
Electronic Theses and Dissertations, 2020-

2020

Seismocardiography - Genesis, and Utilization of Machine Learning for Variability Reduction and Improved Cardiac Health Monitoring

Peshala Thibbotuwawa Gamage
University of Central Florida

 Part of the [Biomechanical Engineering Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd2020>

University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2020- by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Thibbotuwawa Gamage, Peshala, "Seismocardiography - Genesis, and Utilization of Machine Learning for Variability Reduction and Improved Cardiac Health Monitoring" (2020). *Electronic Theses and Dissertations, 2020-*. 622.

<https://stars.library.ucf.edu/etd2020/622>

SEISMOCARDIOGRAPHY - GENESIS, AND UTILIZATION OF MACHINE
LEARNING FOR VARIABILITY REDUCTION AND IMPROVED CARDIAC
HEALTH MONITORING.

by

PESHALA PRIYADARSHANA THIBBOTUWAWA GAMAGE
M.Sc. in Aerospace Engineering, University of Central Florida, USA, 2017
B.Sc. in Mechanical Engineering, University of Moratuwa, Sri Lanka, 2013

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Mechanical and Aerospace Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2020

Major Professor: Hansen A Mansy

© 2020 PESHALA PRIYADARSHANA THIBBOTUWAWA GAMAGE

ABSTRACT

Seismocardiography (SCG) is the measured chest surface vibrations resulting from cardiac activity. Although SCG can contain information that correlate with cardiac health, its utility may be limited by lack of understanding of the signal genesis and a variability that can mask subtle SCG changes. The current research utilized medical imaging reconstruction and finite element method (FEM) to simulate SCG by modeling the propagation of myocardial movements to the chest surface. FEM analysis provided a link between myocardial movements and the SCG signals measured at the chest surface and suggested that myocardial movement is a primary source of SCG. Increased understanding of the sources and propagation of SCG may help increase the utility of SCG as a cardiac monitoring tool.

To reduce the variability of SCG measured in human subjects, unsupervised machine learning (ML) was implemented to group SCG beats into clusters with minimal intra-cluster heterogeneity. The clustering helped reduce the SCG variability and unveiled consistent relations with the respiratory phases and SCG morphology. This clustering reduced noise in calculating signal features and provided additional useful features. The study also analyzed longitudinal SCG from heart failure (HF) patients in order to predict HF readmission. Here, many time- and frequency-domain SCG features were extracted. Certain features showed good correlations with readmission. Using supervised ML algorithms, high classification accuracies (up to 100%) were achieved suggesting high SCG utility for monitoring HF patients and possibly other heart conditions. Effective

monitoring followed by timely intervention can lead to improved patient management and reduced mortality.

ACKNOWLEDGMENT

First and foremost, I wish to thank my advisor, Dr. Hansen Mansy, for his guidance and support throughout my graduate career. I am grateful to Dr. Richard Sandler, Dr. Alain Kassab, and Dr. Dajhong Wu, who served as my committee members and gave me valuable feedback towards the successful completion of my dissertation. Further, I thank all the great teachers I had since my childhood, especially Dr. Y.W.R Amarasinghe for his guidance and support at University of Moratuwa, Sri Lanka.

I'd like to thank my friend and colleague, Md Khurshidul Azad, who immensely supported throughout my research and constantly motivated me. I also like to thank my friends, Dr. Fardin Khalili and Dr. Amirtaha Taebi who encouraged me with valuable advice. Special thanks goes to Andrew Spiewak who made the research more joyful and for being a great friend. I'd like to express thanks to my colleagues Rajkumar Dhar and Tanvir Hassan for their support in research.

I take this opportunity to thank my Sri Lankan friends, who have been there for me in different phases of my life. I especially thank Navindra Wijeyeratne, Kalpana Madushan, Singith Abey Siriwardena, Nadun Kuruppumullage, Thiwanka Sanjitha, Mafaz Zafer, Kasun Jayaneththi and Milinda Yapa, for being there with me at rough times of my life.

Special thanks to my parents and my siblings for continuously encouraging me and being there for me at every point of my life. I am genuinely blissful for having you all in my life, and nothing in my life would have been possible without you. I dedicate my dissertation to my Father who was the greatest person I ever knew.

TABLE OF CONTENTS

LIST OF FIGURES	x
LIST OF TABLES	xix
CHAPTER 1 - INTRODUCTION	1
Reducing the Variability of Seismocardiographic (SCG) Signals.....	2
Numerical Modeling of SCG	3
Prediction of Heart Failure (HF) patient readmission using SCG	4
CHAPTER 2 – REDUCING SCG MORPHOLOGICAL VARIABILITY	5
Methods.....	9
Experimental measurements	10
Preprocessing	11
Filtering	11
SCG segmentation	12
Unsupervised machine learning	13
Averaging SCG beats.....	17
Results and Discussion	23
Optimum number of clusters	23
Purity of clustering with labels HLV/LLV and INS/EXP	25
Analyzing cluster distribution with respiratory phases.....	28

Cluster switching	30
Relation between heart rate and clustering	33
Intra-cluster variability.....	33
Conclusion	38
CHAPTER 3 – NUMERICAL MODELING OF SCG	40
Numerical Methods.....	42
Medical imaging	43
Motion tracking.....	45
Validation of motion tracking.....	48
Computational modeling.....	52
Results and Discussion	56
Simulated SCG waveform	56
Effect of chest soft tissue stiffness.....	58
Surface distribution of SCG.....	59
Effect of heart position (due to respiratory motion)	64
Acceleration over the heart wall	67
Summary and Conclusions	68
CHAPTER 4 – PREDICTION OF HEART FAILURE PATIENT READMISSION USING SCG.....	71

Methodology	73
Study protocol	73
Data acquisition.....	74
Pre-processing of SCG data	75
Clustering of SCG morphology for accurate feature extraction	76
SCG Features	77
SCG waveform variability features	79
Spectral features	80
Spectro-temporal features	82
Time-amplitude features	88
Instantaneous frequency SCG features	88
Heart rate variability features.....	90
Feature Analysis and Selection.....	90
Minimum redundancy maximum relevance algorithm.....	93
Principal component analysis for visualizing the features.....	94
Linear discriminant analysis	97
Classification Algorithms	98
Support vector machine	99
K- nearest neighbor classification.....	101

Results and Discussion	101
Feature sorting using statistical analysis.....	102
Feature derivation using LDA	105
Classification using machine learning (ML).....	107
HF patients longitudinal trajectory prediction using ML	111
Use of clinical data for predicting HF readmission	115
Conclusion	117
LIST OF REFERENCES	118

LIST OF FIGURES

Figure 1- The cardiac events identified in the SCG signal as proposed by Crow et al(1994).The abbreviations are: AS-atrial systole, MC-mitral valve closure, IM-isovolumic movement, AO-aortic valve opening, IC-isotonic contraction, RE- rapid ejection, AC-aortic valve closure, MO-mitral valve opening, RF- rapid filling	5
Figure 2- Two different SCG morphologies observed in a healthy subject during a single recording session	6
Figure 3- Methodology workflow.....	9
Figure 4- Experimental setup.....	11
Figure 5 – Original and filtered SCG signal in (a) time domain (b) zoomed in time domain (c) frequency domain (d) zoomed in frequency domain.....	12
Figure 6 – Segmentation of the SCG signal using ECG beats.....	13
Figure 7-(a) Variations of the length of segmented SCG events due to heart rate (b) Zoomed view	14
Figure 8- Distance matrix and the optimum warping path for signals X and Y ...	15
Figure 9- Associated points between signals X and Y when the dissimilarity is measured with (a) Euclidean and (b) DTW measures (Zhang et al., 2014).....	16
Figure 10- DBA average calculated for a set of SCG beats	19
Figure 11- Average SOD for different number of clusters. Two clusters were chosen using the elbow method	24
Figure 12- Box-whiskers plot of average Silhouette value (calculated for all 17 subjects) for different number of clusters	25

Figure 13- Labeling SCG beats, HLV=High lung volume, LLV=Low lung volume, EXP=expiration, INSP=Inspiration, Red trace=Lung volume, Blue trace=Respiratory flow rate.....	26
Figure 14 -Purity values for two labelling criteria; INS/EXP and HLV/LLV. HLV/LLV labeling provided higher purity levels.	27
Figure 15- SOD values before clustering and when the clusters were separated based on HLV/LLV, INS/EXP and k-medoid solution. K-medoid had least SOD in all subjects.....	28
Figure 16- Cluster results (event locations) plotted on (a) Respiratory flowrate cycle (b) Lung volume cycle (c) both flowrate and lung volume cycles in 3D	29
Figure 17-(a) Four respiratory phases identified in a simplified lung volume waveform (b) Cluster distribution in lung volume and flowrate space in a typical subject	30
Figure 18: The cut-off function between the two clusters calculated from linear SVM theory in a typical subject.....	31
Figure 19- Cluster distributions and the cluster cut-off equations plotted in flowrate vs. lung volume space for all study subjects	32
Figure 20- (a) Defining respiratory phase angle (b) cluster distribution with heart rate as a function of respiratory phase angle (for one subject)	33
Figure 21- Clustered SCG beats plotted separately for cluster 1 and cluster 2 (for one subject)	34

Figure 22- Distance plotted from cluster medoids to beats in the same cluster in ascending order for (a) cluster 1 (b) cluster 2. Closest (10%) of the beats to the medoid of (c) cluster 1 (d) cluster 2. Furthest (10%) of the beats from the medoid of (e) cluster 1 (f) cluster2.....	35
Figure 23-Calculated DBA for 2 clusters for all subjects.....	38
Figure 24- Methodology	42
Figure 25- ECG tracing for cine MRI imaging gating. Different colors refer to acquisition time windows. The windows with the same color are averaged to produce the cine MRI images.	44
Figure 26- MRI images (a) 2- chamber view (b) 4-chamber view and (c) short axis views in different planes, (d) voxel volume and (e) the slice orientation with respect to the chest denoted by 2 parallel lines	45
Figure 27-Movement of $n \times n \times n$ neighborhood from frame to frame	46
Figure 28-Tracked ground truth of landmarks on ventricle wall by two observers (Tobon-Gomez et al., 2013).....	49
Figure 29- Locations of ground truth (GT) and tracked locations at (a) end systolic and (b) close to end diastolic frames in 3D Euclidean space (all dimensions are in mm) 50	
Figure 30-Tracked contours at the ventricle walls (endocardium) with the displacement vectors (represented by red arrows) at the end systolic and end diastolic states. Top: short axis view, Bottom: Another view to increase clarity. The displacement between consecutive frames is shown by the arrows. The arrow length equals 8 times the displacement to make arrows more visible.....	51

Figure 31- Segmentation and 3D model construction. (a) Segmented voxels from the MRI slices, (b) 3D model constructed by combining voxels.....	53
Figure 32- Chest location and detailed structures in the 3D modeled region (a) Muscular, bony and lung regions (b) Ribs, ICM, cartilage, sternal and xiphoid (c) location and orientation of the modeled region.	54
Figure 33- 3D representations of the computational model domain: (a) Computational mesh, (b) MRI-mapped displacements of heart surface and (c) and (d) more 3D views. The rib edges are colored in red.	54
Figure 34-Right: Calculated chest surface acceleration waveforms (in dorso-ventral direction) at several locations. The vertical dashed line indicates the timing of the ECG R-peak extracted from MRI. Left: Location of the 4 precordial points. The calculated waveforms had similarities with measured SCG.	56
Figure 35- Feature points identified in the calculated SCG waveform (at 5 th intercostal space) with the corresponding LV volume variation and R-wave timing. The feature points appeared consistent with the LV volume variation and previous studies (Crow et al. 1994)	57
Figure 36- The effect of chest soft tissue stiffness on the SCG waveform at 4 th ICS. Similar trends were observed at other chest surface locations. Higher stiffness appears to increase the waveform amplitude, which is consistent with simplified 1-D models.....	58
Figure 37- Simplified representation of a spring mass system with translational motion from a moving boundary	59

Figure 38- Surface acceleration distribution in dorso-ventral direction at: (a) &(c) the chest surface and (b) & (d) underneath the skin and pectoralis major. Sub-plots (a)&(b) show results at the SCG 1 peak timing while (c) and (d) are at SCG 2 peak timing. Timing was determined from the SCG signal at 4th intercostal space. The calculated surface SCG amplitude was comparable to measured values. The locations of the maximum SCG1 and SCG2 were around the 4th and 3rd ICS, respectively. The surface acceleration was less loud and more uniform than the subsurface SCG. There was less than 15% change in surface SCG amplitude in the 3cm diameter area surrounding the maximum amplitude location. This suggests that a sensor positioning error of 3 cm may be tolerable; all the measurements are in mm/s² 60

Figure 39-(a) Array of sensors attached to the chest surface (b) sensor locations with index. (Refer with SCG amplitude maps shown in Fig. 17)..... 62

Figure 40- Computational surface SCG amplitude maps at the timing of feature points: (a) MC (b) IC (c) AO (d) AC (e) MO and (f) identified feature points on the SCG signal measured at 5th ICS; all the measurements are in mm/s² 62

Figure 41- Experimental surface SCG amplitude maps at the timing of feature points: (a) MC (b) IC (c) AO (d) AC (e) MO and (f) identified feature points on the SCG signal measured at 5th ICS. (Small circles on Figures (a) to (e) represent the sensor locations as shown in Fig. 39 (b). The dashed line encloses the region to be compared with the computational model); all the measurements are in mm/s² 63

Figure 42- The heart location during respiration in (a) posterior-anterior projection and (b) lateral projection (Bogren, Lantz, Miller, & Mason, 1977) 64

Figure 43- Surface acceleration distribution in dorso-ventral direction at: (a) &(c) End inspiration state; and (b) & (d) end expiration during both SCG 1 and SCG 2 peak timing. The SCG peak tended to be louder at end inspiration. The location of the SCG1 peak moved superiorly at end expiration by about 3cm, which is comparable with the corresponding upward shift of the heart location; all the measurements are in mm/s² 65

Figure 44- Simulated SCG signals at different chest locations for the end inspiration and end expiration states. The waveform variability (calculated as the rms of the waveform difference) was comparable with experimental results..... 66

Figure 45- Calculated SCG at 4th ICS, 5th ICS, and sternum (a,b,c, respectively) and corresponding acceleration on the pericardium (d,e,f). The latter locations were determined by projecting the surface locations on the pericardium (The surface point locations are shown in Fig. 34). There was some similarity between surface and subsurface waveforms although the surface amplitudes were about 10 times lower than those at the pericardium, which is consistent with a previous study (Bombardini et al. 2007). 67

Figure 46- SCG measured at pericardium of a dog (measured using an accelerometer attached at left ventricular wall near interventricular septum) with corresponding ECG (Ozawa, Smith, and Craige. "Origin of the third heart sound. II. Studies in dogs." Circulation 67.2 (1983): 399-404). This measured acceleration shows characteristics similar to MRI-extracted signals, especially that of Fig 19 f..... 68

Figure 47-Methodology 73

Figure 48- Data session distribution of (a) readmitted and (b) non readmitted patients. Numbers represent the numbers assigned 74

Figure 49- Data acquisition setup.	75
Figure 50- Events closest to the medoid with their average waveforms (from same session)	77
Figure 51- Top: SCG waveform acquired from a patient, bottom: Time frequency distribution of SCG generated by PCT with identified high energy regions	85
Figure 52- Examples of two SCG waveforms with SCG1, SCG2 and SCG3 locations	87
Figure 53- (top): Instantaneous frequency signal with identified features (bottom): SCG heart cycle	89
Figure 54- An example of longitudinal variation of a feature in each patients. Red: re-admitted, Green: non readmitted. Data points belongs to each patient is linked with a dashed line.	91
Figure 55- Highlighted regions defining R and nR groups for feature analysis...	92
Figure 56- An Example of analyzing the differences of a feature value (SCG3 amplitude) distributions between R and Re groups.. Features from a healthy group ‘H’ is also included for comparison purposes. Distributions are represented by box and whisker plots and the p values between the groups are presented. SCG3 amplitude shows significant differences ($P < 0.05$) between H and the HF patient groups (i.e., R and nR) while no significance difference is observed between R and nR. Such observation for SCG3 amplitude is consistent with the previous study (Siejko et al., 2013).....	93
Figure 57- Example of PCA components generated using 2D dataset (i.e., 2 features) (ref: https://towardsdatascience.com)	95

Figure 58- Simplified illustration of PCA and LDA component axes (https://sebastianraschka.com/Articles/2014_python_lda.html)	97
Figure 59- Simplified illustration of SVM hyperplanes	100
Figure 60- K-NN classifier illustration (Sun, Shaohui. "Automatic 3D building detection and modeling from airborne LiDAR Point clouds." (2013).)	101
Figure 61- 2D biplot with first 2 PCA components representing PCA analysis of sorted features and class distribution	104
Figure 62- 3D biplot with first 3 PCA components representing PCA analysis of sorted features and class distribution	104
Figure 63- Box plots of sorted features and LDA derived feature	106
Figure 64- LDA derived feature (red-R , blue nR)	107
Figure 65- Variation of the training accuracy and the leave one out validation accuracy vs. number of features used (sorted using MrMR algorithm)	109
Figure 66- Variation of the training accuracy and the 10-fold cross validation accuracy vs. number of features used (sorted using MrMR algorithm)	109
Figure 67- Representation of SVM hyperplane with first 3 sorted features. Observation points of classes R and nR are represented by red and blue circles respectively.	110
Figure 68- Representation of SVM hyperplane with first 2 sorted features. Observation points of classes R and nR are represented by red and blue circles respectively.	110

Figure 69- Class trajectory of each patient predicted using SVM. (blue-nR, red-R)	111
Figure 70- Class trajectory of each patient predicted using K-NN. (blue-nR, red-R)	112
Figure 71- Class trajectory predictions of (a) readmitted and (b) non-readmitted HF patients when patients are divided into training and testing sets. (i.e., Patient set 33,35,36,41 and patient set 10,14,24 are used as the testing set for readmitted and non-readmitted groups,respectively)	114
Figure 72- ML predictions for the HF patients who have not completed the 6 months of study	115
Figure 73- plots of longitudinal variation of MAGGIC score in readmitted and non-readmitted HF patients.	116
Figure 74- Thoracic impedance (Z_0) values of HF patients	117

LIST OF TABLES

Table 1- Subject characteristics	10
Table 2- Material properties.....	55
Table 3- Data recording sessions longitudinally categorized in time for re-admitted and not re-admitted patients from data gathered in a 6 months duration after initial discharge.	74
Table 4- Feature description	78
Table 5- Sorted features based on unpaired t-test ($p < 0.05$)	103
Table 6- Validation accuracies of ML classification	107

CHAPTER 1 - INTRODUCTION

Seismocardiography (SCG) is the process of measuring the chest surface vibrations resulting from cardiac activity. These vibrations are typically measured by an accelerometer and believed to be primarily caused by valve closure and opening, blood momentum changes and myocardial movements (Gurev et al., 2012; Korzeniowska-Kubacka et al., 2006; Amirtahà Taebi, Solar, Bomar, Sandler, & Mansy, 2019). The method was initially introduced in 1957 by Eliot et al (Mounsey, 1957) and the term “Seismocardiography” was used by Bozhenko (Bozhenko, 1961) in 1961 who studied the use of SCG in space flights. Although, SCG is a promising method for the detection of important cardiac information such as cardiac timing intervals and cardiac contractility, its use was initially limited by the heavy accelerometers available at the time, while more rapid advancement were made in electrocardiography (ECG) and other medical imaging methods (Inan et al., 2015; Wilson, Bamrah, Lindsay Jr, Schwaiger, & Morganroth, 1993). Later, the progress in MEMS technology produced much lighter, highly sensitive and low-cost accelerometers, which has promoted the use of SCG as a feasible method for cardiac monitoring systems (Di Rienzo et al., 2011).

