach to Regression, ANOVA, And Beyond*. Routledge, 2017.
- [40] W.Y. Loh, "Regression Tress with Unbiased Variable Selection and Interaction Detection" *Statistica Sinica*, pp.361-386, 2002.
- [41] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random Forests and Decision Trees" *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 5, p.272, 2012.
- [42] M. Dolgin, NYH Association, A.C. Fox, R. Gorlin, RI Levin, New York Heart Association. Criteria Committee. Nomenclature and Criteria for Diagnosis of Diseases Of The Heart And Great Vessels. 9th edition, Boston, MA: Lippincott Williams and Wilkins; March 1, 1994.
- [43] Criteria Committee, New York Heart Association, Inc. Diseases of The Heart And Blood Vessels. Nomenclature And Criteria For Diagnosis, 6th edition Boston, Little, Brown and Co. 1964, p. 114.
- [44] F. Khosrow-khavar, *et al.*, "Automatic Annotation of Seismocardiogram with High-Frequency Precordial Accelerations," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1428-1434, July 2015.
- [45] F. Khosrow-khavar, K. Tavakolian and C. Menon, "Moving Toward Automatic and Standalone Delineation of Seismocardiogram Signal," *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015*, pp. 7163-7166.
- [46] A. Laurin, et al. "Accurate and Consistent Automatic Seismocardiogram Annotation without Concurrent ECG," *2015 Computing in Cardiology Conference (CinC), Nice, 2015*, pp. 25-28.
- [47] C. Yang, S. Tang and N. Tavassolian, "Utilizing Gyroscopes Towards the Automatic Annotation of Seismocardiograms," *IEEE Sensors Journal*, vol. 17, no. 7, pp. 2129-2136, April1, 1, 2017.
- [48] R.J. Samworth, "Optimal Weighted Nearest Neighbour Classifiers," *The Annals of Statistics*, vol. 40, no. 5, pp.2733-2763, 2012.
- [49] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.
- [50] J. Bergstra and Y. Bengio. "Random Search for Hyper-Parameter Optimization." *Journal of Machine Learning Research*, vol. 13, pp.281–305, 2012.
- [51] H. Liu, and H. Motoda, *Computational methods of feature selection*. CRC Press, 2007.
- [52] G. Litjens, T. Kooi, B.E. Bejnordi, A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A Survey on Deep Learning in Medical Image Analysis." *Medical image analysis*, vol. 42, pp.60-88, 2017.
- [53] G. Tsoumakas, and I. Katakis, "Multi-label Classification: An Overview," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 3, no. 3, pp.1-13, 2007.
- [54] A. Wosiak, K. Glinka, and D. Zakrzewska, "Multi-label Classification Methods for Improving Comorbidities Identification." *Computers in biology and medicine*, vol. 100, pp.279-288, 2018.