SCG provides information about the mechanical cardiac activity and, when combined with ECG which is indicative of the electrical activity, can provide a more complete picture of the cardiac health. This includes determining electromechanical time intervals such as pre-ejection period (PEP) which was found to correlate with cardiac health (Sahoo, Thakkar, Lin, Chang, & Lee, 2018). Results from several other studies have also suggested correlations between SCG signal features and different cardiac pathologies (Wick et al., 2012) including heart failure (Krishnan, Mansy,

Berson, Mentz, & Sandler, 2018; Amirtaha Taebi & Mansy, 2017a). Hence, SCG has high potential utility for diagnosis and monitoring of cardiac conditions. In addition, SCG has many advantages including being non-invasive, inexpensive, compatible with telemedicine, and not requiring highly trained operators.

Analysis of SCG waveforms usually focuses on both time and frequency features to gain better understanding of heart function and for classification of SCG under different heart conditions (Amirtaha Taebi, 2018; Amirtaha Taebi & Mansy, 2017b). In addition, SCG can provide information about the interactions between cardiovascular and pulmonary systems (Amirtaha Taebi & Mansy, 2017a; Amirtaha Taebi, Solar, & Mansy, 2018). Some studies have also employed SCG to monitor sleep apnea (Morillo, Ojeda, Foix, & Jiménez, 2010) and for estimating respiratory rate (Reinvuo, Hannula, Sorvoja, Alasaarela, & Myllyla, 2006).

The proposed research work is focused on three areas of SCG research not adequately presented in previous studies.

Reducing the Variability of Seismocardiographic (SCG) Signals

As SCG signals are associated with the mechanical movement (rather than electrical activity) measured over chest surface, SCG signal morphology is affected by different factors such as respiration (e.g., changes in lung volume), heart rate and cardiac contractility (Inan et al., 2015; Tavakolian et al., 2012). These factors may cause signal variabilities that mask subtle SCG changes that may be of diagnostic value. To reduce these variabilities, SCG waveforms can be separated into different groups (with each group having similar waveform morphology). This can help provide more accurate signal features, which may increase the diagnostic value of SCG.

Machine learning (ML) is a convenient tool to classify SCG events based on their morphological features without a need to have full understanding of the underlying mechanisms. Better classification may provide more insights into the effect of cardio-pulmonary interaction on SCG morphology. After SCG events are clustered into different groups, the phase of respiration (i.e., lung volume, inspiration, or expiration) of each event is examined to give insights into the effect of respiratory phases on SCG morphology thereby improving the utility of SCG monitoring for cardiac conditions (such as heart failure deterioration).

Numerical Modeling of SCG

Accurate understanding of how the movement of the heart is related to the morphology of the SCG signal measured on the chest surface will help to improve the usage of SCG as a reliable diagnostic tool to detect heart conditions. Certain studies have used electrocardiography to describe different peaks observed in a SCG heart cycle (Crow, Hannan, Jacobs, Hedquist, & Salerno, 1994; Giorgis et al., 2008), but these conclusions remain debatable (Akhbardeh et al., 2009). Relating the cardiac movements to SCG morphology is a challenging task, considering the complex movements of heart walls where certain cardiac events may superimpose on each other where they will amplify or nullify their effect on corresponding SCG morphology. However, the use of 4D (i.e., 3D+time) cardiac Cine-MRI imaging and image processing techniques to capture heart movements and relating them with SCG will help to unveil the origins of SCG signal. Furthermore, accurate capturing of 3D heart movements can provide boundary conditions for developing a computational model to study the propagation of heart movements to the chest surface.. The current study is

focused on developing a computational model to simulate the propagation of cardiac movements to the chest surface to accurately simulate the SCG signal. Such computational model can provide a platform to study the link between the SCG and cardiac movements and help enhance the understanding of SCG genesis.

Prediction of Heart Failure (HF) patient readmission using SCG

More than 20% of patients admitted for HF are readmitted within 30 days and up to 50% by 6 months (O'Connor, 2017). Accurate prediction of HF readmission will help doctors to pay close attention to HF patients who are likely to be readmitted and improve their quality of life. Extracting relevant SCG features from HF patients and the application of machine learning algorithms can be used to predict which HF patients are likely to be readmitted (ex:- within 30 days or within 60 days). The current study is focused on the use of SCG in predicting the readmission of reduced ejection fraction HF patients after their initial discharge.

CHAPTER 2 – REDUCING SCG MORPHOLOGICAL VARIABILITY

Fiducial points (i.e. certain SCG peaks) can be defined for SCG and can contain useful diagnostic information (Akhbardeh et al., 2009). For instance, several studies showed the consistent use of SCG to find the cardiac events (shown in Fig. 1) such as Mitral valve closing (MC), Aortic valve opening (AO), etc. (Akhbardeh et al., 2009; J. Zanetti, Poliac, & Crow, 1991; J. M. Zanetti & Tavakolian, 2013). However, some studies have reported inconsistencies in the location of the certain fiducial points (Akhbardeh et al., 2009; Khosrow-Khavar et al., 2015). The inconsistency and uncertainty of the correlation between fiducial points and cardiac events have posed limitations on the clinical application of SCG (Sørensen, Schmidt, Jensen, Søgaaard, & Struijk, 2018). This may be mainly due to the low time resolution in medical imaging modalities (Akhbardeh et al., 2009; Crow et al., 1994) that are usually used to establish the above correlation and also possibly due to the high variability of the SCG signal.

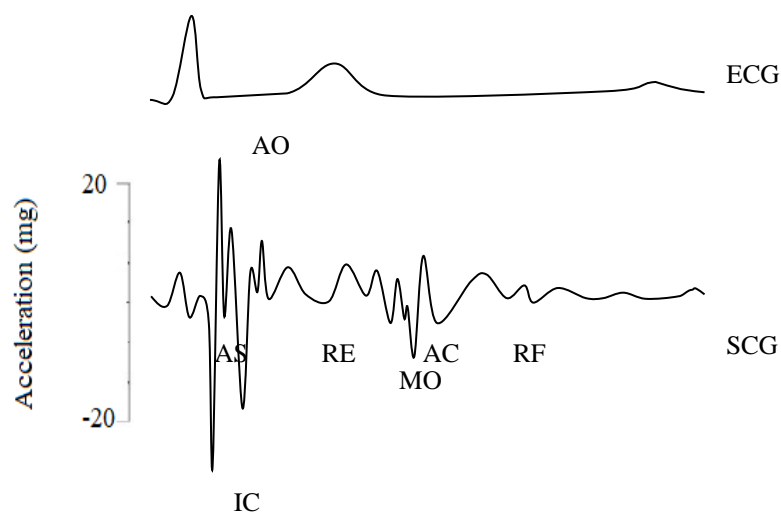


Figure 1- The cardiac events identified in the SCG signal as proposed by Crow et al(1994).The abbreviations are: AS-atrial systole, MC-mitral valve closure, IM- isovolumic movement, AO-aortic valve opening, IC-isotonic contraction, RE- rapid ejection, AC- aortic valve closure, MO-mitral valve opening, RF- rapid filling

SCG signals have a high inter and intra-subject variability. Signal morphology may also change among consecutive beats in the same data collection session of the same subject. Fig. 2 shows an example of two different SCG morphologies observed in measurements performed on a healthy subject during a single recording session. Unlike ECG, SCG signals are associated with the mechanical movement (rather than electrical activity) measured over chest surface. Consequently, SCG signal morphology can be largely affected by different factors such as respiration (which involves lung volume and intra thoracic pressure changes), heart rate, and cardiac contractility (Inan et al., 2015; Tavakolian et al., 2012). Effects of respiration on SCG morphology have been reported. An earlier study reported SCG waveform morphological changes for the inspiratory vs expiratory phases (Tavakolian, Vaseghi, & Kaminska, 2008). A recent study (Taebi & Mansy, 2017a) calculated waveform dissimilarities between SCG events and concluded that SCG morphology appeared to change with lung volume (which may correlate with intra-thoracic pressure) more than respiratory flow direction (inspiration vs. expiration).

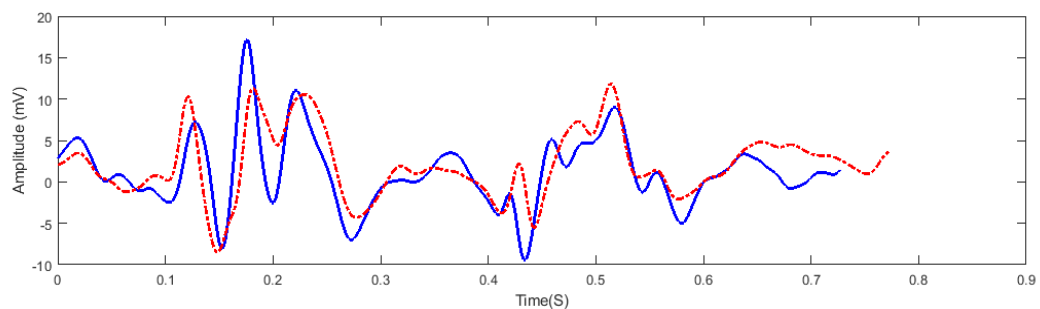


Figure 2- Two different SCG morphologies observed in a healthy subject during a single recording session

Interrelated mechanisms that may cause SCG changes with respiration include:

- 1) Heart position changes: During breathing, the diaphragm and lungs move leading to changes in the heart position relative to the SCG sensor that has a fixed

location at the chest surface. Lung and heart movements also change the structural composition of the thorax and consequently affect the properties of transmission path between the sensor and the sources of vibrations, which would affect the signal measured at the chest wall (Dai et al., 2014). For different subjects, this effect may vary based on their breathing pattern and tissue dimensions and properties.

2) Intrathoracic pressure variation: This pressure affects filling and ejection of blood flow from the different heart chambers. For example, the negative intrathoracic pressure induced during inspiration causes higher right heart filling and increases right heart output into the more compliant lungs (Cheuk & Sanderson, 1997) . Conversely, the positive expiratory intrathoracic pressure on the lungs inhibits the right heart filling and ejection during expiration. The resulting changes in blood flow would, in turn, cause morphological differences in the measured SCG signal.

Other processes such as chest wall and diaphragm contraction and relaxation may further interact with these mechanisms leading to complex variation in the SCG morphology that can be of diagnostic value. It is to be noted, however, that increased variability in the SCG waveform would reduce the precision of estimating the mean SCG waveform, which may interfere with accurate determination of SCG features and reduce SCG diagnostic utility.

To provide averages of SCG waveforms that would be optimally representative of SCG events, it is helpful to separate SCG waveforms into groups with minimum maximum intra- and inter-group dissimilarity, respectively. Hence, this study investigates the use of machine learning for optimal clustering of the SCG waveforms, which can help provide more accurate signal features and increase the diagnostic value

of SCG. While the focus here will be on SCG time domain morphology, similar analysis can be performed in the frequency domain.

Machine learning (ML) is a potential tool for grouping SCG events based on their temporal morphological features without a need to have full understanding of the underlying mechanisms of SCG variability (Gamage, Azad, Taebi, Sandler, & Mansy, 2018). A few studies have used supervised machine learning methods, such as support vector machine (SVM) and random forest (RF) to classify SCG events into respiratory phases such as inspiration vs. expiration or high lung volume vs. low lung volume (Solar, Taebi, & Mansy, 2017; Taebi et al., 2018; Zakeri et al., 2017). In these supervised classification studies, SCG morphology grouping is assumed a priori and the machine algorithm will iteratively try to minimize a loss function (using training and validating dataset) to optimize the accuracy of the classifier. Hence, the accuracy of these classifiers is not necessarily an indication of optimal grouping into classes with minimum and maximum intra- and inter-class variability, but rather indicates how well the algorithm can classify SCG events in to predefined groups (ex: Inspiratory /Expiratory groups).

In contrast to supervised ML, unsupervised ML is capable of clustering the input data into groups without defining the grouping a priori. Here, the unsupervised ML algorithms will optimize a function to separate the input data into clusters such that the data in a cluster are internally similar while dissimilarity between clusters are maximized. In this study, we employ an unsupervised machine learning approach to cluster SCG beats based on their temporal morphology. The main objectives of our study are:

- Clustering SCG events to minimize intra-cluster variability: Unsupervised ML is used for clustering SCG temporal waveforms in healthy subjects. The optimum number of clusters is decided by analyzing the variance (within and between) the clusters.

- Finding relations between cluster boundaries and respiratory phases and heart rate: The timing of the clustered SCG waveforms are compared with their respiratory phases (i.e., inspiratory vs. expiratory and high vs. low lung volume phases). The timing when SCG beats switch from one cluster to another is compared with the respiratory phases and heart rate changes.

- Calculating a representative SCG event for each cluster: After the clusters are defined, a beat that is a representative of the morphology of each cluster is calculated using an advanced shape averaging method. Ultimately, this representative SCG event may be used to define fiducial points for diagnostic purposes.

Methods

Fig. 3 summarizes the methodology employed in this study while more details are provided in following sections.

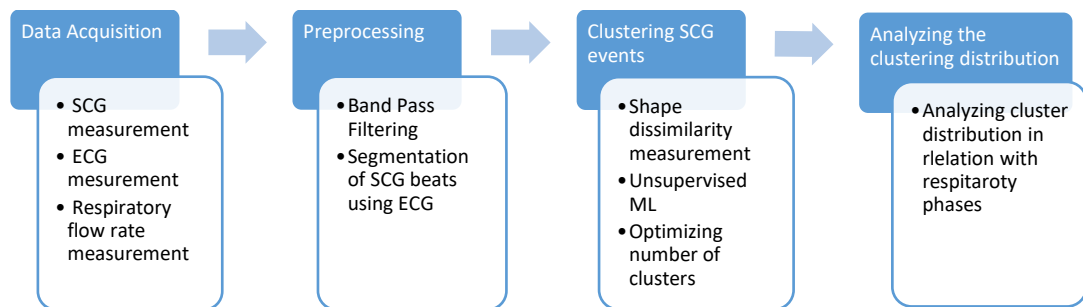


Figure 3- Methodology workflow

Experimental measurements

SCG signals were acquired from 17 healthy subjects after Institutional Review Board (IRB) approval. Subject characteristics are listed in table 1.

Table 1- Subject characteristics

Age (years)	23 ± 3.5
Height (cm)	168.5 ± 9
Weight (kg)	70 ± 13
BMI	24.5 ± 3.9

SCG was measured using a tri-axial accelerometer (Model: 356A32, PCB Piezotronics, Depew, NY) affixed to the chest surface using double sided medical-grade tape (B205-1, 3M, Minneapolis, MN) such that the measured z-component of the acceleration was normal to the chest surface (i.e., dorso-ventral component). The sensor was placed at the 4th intercostal space at the left lower sternal border. The signal from the accelerometer was amplified using a charge amplifier (Model: 482C, PCB Piezotronics, Depew NY) and then acquired using a data acquisition module (Model: IX-TA-220, iWorx Systems Inc, Dover, NH). The current SCG sensor is sensitive to chest wall movement due to respiration. While this movement is an artifact that can corrupt SCG, that artifact has a much lower frequency. This makes it easy to remove that artifact by low pass filtering, which is the approach implemented in this study.

Two other signals were simultaneously acquired. These include ECG (in the lead two arrangement, Model: IX-B3G, iWorx Systems, Inc., Dover, NH) and respiratory flow rate (via a mouthpiece using a spirometer, Model: A-FH-300, iWorx Systems, Inc., Dover, NH). A sampling rate of 10 kHz was used for data acquisition.

Subjects rested comfortably on 45-degree inclined bed during data collection. A diagram of the experimental setup is shown in Fig. 4.

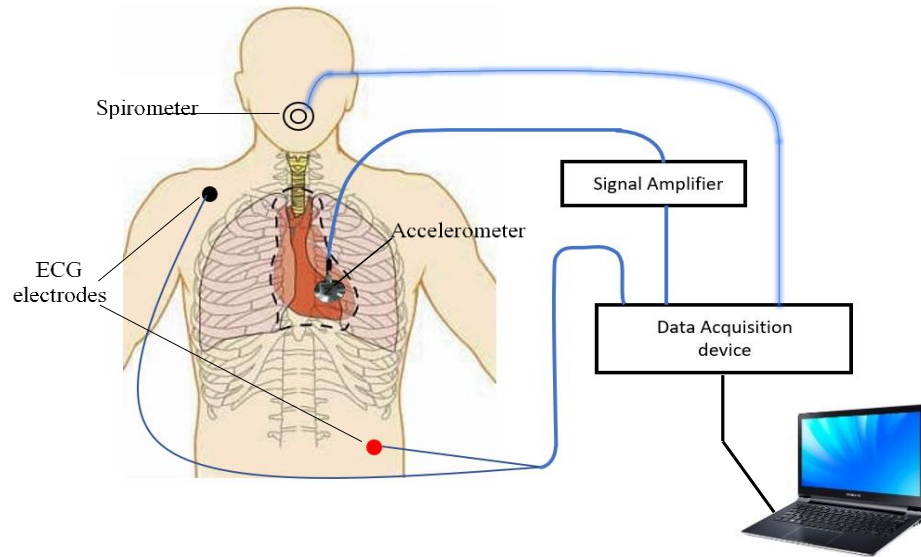


Figure 4- Experimental setup

Preprocessing

Filtering

The code for all signal processing steps was written in MATLAB (2017b, The MathWorks, Inc., MA). SCG and ECG signals were forward-backward filtered using a 4th order Chebyshev 2 type band-pass filter (0.5-50 Hz) to reduce the background noise and baseline wondering (i.e., variation) due to respiration. In addition, a moving average filter of order 5 (low-pass with cut-off ~ 2kHz) was employed to further smooth the signal. For each subject, the original and the filtered signals were compared in time and frequency domains to make sure filtering had minimal distortion on the SCG event amplitudes. Fig. 5 shows an example of the original and filtered SCG data in the time and frequency domains.

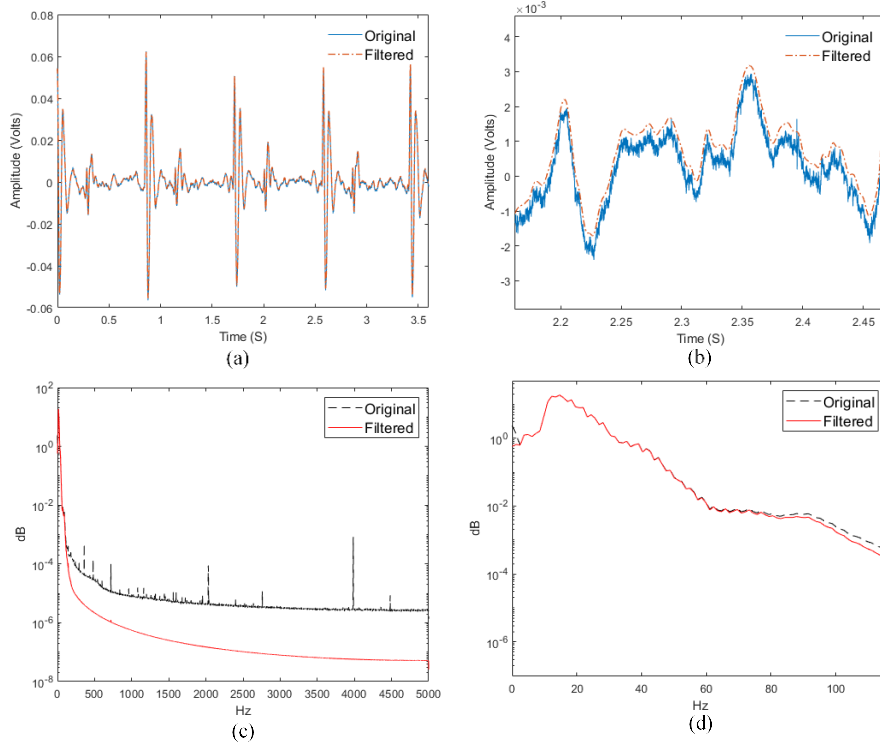


Figure 5 – Original and filtered SCG signal in (a) time domain (b) zoomed in time domain (c) frequency domain (d) zoomed in frequency domain

SCG segmentation

The SCG signal was segmented into SCG beats (also called events in this manuscript) using the R peaks of the ECG signal, which were detected using Pan Tomkins algorithm (Tompkins, 1985). Each SCG beat was selected to start 0.1 seconds before the R peak of the corresponding ECG, while the end point of SCG beat was selected 0.1 seconds before the R peak of the following ECG complex. Since the R-R interval varies over time, this approach resulted in SCG beats with varying duration, which is different from SCG studies that fix the duration of SCG beats (Sørensen et al., 2018; Amirtaha Taebi, 2018; Zakeri et al., 2017). Fig. 6 shows an example of the segmented SCG signal.

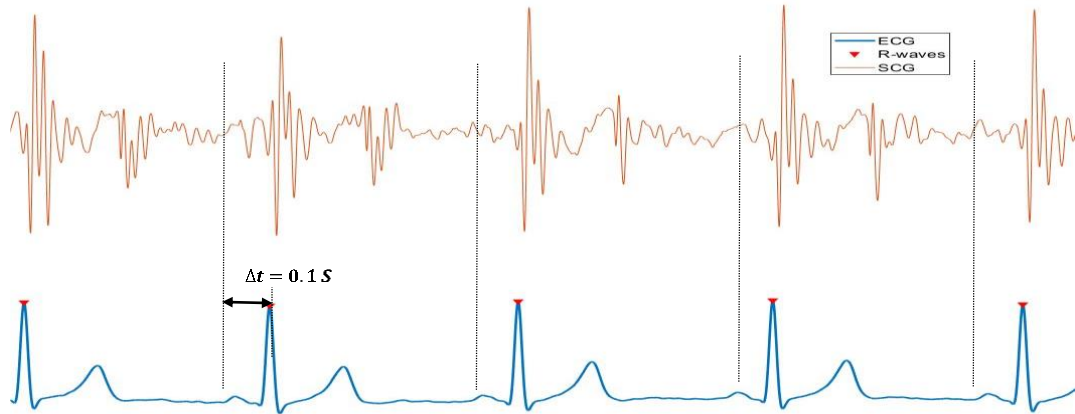


Figure 6 – Segmentation of the SCG signal using ECG beats

Unsupervised machine learning

Clustering SCG morphology

The morphology of a SCG beat may be best described by the signal amplitudes (at each data point of the beat) and the dissimilarity between the SCG morphologies can be quantified by calculating the differences between the signal amplitudes (after signal alignment). Hence, the amplitude values of SCG beats are an appropriate first choice as the input feature vector for the clustering algorithm, which is also called raw-based method in time series clustering (Paparrizos & Gravano, 2017).

The clustering algorithm will separate SCG beats into clusters by measuring the distance (i.e., dissimilarity) between the respective feature vectors using a distance measure. Traditional algorithms typically utilize Euclidean distance in these calculations (Fig. 9 (a)). When clustering, SCG beats need to be accurately aligned in time, otherwise errors in distance measurements will increase and beats with similar morphologies may be assigned to a different cluster. However, alignment of SCG beats become more complicated when signals have inherent morphologic variability. For example, SCG beat morphology may non-linearly stretch or compress due to heart rate

variability. This phenomenon can be seen in consecutive beats shown in Fig. 7(a) and (b). Even if a fiducial point is selected for the alignment of SCG beats, changes in the heart rate and associated intervals would cause misalignment of other fiducial points, which would lead to overestimating dissimilarity calculated by the Euclidean distance. Furthermore, dissimilarity calculated using the classic Euclidean distance assumes equal length of the SCG beats. However, beats have different lengths due to heart rate variability, and longer beats should be trimmed or compressed (to the length of the shortest beats, which corresponds to the highest heart rate) to have a constant length among all SCG beats. This process may remove valuable information contained in the part of beat that is removed.

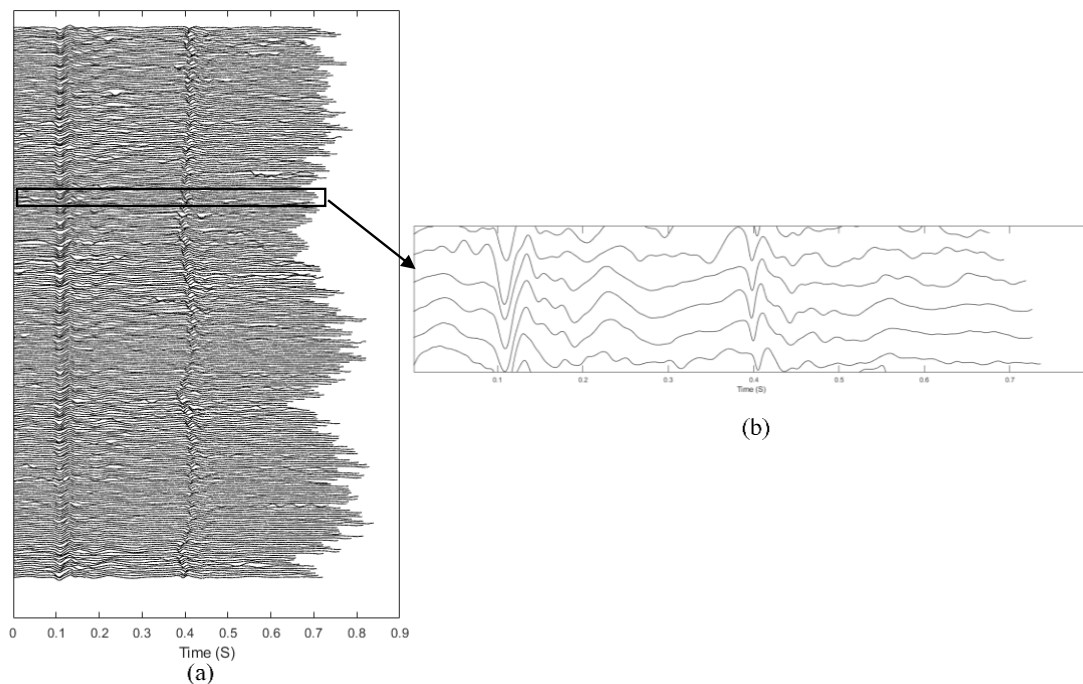


Figure 7-(a) Variations of the length of segmented SCG events due to heart rate (b) Zoomed view

Dynamic Time Warping (DTW)

DTW is a widely used measure of the similarity between two time series. It was originally designed for automatic speech recognition (Sakoe & Chiba, 1978) to identify the same word spoken at different speeds. DTW determines the optimal “global alignment” between two-time sequences by exploiting the temporal distortions between them (Sakoe & Chiba, 1978; Silva & Batista, 2016). DTW non-linearly “warps” the two sequences in the time domain to determine a measure of their similarity (Sakoe & Chiba, 1978). This dissimilarity measure is often used in time series clustering (Paparrizos & Gravano, 2017). The steps for calculating the DTW distance between two time series with different lengths, X and Y , are as follows

$$X = \{x_1, x_2, \dots, x_i, \dots, x_n\} \quad (1)$$

$$Y = \{y_1, y_2, \dots, y_j, \dots, y_m\} \quad (2)$$

where n and m are the lengths of the two signals. A “distance matrix” for X and Y is then generated as shown in Fig. 8.

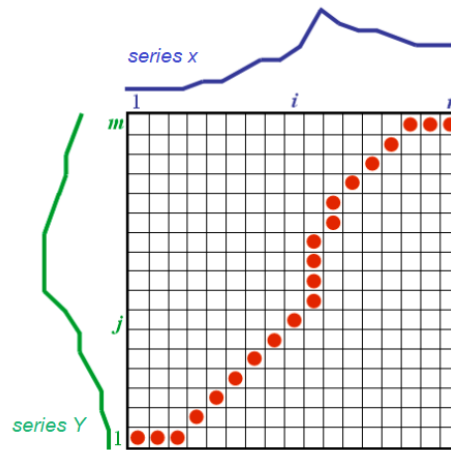


Figure 8- Distance matrix and the optimum warping path for signals X and Y

This distance matrix is recursively filled using following formula,

$$D(i, j) = \delta(x_i, y_j) + \min \begin{cases} D(i, j-1) \\ D(i-1, j) \\ D(i-1, j-1) \end{cases} \quad \text{where } \delta(x_i, y_j) = (x_i - y_j)^2 \text{ or } |x_i - y_j| \quad (3)$$

An optimal alignment (warping path) $W = \{w_1, w_2, \dots, w_k, \dots, w_N\}$ is to be found where $w_k = (i, j)$ represent the alignment between i^{th} point of X and j^{th} point of Y .

The optimal warping path is found such that it minimizes,

$$DTW(X, Y) = \operatorname{argmin} \sum_{k=1}^{k=N} D(w) \quad (4)$$

where, the warping path should satisfy the following three conditions (Zhang, Tang, Huo, & Zhou, 2014).

Boundary constraint: $w_1 = (1, 1), w_N = (n, m)$

Monotonicity constraint: $w_k = (i, j), w_{k+1} = (i', j')$ where $i' \geq i$ and $j' \geq j$

Continuity constraint: $w_k = (i, j), w_{k+1} = (i', j')$ where $i' \leq i + 1$ and $j' \leq j + 1$

The computed $DTW(X, Y)$ reflects the dissimilarity between X and Y . Fig. 9 shows the difference between using Euclidean distance and DTW as a dissimilarity measure. As can be seen in Fig.9, associated points are concurrent for Euclidean distance. In DTW, associated points are related non-linearly in time.

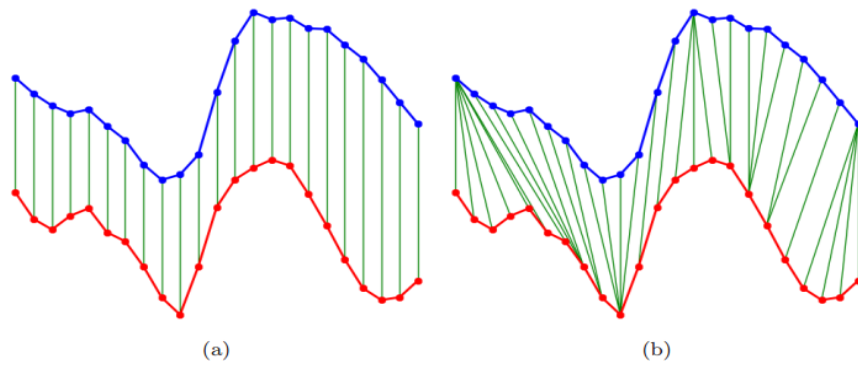


Figure 9- Associated points between signals X and Y when the dissimilarity is measured with (a) Euclidean and (b) DTW measures (Zhang et al., 2014)

Hence, for accurate clustering of SCG morphology a distance measure that accounts for aforementioned distortions in SCG beats is needed. Previous studies of clustering time series data (based on their temporal morphology) have encountered similar issues and proposed the use of dynamic time warping (DTW), which is a similarity measure that delivers more superior accuracy than Euclidean distance (Paparrizos & Gravano, 2015, 2017).

Averaging SCG beats

Averaging of SCG beats is performed in order to find a representative beat from each set of similar SCG beats. The representative beat is then used to derive features (e.g. determine fiducial points) for diagnostic purposes. Hence, the SCG beat should accurately represent the morphology of the set of SCG beats. In previous studies, fixed length SCG beats were averaged after aligning relative to a certain peak (such as R peak of ECG, maximum peak of the systolic portion of SCG, S1 or S2 peaks of the phonocardiography (PCG)) (Peshala T Gamage, Azad, Taebi, Sandler, & Mansy, 2018; Sørensen et al., 2018; Amirtaha Taebi, 2018). However, calculating an accurate average while conserving the original morphology (of the non-linearly stretched or truncated SCG beats) is not an easy task. For traditional averaging methods in Euclidean space, morphological features are more conserved close to the aligned location while features may be lost in other regions of the beat. For example, if the SCG beats are aligned with respect to R-peaks, morphological features may be conserved in the systolic region close to the R-peak while beat morphology may get smeared in the diastolic region due to decreased alignment accuracy. To overcome this issue, a recent study calculated two different averages by aligning the signals relative to R peaks in ECG and S2 peaks of

PCG and analyzed the fiducial points in systolic and diastolic regions, separately (Sørensen et al., 2018). Furthermore, if the average is calculated between SCG beats with different morphologies, the average beat may have a different morphology that may mask some important diagnostic features that are contained in original SCG beats.

In the current study, we propose the use of a shape-based averaging technique after SCG beats are optimally clustered into different groups. When employing DTW, several shape-based averaging methods were suggested in literature including: Nonlinear Alignment and Averaging Filters (NLAAF), Prioritized Shape Averaging (PSA) and DTW barycenter averaging (DBA) (Petitjean et al., 2014; Petitjean, Ketterlin, & Gançarski, 2011). The first two methods use a pairwise averaging strategy that suffer from the growth of the length of the resulting average (to almost double) in each step and the process of reducing the length may lead to loss of information (Petitjean et al., 2014). The recently introduced DBA method eliminates these drawbacks and appears most accurate and efficient (Paparrizos & Gravano, 2017). Therefore, DBA will be employed in this study.

DTW Barycenter Averaging (DBA)

DBA tries to find an optimum average for a set of time sequences (e.g., SCG beats) in DTW space. This average is such that it has minimum DTW distance from the set of sequences. The method starts by selecting an arbitrary average and iteratively updating the average to minimize the sum (DTW) of distances from the set of sequences to the sequence average.

When calculating the DTW distance between the average beat and the set beats, many points of the set of beats may be associated with the same point of the average

beat (and vice versa) as seen in Fig. 9. The higher the number of these points, the higher the DTW distance between the set of beats and the average beat (because these distances are added). To reduce the associated number of points, DBA suggests updating the point of the average beat by taking the barycenter average (barycenter is the center of mass of the points, assuming equal mass for the points) of the associated points. Iteratively, the same procedure is followed with the updated average till the sum of the DTW distances between the average and the set of beats converges within a predefined tolerance (Petitjean et al., 2014). Fig. 10 shows a DBA average calculated for a set of SCG beats.

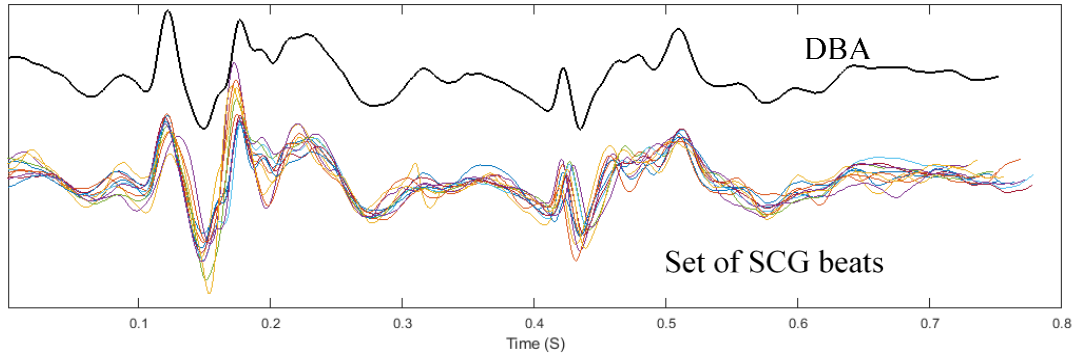


Figure 10- DBA average calculated for a set of SCG beats

However, DBA has a time complexity of $\theta(I.N.l^2)$, where I, N and l denote the number of iterations, number of sequences and the length of the sequences, respectively (Petitjean et al., 2014). In addition, in some cases DBA may not converge to a smooth signal (which was observed when DBA was calculated at sampling rate of 500 Hz in our study) due to non-linear distortions in DTW calculations (Petitjean et al., 2014). To avoid such limitations, the medoid can be also used as a possible alternative representative of the cluster the morphology.

Clustering algorithms

Unsupervised machine learning is employed in the current study to cluster SCG morphology into clusters with highest intra-cluster similarity and highest inter-cluster dissimilarity. In time series clustering, three main types of clustering methods are used, namely, hierarchical, spectral and partitional clustering (Paparrizos & Gravano, 2015).

Hierarchical clustering generates a cluster hierarchy by (a) combining most similar clusters pairwise (starting from the individual sequences) till all sequences (i.e., SCG beats) are merged in to a single cluster (agglomerative clustering) or (b) dividing the clusters in to pairs until each leaf cluster contains only one object (divisive clustering). However, hierarchical clustering performs clustering in a local level and no global objective function is directly minimized as in “partitional” based methods (discussed below). In addition, hierarchical clustering is very sensitive to outliers. Furthermore, during the clustering process, once two clusters are merged/split they cannot be undone in a later step for a more suitable merge or split (Rokach & Maimon, 2005). Hence, hierarchical clustering may yield a suboptimal solution and the dendrogram may not necessarily represent the natural clusters in the data.

Spectral based clustering is typically used in graphical applications such as image processing, computer vision and computational biology. This clustering is a graph-based approach where the sequences are treated as nodes in a graph that are mapped to a low dimensional space using the spectrum (Eigen values) of the input data similarity matrix (Von Luxburg, 2007; Zelnik-Manor & Perona, 2005). Spectral based clustering is very sensitive to the initial conditions. In addition, it's sensitive to the similarity parameters used to define the “connectedness” in the similarity graph which

are used in the Gaussian similarity function (Von Luxburg, 2007). Also, this method is computationally expensive for large datasets.

In partitional clustering, each cluster is represented by the center of the cluster (e.g. k-means) or the member of the cluster with the minimum average dissimilarity to all other members (e.g. k-medoids). These algorithms work on assigning the sequences to the closest cluster representative (means or medoids), then updating the mean or medoid. Ideally, this process is repeated till no changes in cluster assignments are observed. The convergence of partitional clustering is usually monitored by the “sum of distances” (SOD), which is the total summation of the distances between each observation to its cluster medoid/center (see Eqn. 8). Partitional clustering may converge to a local minimum since the algorithms heavily depend on the initial conditions (e.g., initial medoids). However, this issue can be controlled by appropriately tuning the initial conditions by monitoring the convergence criteria for several randomly selected initial conditions where certain initial conditions may converge to a higher SOD values (local minima) and other will converge to a low SOD value (global minima). Furthermore, by monitoring the number of iterations for convergence, a good initial condition can be selected. In addition, the partitional clustering will yield a simple representation of the clustered results since each cluster is represented by the medoids or average. This study will use partitional clustering to cluster the SCG morphology.

Partitional based methods have been extensively used in shape-based time series clustering with DTW as a distance measure where generally k-medoid method is preferred over k-means to avoid the effect of outliers. In addition, the k-means may require more computations since it requires computing an artificial average sequence

as the centroid for each iteration (Liao, 2005). The centroid is then computed by a shape-based averaging method such as DBA, which requires additional computational expense. In contrast, k-medoid will determine the medoid sequence, which can be easily implemented with DTW measure. A recent study (Paparrizos & Gravano, 2017) reported higher accuracy for k-medoid clustering when compared with other clustering methods for shape-based clustering of time series. The details of proposed algorithm are below.

Inputs: Number of clusters= K . Set of sequences: $S = \{X_1, X_2, X_3, \dots, X_i, \dots, X_N\}$ where each sequence is defined by its feature vector (amplitude) as $X_i =$

$$\{x_1, x_2, x_3, \dots, x_{l_i}\}$$

Step 1: Initialize $C_1, \dots, C_j, \dots, C_k$ as the medoids for each cluster

Step 2: For each X_i find the nearest C_j using DTW as the distance measure and assign X_i to cluster j

Step 3: Update C_j based on the clustered events from step 2 using Eqn. 7.

$$C_j = \underset{y \in \{X_{1j}, X_{2j}, \dots, X_{ij}, \dots, X_{nj}\}}{\operatorname{argmin}} \sum_{i=1}^{n_j} \operatorname{dtw}(y, X_{ij}) \quad (7)$$

where, X_{ij} is the i^{th} sequence belongs to cluster j and n_j is the number of sequences belong to C_j after step 2.

Step 4: Repeat step 2 and 3 till none of the cluster assignments change.

The time complexity of DTW is $\theta(l^2)$ where l denotes the length of a sequence and when DTW is calculated N times, the complexity becomes $\theta(N.l^2)$ (Petitjean et al., 2014).

Hence, to reduce the time complexity of clustering in the current study, SCG beats were down sampled to 500 Hz (after filtering and segmentation). The SCG beats

were also normalized by their maximum amplitudes. This is not expected to affect DTW measure of similarity.

Results and Discussion

In this section we discuss the clustering results in relation to respiratory phases and heart rate. Cluster switching timing and the variability are also discussed.

Optimum number of clusters

Optimum number of clusters were decided using the elbow method and analysis of the average silhouette value. Elbow method is commonly used for determination of the fewest number of clusters that optimizes intra-cluster variance. The variance (or heterogeneity) of the clustering is often measured by calculating the sum of distances (SOD) between the observation points (i.e., SCG beats) and their cluster medoids using Eqn. 8.

$$SOD = \frac{1}{N} \sum_{j=1}^k \sum_{i=1}^{n_j} dtw(C_j, X_{ij}) \quad (8)$$

where, X_{ij} is the i^{th} sequence belongs to cluster medoid C_j and n_j is the number of sequences belong to C_j . N is the total number of sequences used for clustering. With the increase number of clusters, the observation points get closer to their centroids and the SOD will decrease (e.g., for N clusters, SOD reaches zero). When SOD is plotted against the number of clusters, SOD will decline rapidly, and then at a slower rate creating an elbow shape in the plot. The number of clusters at the elbow point can then be selected as the optimum number of clusters. Fig. 11 shows the mean SOD of 17 subjects for different number of clusters. The elbow shape can be seen around a cluster number of 2.

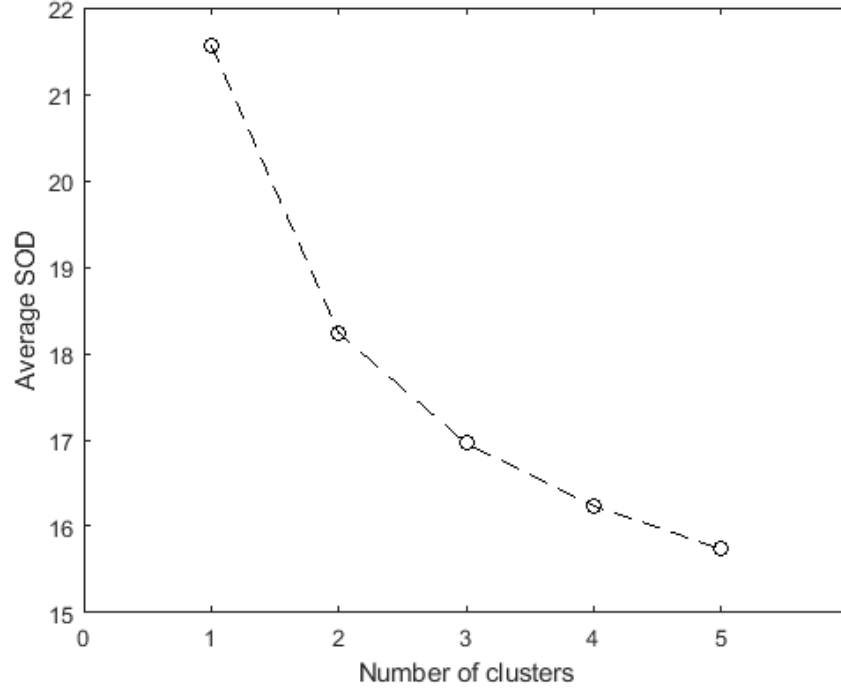


Figure 11- Average SOD for different number of clusters. Two clusters were chosen using the elbow method

The average Silhouette value (Si) of the clustered data is another method to determine the optimum number of clusters. The Si value of an observation point (i.e., SCG beat) in the cluster is a relative measure of how well that observation point is placed within its own cluster. The Silhouette value (Si) of the point i in the clustering can be expressed as,

$$Si = (b_i - a_i) / \max(a_i, b_i) \quad (9)$$

where a_i denote the average distance measured from the i th point to the other points in the same cluster (that point i belongs to) and b_i denotes the average of the minimum distances measured from i th point to all other points in other clusters. Si values range from -1 to 1 where a positive Si value closer to 1 indicates a point is well

inside in its own cluster (away from the boundaries of other clusters) and a negative Si value closer to -1 indicates that the point is closer (to the boundaries) of other clusters.

Fig. 12 shows the average Si values calculated for 17 number of subjects. The results indicated that the highest average Si value was observed when the data are clustered into two groups. With the increase number of clusters, the average Si value decreased. Based on the results from the elbow method and average Si value, 2 clusters were selected for the study.

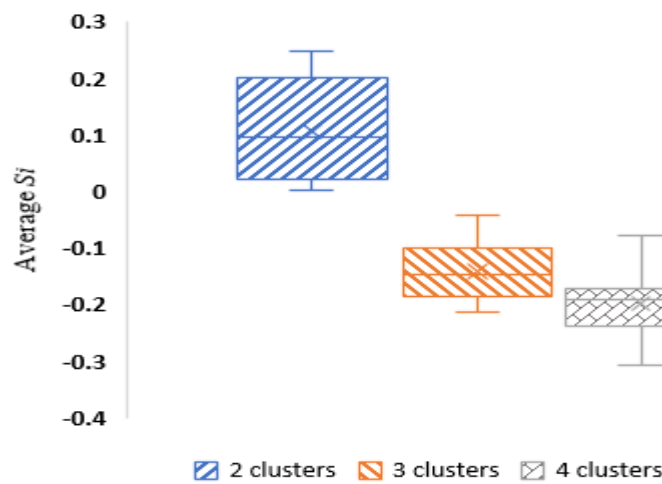


Figure 12- Box-whiskers plot of average Silhouette value (calculated for all 17 subjects) for different number of clusters

Purity of clustering with labels HLV/LLV and INS/EXP

Some previous studies categorized SCG signal into respiratory phases to minimize the variability of SCG beats prior to feature detection. While some studies have chosen to group SCG beats based on respiratory flow direction (inhale /exhale) (Amirtaha Taebi & Mansy, 2017a; Amirtaha Taebi et al., 2018), others grouped SCG beats according to the lung volume (Peshala T Gamage et al., 2018; Amirtaha Taebi, 2018) or to the movement of the chest measured using a chest belt (Zakeri et al., 2017). To study the efficiency of these grouping criteria, purity values of the clustered data

were calculated for two labeling criteria. The labels were named inspiration/expiration (INS/EXP) and high/low lung volume (HLV/LLV). The positive/negative values of the flowrate signal (measured using a spirometer) indicated inspiration/expiration, respectively. The flowrate signal was integrated to get the lung volume signal and the positive/negative values of that signal were indicative of high /low lung volume, respectively. SCG beats were labeled based on the timing of the corresponding R peak on the respiratory signals as shown in Fig. 13.

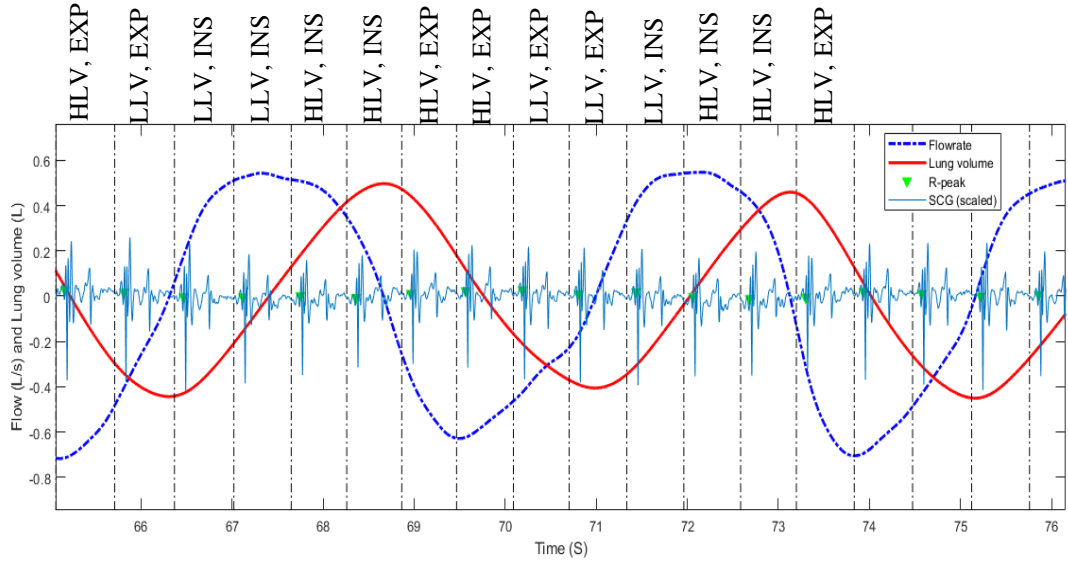


Figure 13- Labeling SCG beats, HLV=High lung volume, LLV=Low lung volume, EXP=expiration, INSP=Inspiration, Red trace=Lung volume, Blue trace=Respiratory flow rate

The purity value indicates how well the labelling criteria fits with the cluster result and it is defined as,

$$Purity = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

where, TP is the number of true positives and TN is the number of true negatives. FP and FN indicate the number of false positives and false negatives, respectively. For example, if the labeling criteria is INS/EXP and SCG beats are divided into two clusters, TP indicates the number of correctly labeled SCG beats as INS and TN

indicates number of correctly labeled SCG beats as EXP. Similarly, FP and FN are the number of incorrectly labeled beats as INS and EXP, respectively.

Fig. 14 shows the purity values for 17 subjects. The mean purity value for INS/EXP and HLV/LLV were 0.66 and 0.77, respectively. The higher purity for HLV/LLV classification suggests that the clustering using HLV/LLV criteria would provide better separation of SCG beats than INS/EXP criteria (except for subject 5 and 8).

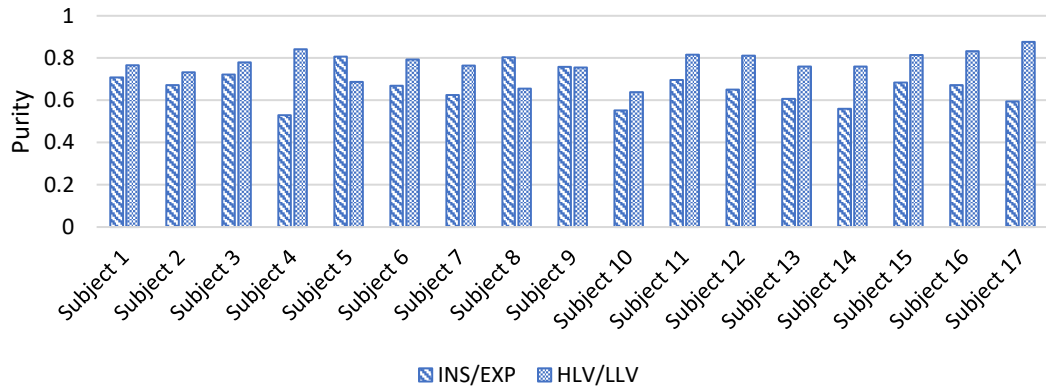


Figure 14 -Purity values for two labelling criteria; INS/EXP and HLV/LLV. HLV/LLV labeling provided higher purity levels.

Further, the intra-group variance (as an indicative of group heterogeneity) before clustering was compared with flowrate direction (i.e., INS/EXP), lung volume (i.e., HLV/LLV), and k-medoid clustering. Here, SOD (Eqn. 8) was used as a measure of the variance (Paparrizos & Gravano, 2015). The mean SOD were 21.57 before clustering and 20.93, 20.22 and 18.24 for INS/EXP, HLV/LLV, and k-medoid, respectively. It can then be concluded that the highest variance existed before clustering and decreased by about 15% when k-medoid clustering was performed (Fig. 15). The variance for the INS/EXP was slightly higher than HLV/LLV grouping, suggesting that HLV/LLV grouping may provide better homogeneity than INS/EXP. Fig. 15 shows higher

heterogeneity for the INSP/EXP compared with HLV/LLV grouping in most subjects (Fig. 15).

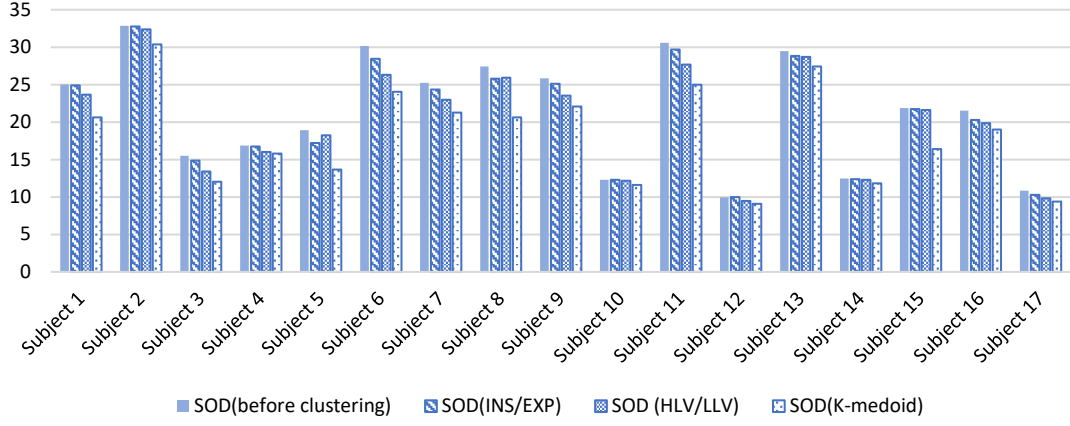


Figure 15- SOD values before clustering and when the clusters were separated based on HLV/LLV, INS/EXP and k-medoid solution. K-medoid had least SOD in all subjects.

Analyzing cluster distribution with respiratory phases

To analyze the possible relations between the cluster distribution and the respiratory cycle, the timing of clustered SCG beats was located on the respiratory flowrate and lung volume waveforms. Here, the respiratory waveforms were normalized according to Eqn. 11.

$$F_{norm} = \frac{F}{F_{max} - F_{min}} \quad (11)$$

where, F is the respiratory waveform and F_{max}, F_{min} are the maximum and minimum values of F . The locations of SCG beats are shown in Fig 16, where beats belongs to cluster 1 and cluster 2 are labeled as blue ‘o’ circles and red ‘v’ triangles, respectively. The shown locations of the SCG beats coincide with the timing of their respective R peaks. Fig. 16 (a) shows the cluster distribution (of one subject) on the normalized flowrate waveform and Fig. 16 (b) shows the cluster distribution on lung

volume waveform. Fig. 16 (c) shows the cluster distribution with both flowrate and lung volume in 3D.

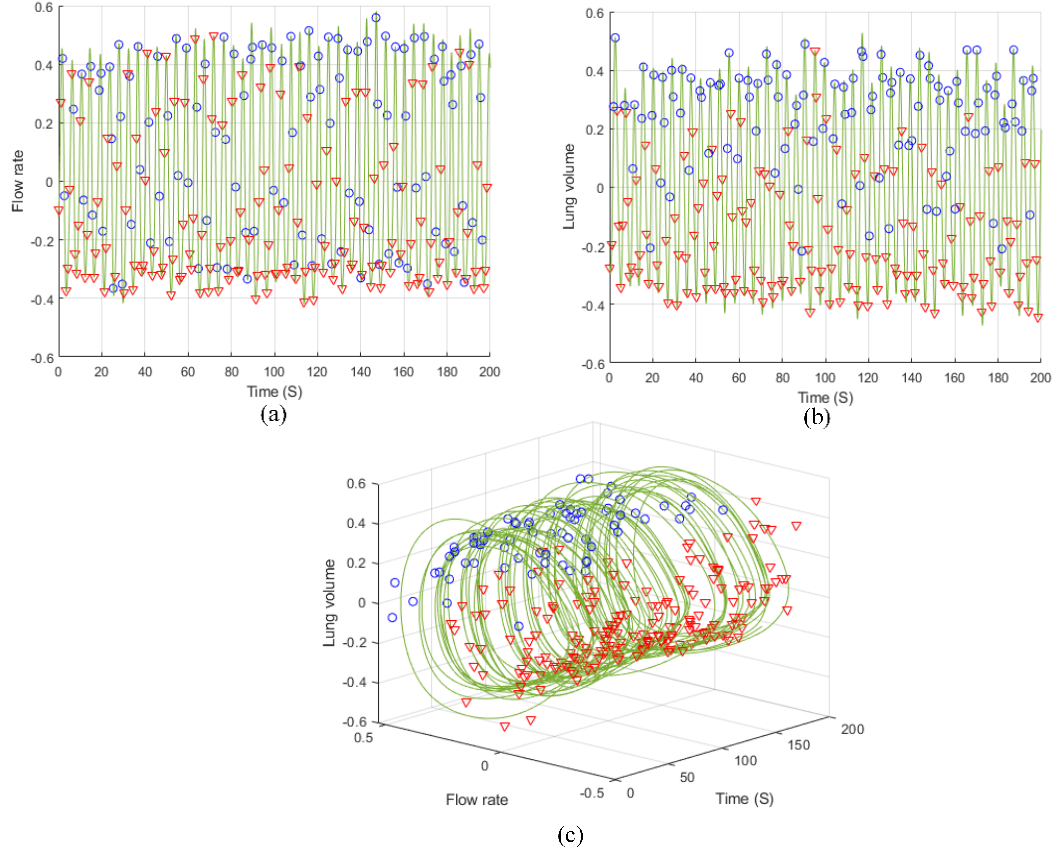


Figure 16- Cluster results (event locations) plotted on (a) Respiratory flowrate cycle (b) Lung volume cycle (c) both flowrate and lung volume cycles in 3D

For a better analysis of the cluster distribution, the respiratory phases were divided into four groups, namely: HLV-INS, HLV-EXP, LLV-EXP and LLV-INS as shown in Fig. 17 (a). A recent study (Peshala T Gamage et al., 2018) used a variance minimization approach to analyze SCG clustering and observed similar trends. The cluster distribution (of one subject) with corresponding respiratory flowrate and lung volume is presented in Fig. 17 (b) and the corresponding four respiratory phases are shown in Fig. 17 (a).

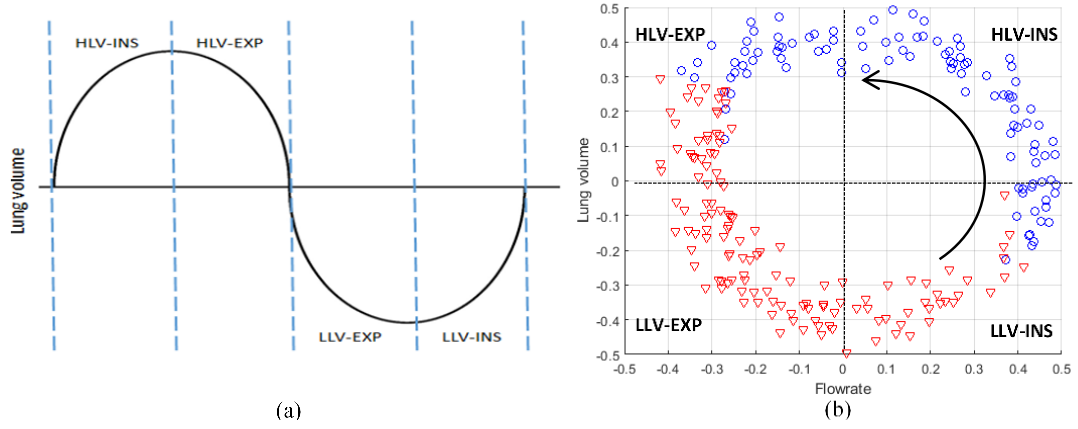


Figure 17-(a) Four respiratory phases identified in a simplified lung volume waveform (b) Cluster distribution in lung volume and flowrate space in a typical subject

As shown in Fig. 17 (b), the clusters are well separated in LLV-EXP and HLV-INS regions while clusters are mixed in LLV-INS and LLV-EXP regions where the cluster switching occurs. This trend was consistent among all subjects.

Cluster switching

The cut-off timing when SCG beat switches between the two clusters was determined by employing a linear support vector machine (SVM) theory (Cortes & Vapnik, 1995). Here, the SVM method was applied on a two-dimensional space using the values of respiratory flowrate and lung volume at the time of the R peak of SCG beats. The linear SVM method finds a hyperplane (i.e., a decision boundary) in the feature space of flowrate-lung volume such that the margin between two clusters is maximized. Fig. 18 shows the decision boundary plotted on flowrate vs. lung volume for one subject. The equation of the decision boundary is a linear function of flowrate and lung volume, which we define as the cluster cut-off equation.

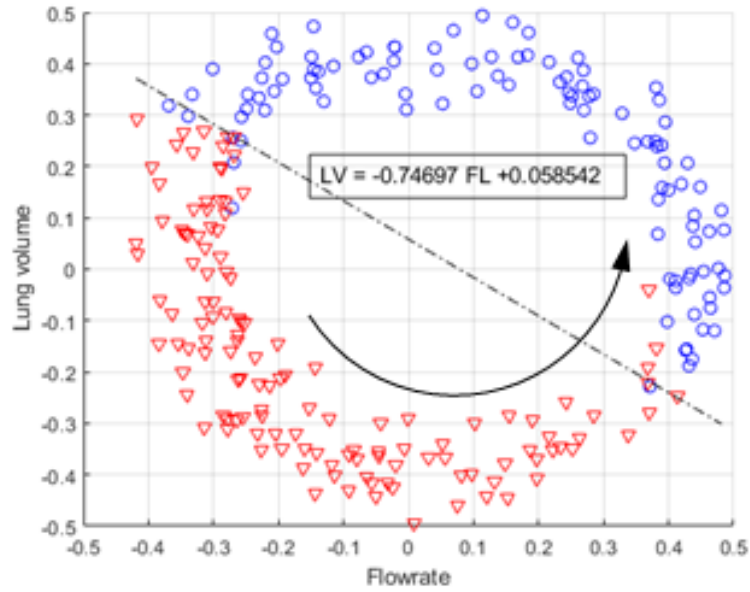
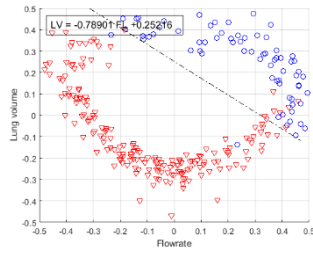
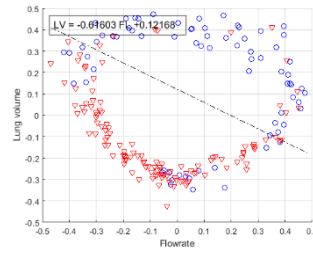


Figure 18: The cut-off function between the two clusters calculated from linear SVM theory in a typical subject

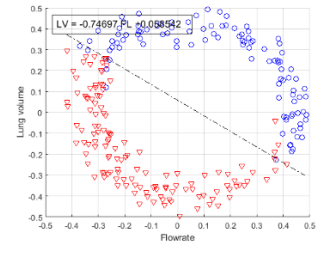
Similarly, cluster cutoff equations were calculated for all the subjects and the results are shown in Fig. 19. These linear equations (in the form $LV=m \times FL+c$) had a positive slope ' m ' of -0.63 ± 0.39 (mean \pm std) and an intercept ' c ' of 0.05 ± 0.05 (mean \pm std).



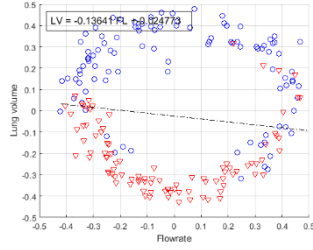
S1



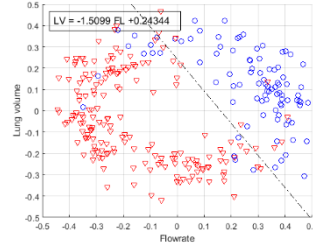
S2



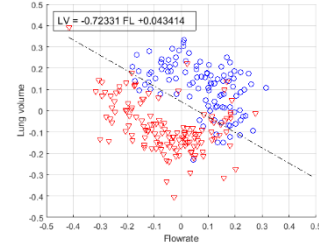
S3



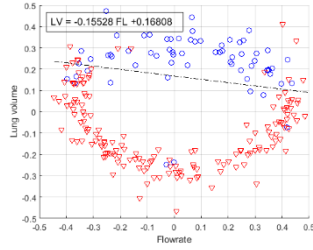
S4



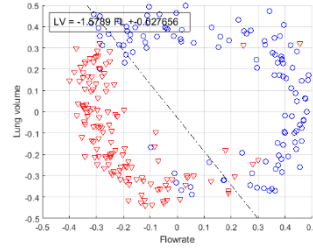
S5



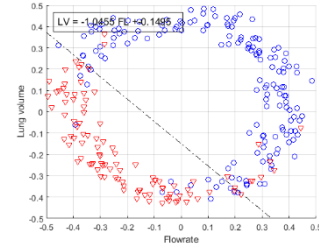
S6



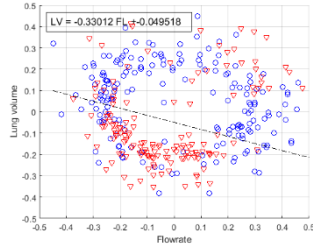
S7



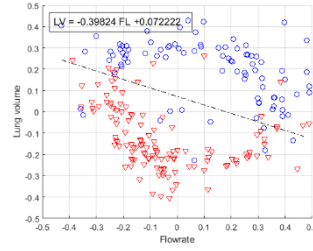
S8



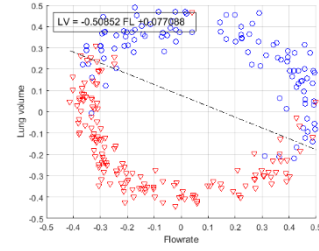
S9



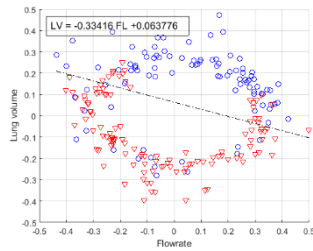
S10



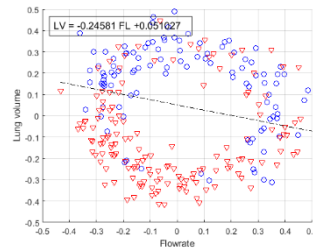
S11



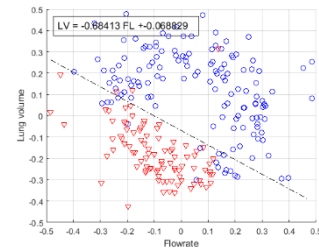
S12



S13



S14



S15

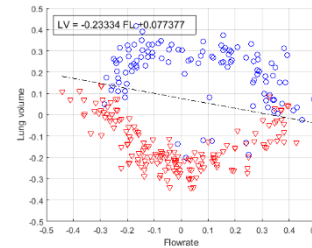
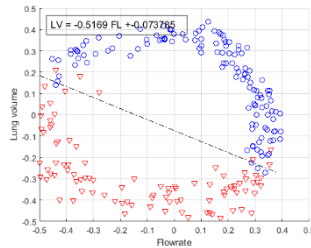


Figure 19- Cluster distributions and the cluster cut-off equations plotted in flowrate vs. lung volume space for all study subjects

Relation between heart rate and clustering

To investigate the relation between heart rate and clustering, cluster distribution was plotted with the heart rate variation as function of breathing as shown in Fig. 20.

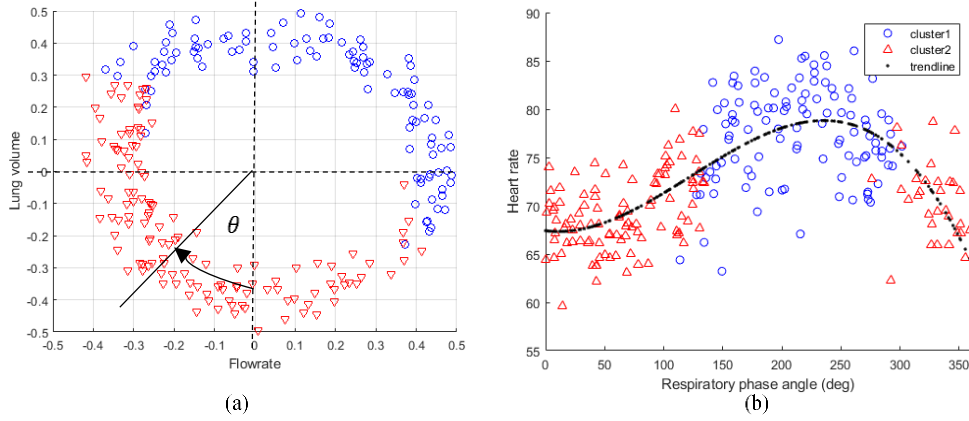


Figure 20- (a) Defining respiratory phase angle (b) cluster distribution with heart rate as a function of respiratory phase angle (for one subject)

As shown in Fig. 20 (b), the trend line of the heart rate showed patterns that are similar to previous studies that investigated the well-known phenomenon of respiratory sinus arrhythmia (RSA) (Angelone & Coulter JR, 1964). The data showed that the average heart rates for the two cluster were 76.95 and 73.36, respectively. The heart rate for the cluster containing HLV-INS phase was $6.03 \% \pm 3.32\%$ (*mean \pm std*), higher than the other cluster. This change in heart rate was significant ($p = 3.04 \times 10^{-7}$).

Intra-cluster variability

In k-medoid clustering, the distance between the observation points and the medoids are measured then observation points (i.e., SCG beats) are assigned to the clusters with the closest medoid. Hence, within a single cluster, some points are closer to the medoid while others are far from the medoid. To demonstrate SCG beat variation

within clusters, Fig. 21 shows SCG beats for cluster 1 and cluster 2 separately. Here, it can be seen that the clustered SCG beats contain dominant morphological features inherent to each cluster while having noticeable variability within the clusters.

By analyzing the distance from the cluster medoid to each point belonging to that cluster, the variance of the cluster can be quantified. In our study, such analysis will help detect outliers as well as to locate SCG beats with the most similar morphologies.

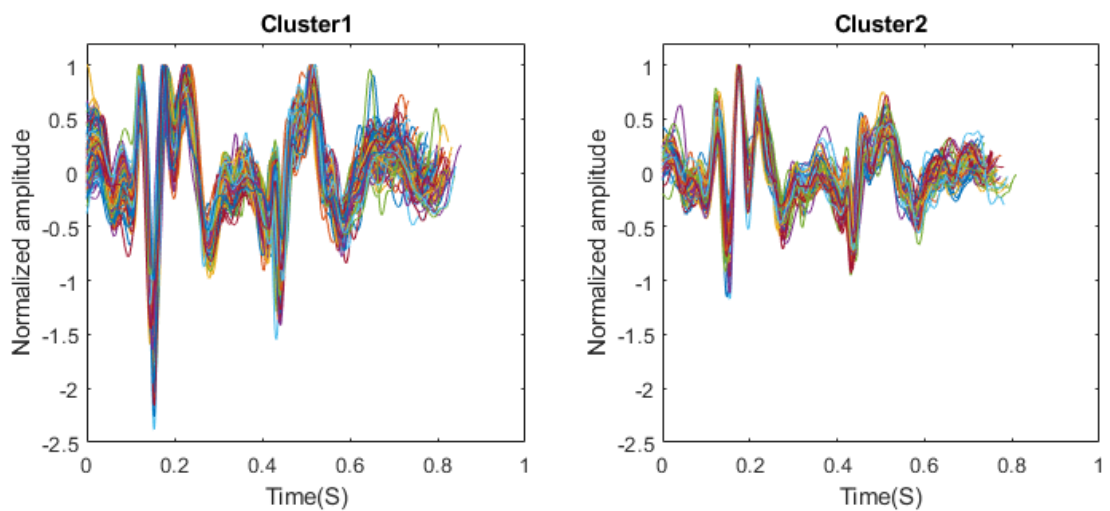


Figure 21- Clustered SCG beats plotted separately for cluster 1 and cluster 2 (for one subject)

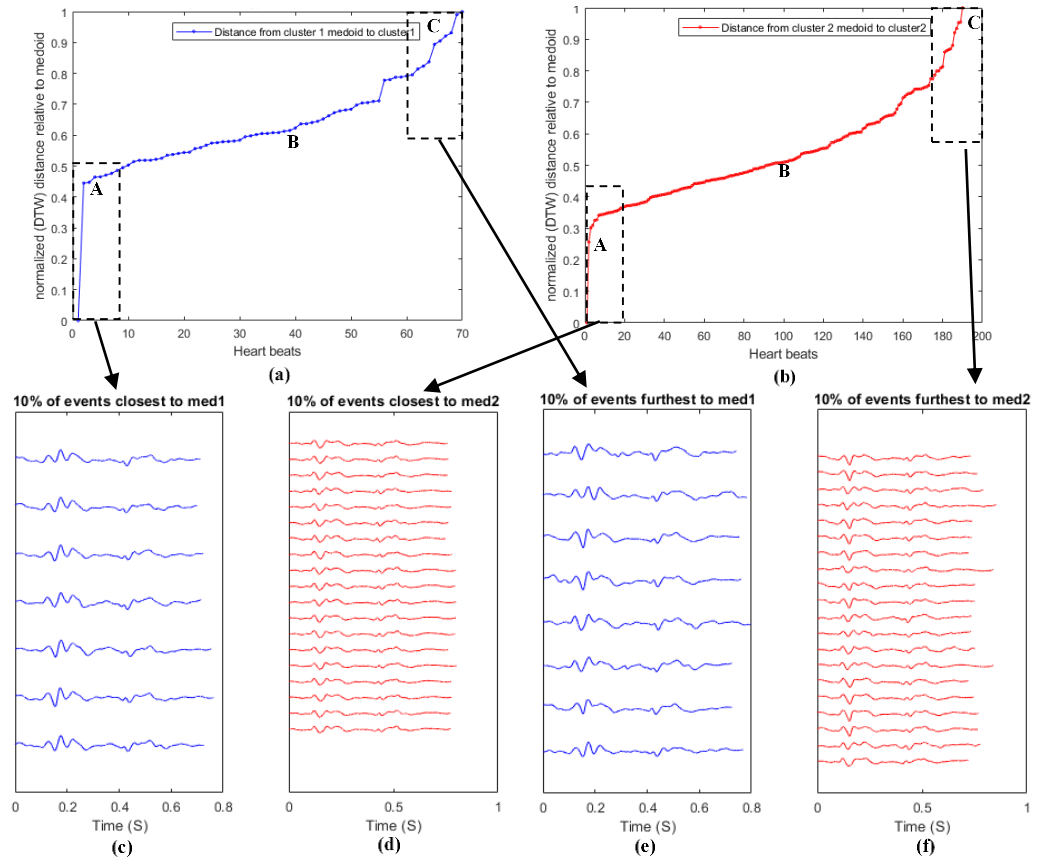


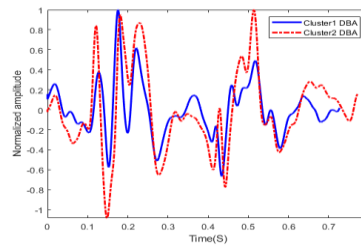
Figure 22- Distance plotted from cluster medoids to beats in the same cluster in ascending order for (a) cluster 1 (b) cluster 2. Closest (10%) of the beats to the medoid of (c) cluster 1 (d) cluster 2. Furthest (10%) of the beats from the medoid of (e) cluster 1 (f) cluster2

Fig. 22 (a) and (b) show a plot of the normalized distance distribution from the medoid to the beats in the same cluster. In Fig. 22, the distances are sorted from the closest to the farthest. The plot takes the form of a tangent like function, which was consistent among all subjects. The first point in the plot represents the medoid itself (distance is zero) followed by the point closest to the medoid (point A). As the points move away from the medoid, the distance gradually increases at an approximately constant rate (e.g., point B). As the distance further increases, a steeper distance increase is seen. This steep gradient appears to continue until (point C) where outliers may be found. This data suggests a relatively small number of outliers.

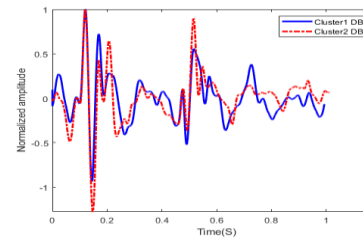
Fig. 22 (c) to (f) show the 10% of SCG beats that are closest and farthest from the medoids for both clusters. As expected, results show that the beats closest to the medoids have more homogenous morphology while beats farthest from the medoid appeared to have more heterogeneity.

In this study, we proposed and used DBA as an averaging method for calculating a representative event of the SCG beats that are closest to the medoid. Here, DBA was calculated for the 10% of the closest beats to the medoid (including the medoid) using the medoid as the initial average. Fig. 23 shows the calculated DBA averages (for a sampling frequency of 10kHz using 10 iterations) of the two clusters for all subjects. These results suggest significant inter-subject variability of SCG beats as well as varying effects of respiration. The noticeable differences between the beats in the two clusters suggest that separation of SCG beats in two cluster would provide a more precise estimate of SCG waveforms.

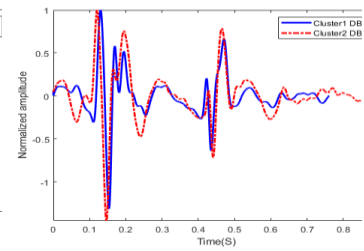
Although the number of participants was relatively small (a few hundred heart beats in each of the 17 participants, which is a limitation of the study), clustering results were consistent in all subjects and reached statistical significance. In addition, the clustering results were consistent with the findings of previous studies that showed similar dependence of SCG morphology on respiratory phases. To further confirm this result, future studies will consider larger number of subjects including those with cardiac conditions.



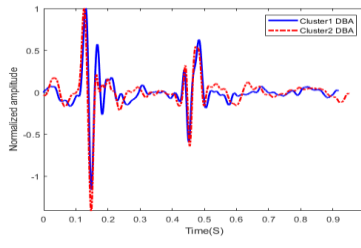
S1



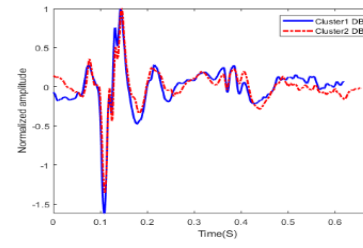
S2



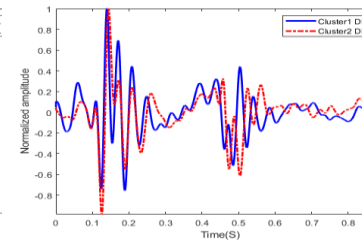
S3



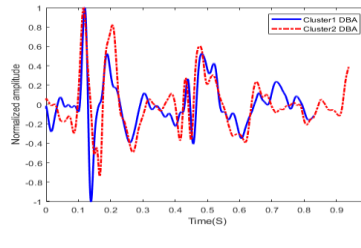
S4



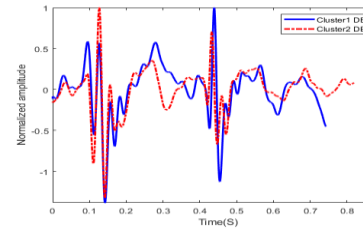
S5



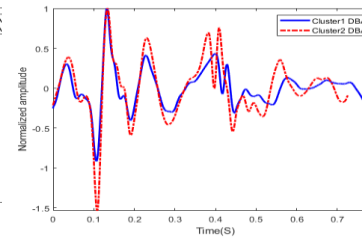
S6



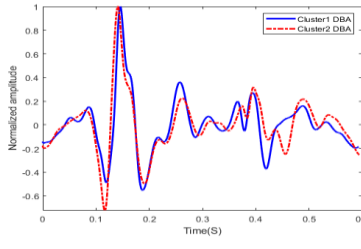
S7



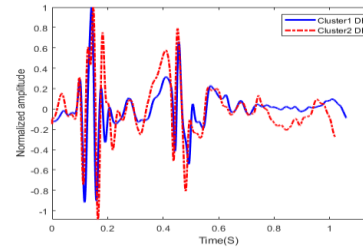
S8



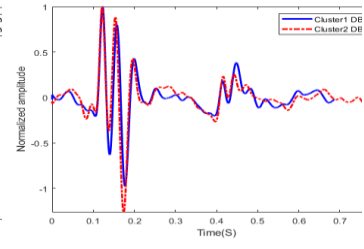
S9



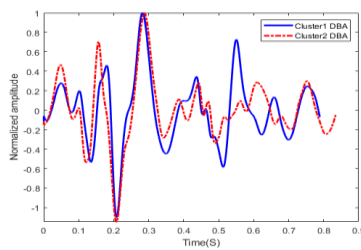
S10



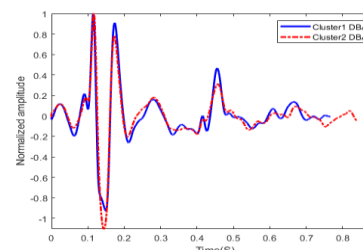
S11



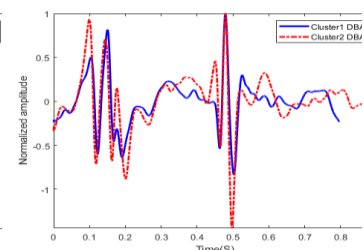
S12



S13



S14



S15

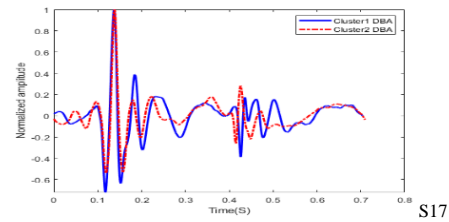
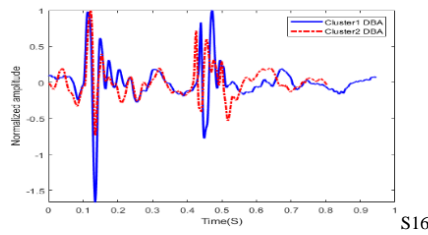


Figure 23-Calculated DBA for 2 clusters for all subjects

Conclusion

This study investigated the utility of unsupervised machine learning in clustering SCG beats to reduce their variability. Seventeen subjects participated in the study. k-medoid clustering was implemented and dynamic time warping was chosen as the dissimilarity measure. The study results showed that the SCG morphology can be optimally separated into two clusters based on the elbow method and the comparison of average silhouette values. The relation between clusters and respiratory phases was investigated. The clusters had better agreement with lung volume phases (i.e., high vs low lung volume) than the respiratory flowrate phases (i.e., inspiration vs. expiration). SCG switching from one cluster to the other consistently occurred during the first half of inspiration and expiration. The relation between SCG switching and heart rate and was also investigated. The average heart rate of the first cluster (containing inspiration and high lung volume) was significantly higher than the other cluster. This suggests that the mechanisms that cause respiratory sinus arrhythmia may be involved in SCG variability.

Waveform differences between clusters were noticeable and varied among subjects. The proposed clustering significantly ($p < 0.01$) decreased SCG variability (by about 15%). The reduced variability can provide more precise average waveforms and,

consequently, more accurate estimates of SCG features. This may yield stronger SCG correlation with heart function, which would enhance clinical utility of SCG. While several studies have shown SCG utility for monitoring cardiac pathology, more studies are actively investigating additional clinically relevant SCG features. If successful, SCG would provide an inexpensive portable noninvasive tool for telehealth and precision medicine applications. It may also provide useful information for big data approaches across volumes of health systems and monitoring data.

CHAPTER 3 – NUMERICAL MODELING OF SCG

Seismocardiography (SCG) is the measurement of the vibrations on the chest surface that are primarily produced by a combination of mechanical activities of the heart such as myocardial movements, valve closure and opening and blood momentum changes (Gurev et al., 2012; Korzeniowska-Kubacka et al., 2006). SCG is typically measured by placing an accelerometer upon the chest surface. Being inexpensive, non-invasive and compatible with telemedicine, SCG may offer significant potential for diagnosis and monitoring of various cardiac diseases SCG analysis of both time and frequency domain have been investigated for cardiac diagnosis (Amirtaha Taebi, 2018; Amirtaha Taebi et al., 2018), while other studies have utilized SCG features to estimate respiratory rate (Reinvuo et al., 2006) and monitor sleep apnea (Morillo et al., 2010).

The SCG signal as measured on the chest wall surface is believed to be a combination of complex 3D movements of the heart during its pumping activity. The complex nature of heart movements, plus vibrations induced by cardiac muscle activity, blood flow, valve openings and closings, and other factors have made it challenging to identify SCG morphology to vibration genesis. Some studies have used medical imaging to link different feature points of the SCG signal by relating their occurrence time with the corresponding cardiac events seen during cardiac imaging (Crow et al., 1994; Giorgis et al., 2008), yet, these findings remain inconclusive (Akhbardeh et al., 2009).

Cardiac movements displace their immediate boundaries (e.g., pericardium, Aorta wall) and surrounding tissues (e.g. lung tissue, ribs, chest muscle and skin) before they propagate to the chest surface. Hence, modeling the propagation of the overall cardiac wall motion to the chest surface may help enhance our understanding of SCG,

help perform parametric studies (e.g., on the effects of the tissue properties, etc.) and assist in the study of SCG genesis by relating feature points to cardiac wall movement.

A few studies have attempted computational modeling of SCG. For example, one study (Akhbardeh et al., 2009) used cine-MRI images to quantify the movements of the ventricles by segmenting the ventricles frame-by-frame and fitting a deformable mesh to the segmented volume. In that study, the tissue movement signal was derived by calculating the displacement of a point on the ventricle wall relative to an observation point. The same study (Akhbardeh et al., 2009) used an electro-mechanical model of a canine heart and reported a SCG-like acceleration at the center of mass of the ventricles. This electromechanical model (Akhbardeh et al., 2009) was implemented in another study (Tavakolian et al., 2012) where simplified cylindrical geometries were used to represent the sternum and internal organs and study their effects on the simulated SCG. These studies predicted SCG waveforms that agree with the general shape of the measured SCG waveform reported in the literature (Amirtahà Taebi et al., 2019).

The current study focuses on modeling the propagation of heart wall vibrations to the chest surface to produce a simulated SCG. Here, the 3D motion of the heart wall was tracked using an optical flow-based motion tracking algorithm on the short-axis images of cardiac cine MRI. The tracked heart wall movements were then used as boundary conditions in the Finite Element Method (FEM) computational model. The model geometry consisted of a region of the thorax that includes part of the lung, rib cage, inter coastal muscles (ICMs), and other chest wall musculature.

The simulated acceleration signals of the chest surface were validated by comparing them with previously described SCG morphological features (Crow et al.,

1994; J. Zanetti et al., 1991). The effect of the rib cage and chest muscle stiffness on SCG were investigated. The SCG distribution over the chest surface was calculated and compared with experimental measurements. In addition, the effect of heart displacement with respiration on SCG was studied by changing the relative position of the heart and rib cage to mimic geometry found during end inspiration and end expiration with these simulated morphological differences compared with actual human SCG measurements. The study enhances our understanding of the SCG genesis, which may lead to increased diagnostic utility of SCG, plus provides potentially useful information on the chest surface distribution of SCG to help guide sensor location placement choices.

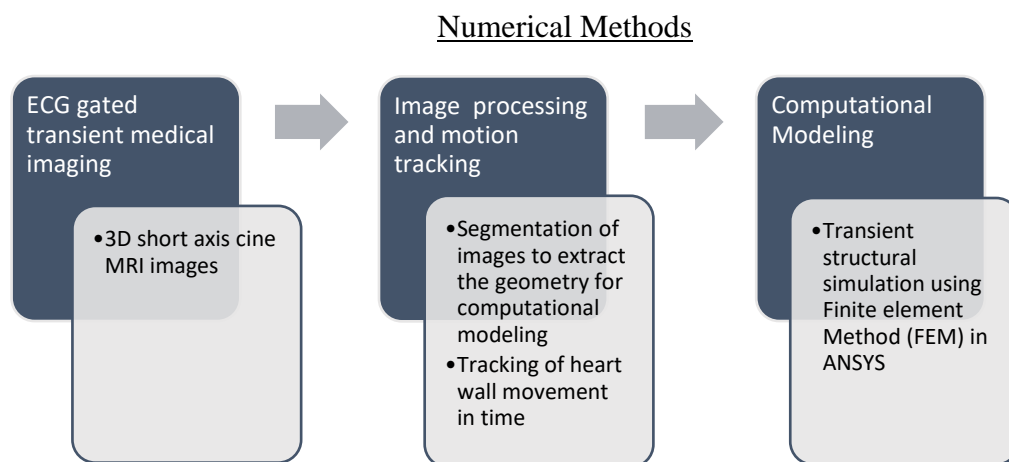


Figure 24- Methodology

The methodology employed in the current study is shown in Fig. 24. First, short-axis images of cine MRI were acquired. These contain gray-scale images of the heart chambers at different times during the cardiac cycle. The short-axis MRI images at different times were combined to generate a 3D geometry (or a voxel volume). Then, a region of interest that contained a part of the thorax adjacent to the heart wall was

manually segmented from the voxel volume to derive the geometry for the computational model. The region was selected such that it contains the surface locations with likely high SCG amplitudes. The SCG signal is known to be the resultant of the heart, valve, and blood flow movements (Amirtahà Taebi et al., 2019). The current study focuses on the propagation of measured ventricle surface movements (which would contain the contribution from valve and blood flow movements) to the chest surface. Short-axis MRI images provides ventricle movements only, which are known to have a dominant effect on the SCG morphology (Crow et al., 1994). Hence, previous studies have used ventricle movements alone to model SCG (Akhbardeh et al., 2009; Tavakolian et al., 2012), this approach is followed in the current study. Furthermore, it can be argued that the movements of main arteries (e.g., movements caused by blood acceleration/deceleration in aorta) contribute to ventricles movements since these organs are mechanically coupled.

The myocardial movement was extracted using a motion tracking algorithm implemented in MATLAB (2017b. The MathWorks, Inc., MA). This algorithm utilized the frame to frame image intensity changes to quantify tissue movement. The tracked 3D motion at the heart wall was then used as the boundary conditions in the computational model, which simulated the myocardial motion propagation to the chest surface.

Medical imaging

Cine MRI frames capture the cardiac motion gated by ECG-triggered segmented imaging. As shown in Fig. 25, the different frames (shown by different color boxes) were acquired during consecutive cardiac cycles and then combined to create

the final movie (or cine) that shows cardiac movements. Hence, the produced movie represents an average (rather than an individual) cardiac cycle.



Figure 25- ECG tracing for cine MRI imaging gating. Different colors refer to acquisition time windows. The windows with the same color are averaged to produce the cine MRI images.

Short-axis cine Steady State Free Precession (SSFP) MRI images available in an online database (Tobon-Gomez et al., 2013) were used in the current study. These images were acquired using a 3T Philips Achieva System (Philips Healthcare, Best, The Netherlands) during breath-holds of approximately 15 s and were gated by ECG (Tobon-Gomez et al., 2013). The selected images were acquired from a 28-year-old healthy male volunteer. The short axis images contained 14 slices with a slice spacing of 8 mm. Each Image contained 256 x 256 pixels with a pixel size of 1.25mm. The images were available in DICOM format and were read using the image processing toolbox in MATLAB (2017b. The MathWorks, Inc., MA). The ECG gated MRI frames contained the timing with respect to the R peak, where the first frame was acquired at that peak and the final frame was acquired at 791 ms after the peak. The cardiac cycle corresponded to 30 frames in this data set. These images were interpolated in both time and Z direction (normal to image plane) such that there were 60 frames per cardiac cycle and 56 slices at each time frame.

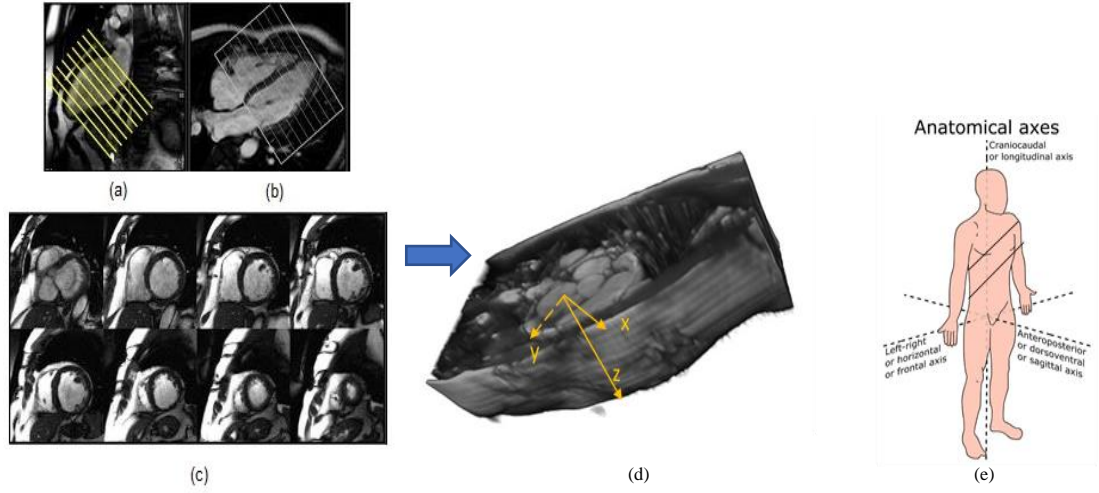


Figure 26- MRI images (a) 2- chamber view (b) 4-chamber view and (c) short axis views in different planes, (d) voxel volume and (e) the slice orientation with respect to the chest denoted by 2 parallel lines

Motion tracking

The motion tracking algorithm employed an optical flow-based method as described below. The coordinates (x, y, z) represent a point in the volumetric domain (i.e, voxel dataset), which is formed by combining the short-axis image slices where x and y are in the directions of width and height of the images and z is the direction normal to the images.

If the intensity of a point (x, y, z) at time t is $I(x, y, z, t)$ and the same point moves to a location $(x + \delta x, y + \delta y, z + \delta z)$ in δt time, optical flow assumes the intensity remains constant as denoted in equation (12).

$$I(x, y, z, t) = I(x + \delta x, y + \delta y, z + \delta z, t + \delta t) \quad (12)$$

Using Taylor series expansion;

$$I(x + \delta x, y + \delta y, z + \delta z, t + \delta t) = I(x, y, z, t) + \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial z} \delta z + \frac{\partial I}{\partial t} \delta t + H.O.T \quad (13)$$

Using equations 12 and 13 and neglecting the higher order terms (H.O.T), the following equation can be derived.

$$\frac{\partial I}{\partial x} \frac{\delta x}{\delta t} + \frac{\partial I}{\partial y} \frac{\delta y}{\delta t} + \frac{\partial I}{\partial z} \frac{\delta z}{\delta t} = - \frac{\partial I}{\partial t} \quad (14)$$

Equation 3 can also be written as:

$$I_x V_x + I_y V_y + I_z V_z = -I_t \quad (15)$$

where I_x, I_y, I_z denote 3D spatial intensity derivatives and I_t is time derivative of the intensity.

Different methods have been proposed to solve equation 15 to find the velocities (V_x, V_y, V_z). In this study, the method proposed by Lucas and Kanade (Lucas & Kanade, 1981) was implemented.

In this method, a $n \times n \times n$ neighborhood block in the 3D space is assumed to have the same motion.

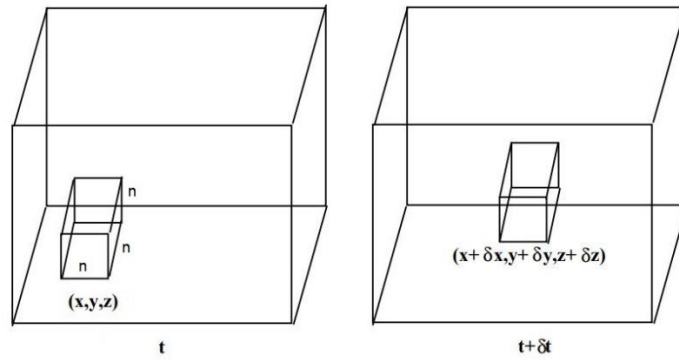


Figure 27-Movement of $n \times n \times n$ neighborhood from frame to frame

Velocity field V can be solved by minimizing equation 16.

$$\sum_{x,y,z \in \Omega} [\nabla I(x, y, z, t) \cdot V^T + I_t(x, y, z, t)]^2 \quad (16)$$

A Gaussian windowing function W (W is an array with $n \times n \times n$ elements containing the Gaussian coefficients) was utilized, such that the priority is given to the centered pixel. With the inclusion of the Gaussian windowing function, equation 16 becomes,

$$\sum_{x,y,z \in \Omega} W^2(x,y,z) [\nabla I(x,y,z,t) \cdot V^T + I_t(x,y,z,t)]^2 \quad (17)$$

Velocity vector V can be solved using equation 18.

$$V = [A^T w^2 A]^{-1} A^T w^2 B \quad (18)$$

where

$$A = [\nabla I(x_1, y_1, z_1), \dots, \nabla I(x_N, y_N, z_N)]$$

$$w = \text{diag}[W(x_1, y_1, z_1), \dots, W(x_N, y_N, z_N)]$$

$$B = -[I_t(x_1, y_1, z_1), \dots, I_t(x_N, y_N, z_N)]$$

$$N = n \times n \times n.$$

$$A^T w^2 A$$

$$= \begin{bmatrix} \sum w^2(x,y,z) I_x^2(x,y,z) & \sum w^2(x,y,z) I_x(x,y,z) I_y(x,y,z) & \sum w^2(x,y,z) I_x(x,y,z) I_z(x,y,z) \\ \sum w^2(x,y,z) I_y(x,y,z) I_x(x,y,z) & \sum w^2(x,y,z) I_y^2(x,y,z) & \sum w^2(x,y,z) I_y(x,y,z) I_z(x,y,z) \\ \sum w^2(x,y,z) I_z(x,y,z) I_x(x,y,z) & \sum w^2(x,y,z) I_z(x,y,z) I_y(x,y,z) & \sum w^2(x,y,z) I_z^2(x,y,z) \end{bmatrix}$$

$$A^T w^2 B = \begin{bmatrix} \sum w^2(x,y,z) I_x(x,y,z) I_t(x,y,z) \\ \sum w^2(x,y,z) I_y(x,y,z) I_t(x,y,z) \\ \sum w^2(x,y,z) I_z(x,y,z) I_t(x,y,z) \end{bmatrix}$$

The following equations were used to numerically calculate the needed derivatives.

$$I_x = \frac{I(x + \Delta x, y, z, t) - I(x - \Delta x, y, z, t)}{2\Delta x}$$

$$I_y = \frac{I(x, y + \Delta y, z, t) - I(x, y - \Delta y, z, t)}{2\Delta y}$$

$$I_z = \frac{I(x, y, z + \Delta z, t) - I(x, y, z - \Delta z, t)}{2\Delta z}$$

$$I_t = \frac{I(x, y, z, t + \Delta t) - I(x, y, z, t - \Delta t)}{2\Delta t}$$

At the image boundaries the I_x , I_y and I_z are calculated from,

$$I_x = \frac{I(x + \Delta x, y, z, t) - I(x, y, z, t)}{\Delta x}$$

$$I_y = \frac{I(x, y + \Delta y, z, t) - I(x, y, z, t)}{\Delta y}$$

$$I_z = \frac{I(x, y, z + \Delta z, t) - I(x, y, z, t)}{\Delta z}$$

After calculating the velocity vector V , the relative position of the next frame is calculated using equation 19.

$$p^T = p_0^T + \Delta t \times V^T \quad (19)$$

where

$p_0 = [x_0 \ y_0 \ z_0]$ is the initial position of a point and $p = [x \ y \ z]$ is the position of the point in next frame after Δt . In the current study, a block size of 5x5x5 was used and the diagonal elements of the Gaussian windowing function W was selected to be unity (Baron, 2004).

Validation of motion tracking

The accuracy of the motion tracking algorithm was tested using a benchmarking framework for validating cardiac motion tracking algorithms (Tobon-Gomez et al., 2013). This benchmarking framework provided cine steady state free precession (SSFP) MRI and 3D tagged MRI data (acquired on the same subject) with the ground truth motion for 12 landmarks in the ventricle wall as shown in Fig. 28.

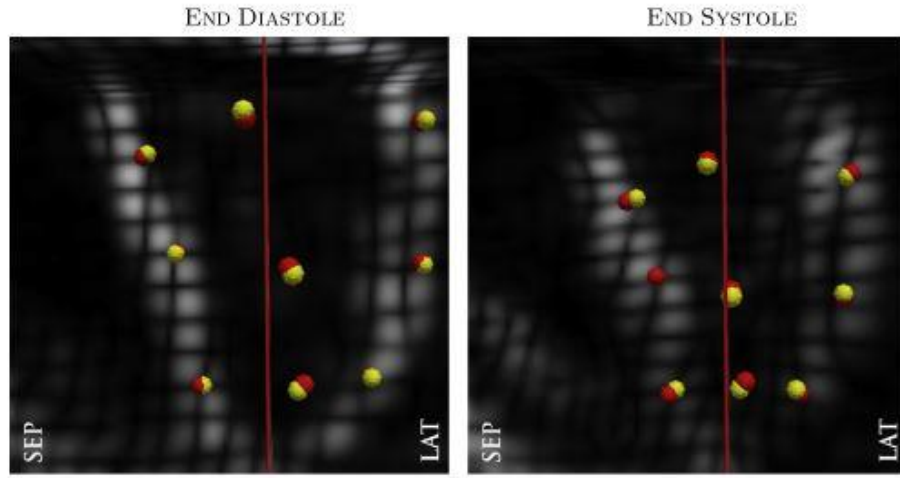


Figure 28-Tracked ground truth of landmarks on ventricle wall by two observers (Tobon-Gomez et al., 2013)

The ground truth motions of these 12 landmark points were manually tracked by two experienced observers using the 3D tagged dataset and the ground truth location data were mapped in to (SSFP) MRI images coordinates by registering 3D tagged MRI coordinates to (SSFP) MRI images coordinates (relative to anatomical coordinate system) (Tobon-Gomez et al., 2013). The current study used the same SSFP MRI images provided in the study (Tobon-Gomez et al., 2013) and the accuracy of the motion tracking algorithm was tested by comparing the tracked motion of the 12 landmark points against the ground truth motion of the 12 points provided in the study (Tobon-Gomez et al., 2013). The distance (in 3D space) between the ground truth and the tracked motion over time can be used to quantify the accuracy level for a motion tracking algorithm. However, due to different temporal resolutions between SSFP and 3D tagged MRI images , the previous study [16] suggested to validate the motion tracking on SSFP MRI images by comparing the Euclidean distance between the tracked and the ground truth locations of the 12 points at the frames closest to end systole and the final frames (Tobon-Gomez et al., 2013). Fig. 29 shows the tracked and

ground truth locations of the 12 landmark points at the frame closest to end systole and the final frame (i.e. frame closest to end diastole).

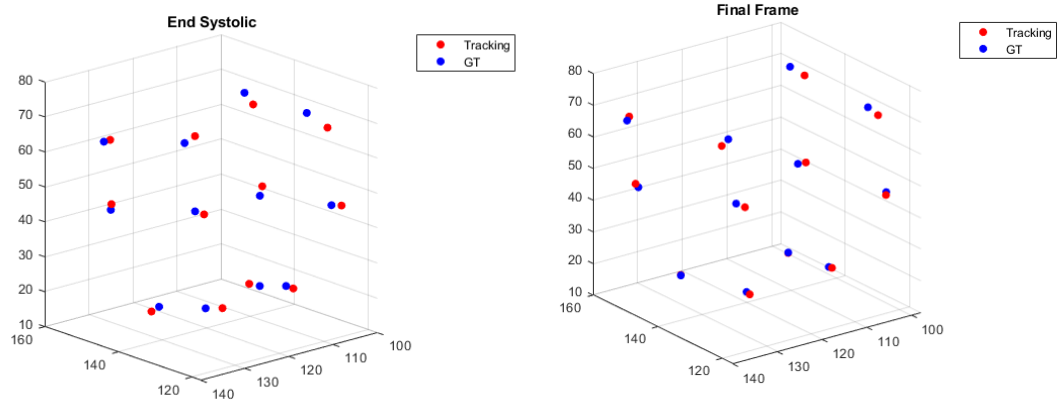


Figure 29- Locations of ground truth (GT) and tracked locations at (a) end systolic and (b) close to end diastolic frames in 3D Euclidean space (all dimensions are in mm)

An Euclidean difference of 4.10 ± 2.29 mm and 2.06 ± 1.36 mm (mean \pm SD) were observed between the tracked and ground truth locations of the 12 points at end systole and end diastole frames, respectively. The error may be partially due to the spatial and temporal resolution and due to variabilities of different acquisition sessions of SSFP and 3D tagged MRI images (Tobon-Gomez et al., 2013). The error in calculated is comparable to acceptable errors in a previous benchmarking study (Tobon-Gomez et al., 2013).

Due to the complex fiber structure orientations of the heart muscles, the heart exhibits a complex movement which involves rotation, twisting and longitudinal movement during cardiac cycle (Ingels Jr, Daughters 2nd, Stinson, & Alderman, 1975). It is important to accurately capture these complex 3D movements to properly model the SCG waveform at the chest surface. Fig. 30 shows a tracked contour at the ventricle walls in the short-axis view (and another view) at end diastole and end systole times. The displacement vectors of the contour points are shown using red arrows.

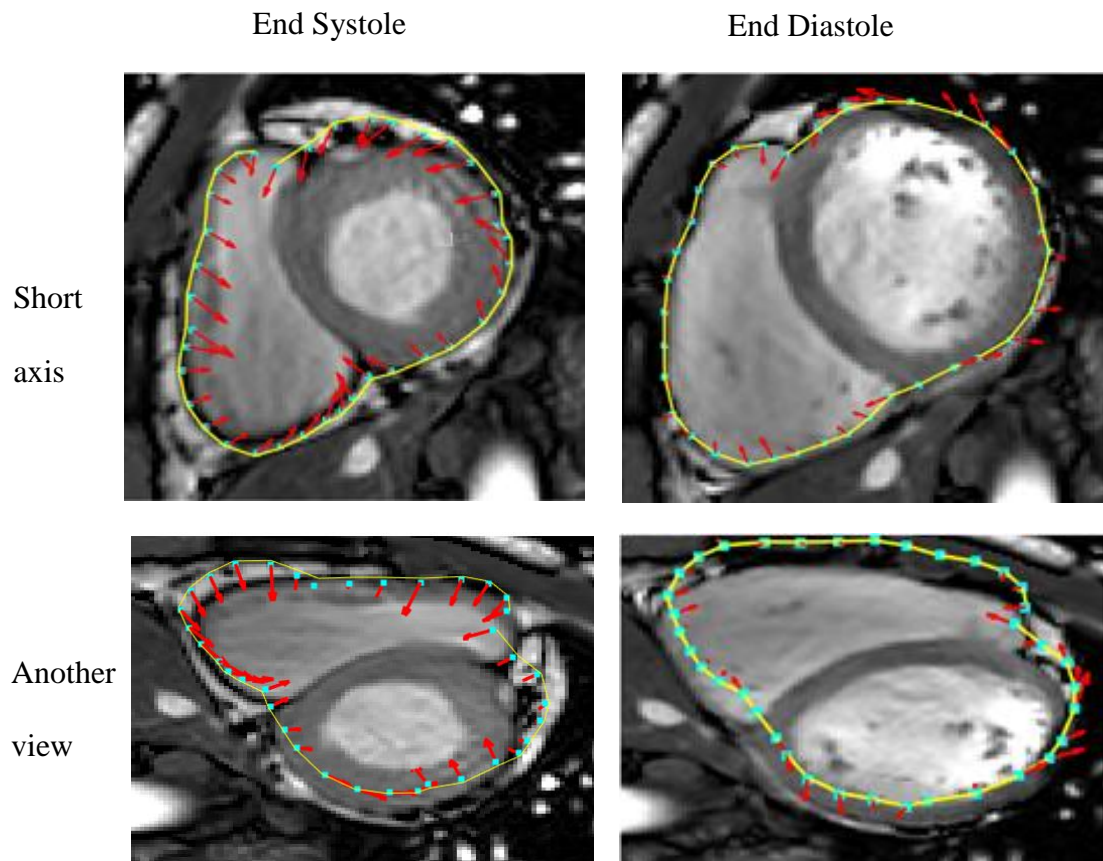


Figure 30-Tracked contours at the ventricle walls (endocardium) with the displacement vectors (represented by red arrows) at the end systolic and end diastolic states. Top: short axis view, Bottom: Another view to increase clarity. The displacement between consecutive frames is shown by the arrows. The arrow length equals 8 times the displacement to make arrows more visible.

The direction of these arrows suggests the ability of the algorithm to capture the rotational and twisting action of the heart wall as well as the movement in the longitudinal direction of the heart.

Computational modeling

The computational domain was generated by segmenting the voxel dataset formed by stacking the short-axis MRI slices. The segmented region is shown in Fig. 31. This region consisted of the chest muscle, ribs, cartilage and lung as shown in Fig. 32 (a) and (b). Although, the position of the sternum could be seen in the MRI images, the intensity contrast levels of the current MRI imaging didn't allow accurate segmentation of the ribs and cartilage. Hence, the ribcage was modeled separately according to the dimensions available in literature (Laurin, Jobin, & Bellemare, 2012). The geometry, was then scaled and placed in the model to match the size and orientation of the sternum and ribs. The intercostal muscles were modeled by connecting the muscle attachments between ribs (Hamzah et al., 2013). The location and orientation of the computational geometry on human chest is shown in Fig. 32 (c). The geometry was limited to that region to reduce computational expense while including areas of the chest surface where SCG is often measured (Amirtahà Taebi et al., 2019). Future studies would expand the simulated region to include the entire thorax.

Computational modeling was implemented in ANSYS (ANSYS Inc, Canonsburg, PA) transient structural analysis module, which uses finite element method (FEM) to solve for the displacements in the computational domain. In FEM, the structural domain is divided into small (finite) elements, which are assembled to

satisfy the required boundary conditions and interfaces (i.e., contacts) between the model components. The forces and displacements acting on each element are governed by the equation of motion. A system of equations is formed by combining the equations from all the elements, which is then solved to find the displacement of all elements (Peshala T Gamage 2019).

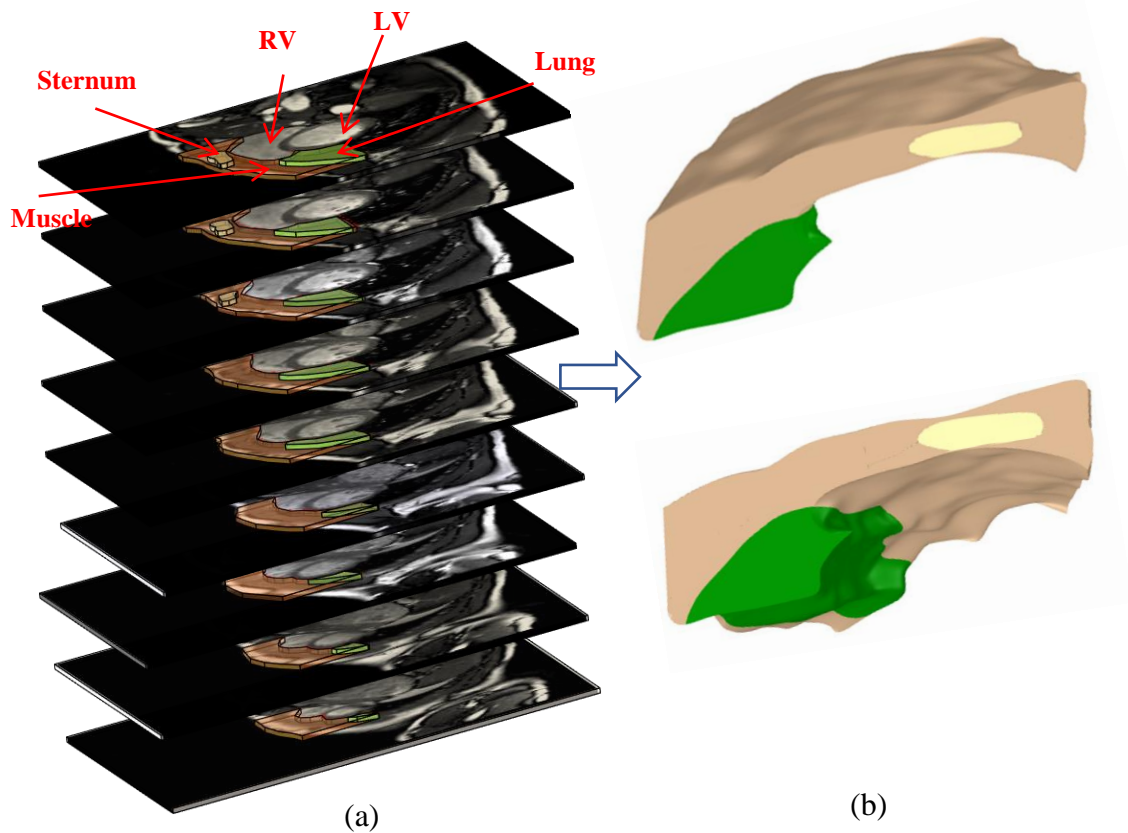


Figure 31- Segmentation and 3D model construction. (a) Segmented voxels from the MRI slices, (b) 3D model constructed by combining voxels.

The generated finite element mesh shown in Fig. 33 (a) contained ~45000 elements and ~84500 nodes and was generated in ANSYS. The tracked motion of the heart surface was mapped to the computational model as shown in Fig. 33 (b). The model was constrained by fixing all degrees of freedoms at the medial and lateral boundaries of the ribs as shown in Fig. 33 (c) and (d). This may be justified by the very low (near zero) movements (during breath-hold) of the rib cage. Bonded contacts were

defined at the rib cage and muscle interfaces. A contact condition that allows sliding between the lung and chest wall was defined at the lung-chest wall interface.

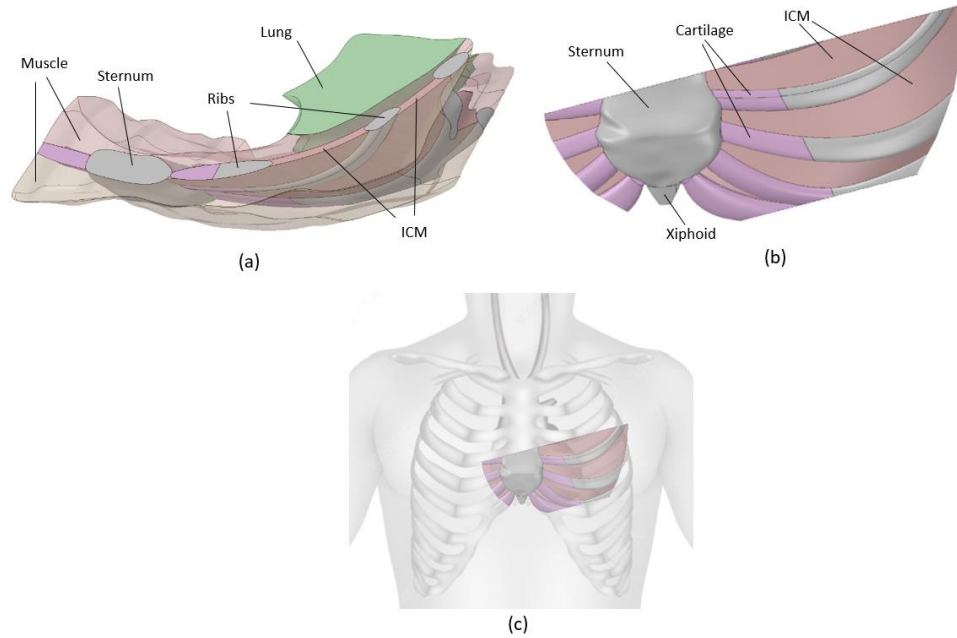


Figure 32- Chest location and detailed structures in the 3D modeled region (a) Muscular, bony and lung regions (b) Ribs, ICM, cartilage, sternal and xiphoid (c) location and orientation of the modeled region.

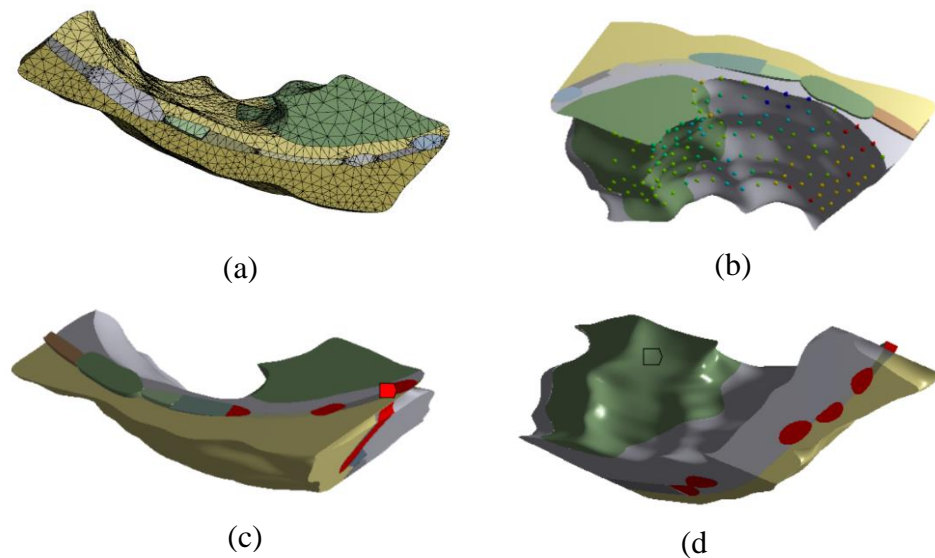


Figure 33- 3D representations of the computational model domain: (a) Computational mesh, (b) MRI-mapped displacements of heart surface and (c) and (d) more 3D views. The rib edges are colored in red.

Table 2- Material properties

Component	Young's modulus	Density (kg/m ³)	Poisson's ratio	References
Chest Muscle (includes pericardial muscle, pectoralis major, Skin)	0.5-2.5 MPa	1000	0.3	(Chawla, Mukherjee, & Karthikeyan, 2006; Gefen & Dilmoney, 2007; Zigras, 2007)
Sternum, Ribs, Xyphoid	12 GPa	2000	0.4	(Sundaram & Feng, 1977)
Costal cartilage	3 GPa	2000	0.4	(Chawla et al., 2006)
Intercostal muscle	3 MPa	1000	0.4	(Chawla et al., 2006)
Lung	Mooney Rivlin 5 th order (C10=-859.78 Pa C01=947.5 Pa C20=1783.2 Pa C11=-5440.5 Pa C02=4633.5 Pa)	1250	-	(Al-Mayah, Moseley, & Brock, 2007; Zeng, Yager, & Fung, 1987)

A linear elastic behavior is assumed for all the components except for the lung, where hyperplastic material properties were fitted in to Mooney-Rivlin 5th order model using available information (Al-Mayah et al., 2007). The material properties for each component are shown in Table 2. Although the chest muscle consists of different components (i.e., pericardial muscle, pectoralis major, skin), the chest muscle and skin were modeled as a single component in this study. Since previous studies (Chawla et al., 2006; Gefen & Dilmoney, 2007; Zigras, 2007) reported a range of elasticity values for these components, the effect of different elasticities on the simulated SCG was investigated.

Results and Discussion

Simulated SCG waveform

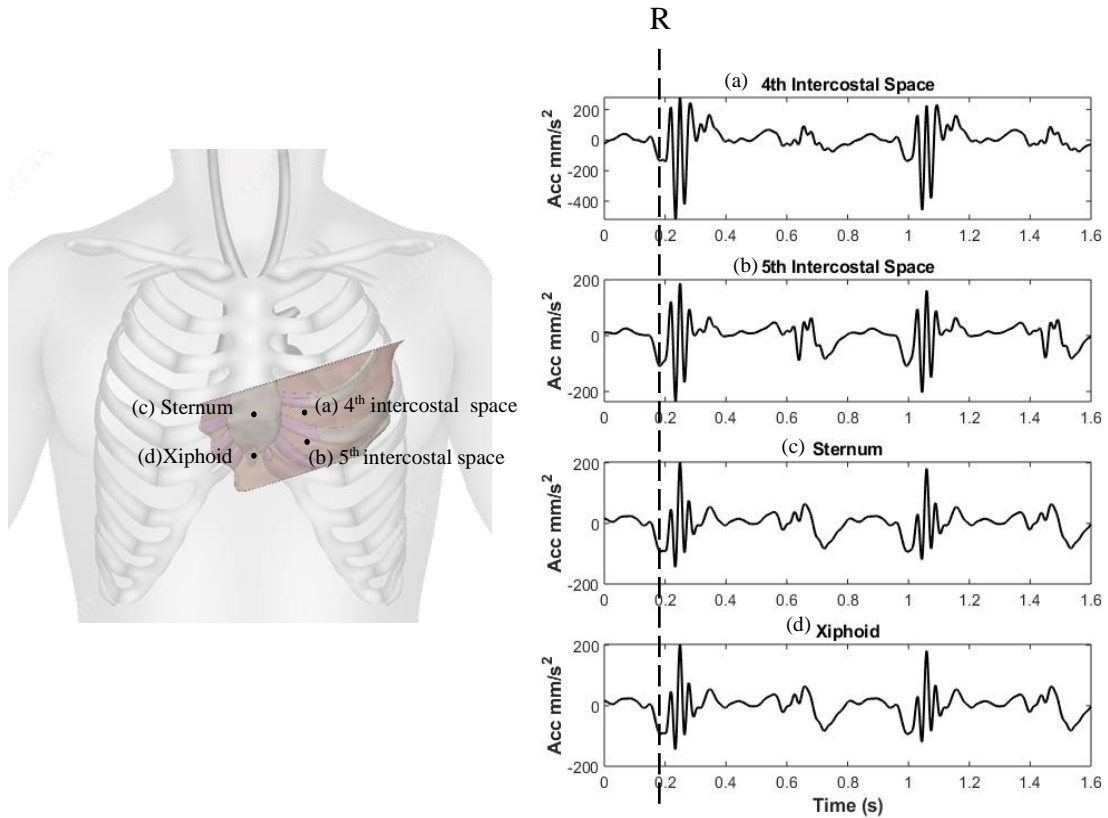


Figure 34-Right: Calculated chest surface acceleration waveforms (in dorso-ventral direction) at several locations. The vertical dashed line indicates the timing of the ECG R-peak extracted from MRI. Left: Location of the 4 precordial points. The calculated waveforms had similarities with measured SCG.

The acceleration in the dorso-ventral direction was calculated at several the chest surface locations (Fig. 34). These signals showed clear peaks as seen in the SCG1 and SCG2 regions (which correspond to heart sounds S1 and S2).

In addition, the blood volume of the left ventricle (i.e., left ventricular volume (LV)) was followed by tracking the epicardium boundary in the cine short-axis MRI images and the calculating the voxel volume inside the tracked boundary. The LV volume variation was used to identify the end systolic and end diastolic states (Kido et

al., 2016) and to relate the observed SCG features with respect to the timing of left ventricular contraction. This approach of validating the SCG timing was previously reported (Akhbardeh et al., 2009; Peshala T Gamage 2019).

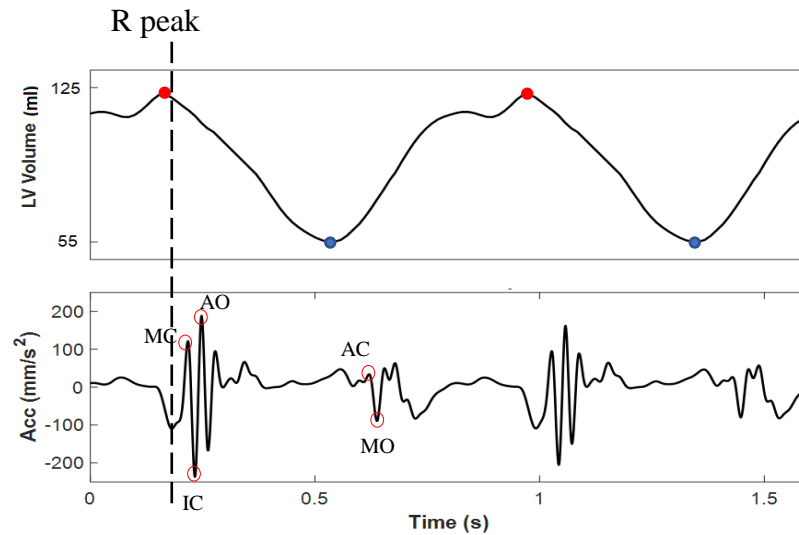


Figure 35- Feature points identified in the calculated SCG waveform (at 5th intercostal space) with the corresponding LV volume variation and R-wave timing. The feature points appeared consistent with the LV volume variation and previous studies (Crow et al. 1994)

Several feature points were identified in the simulated signal in relation to the previously observed features of experimental studies (Crow et al., 1994; Amirtahà Taebi et al., 2019). These feature points of the simulated SCG signal are shown Fig. 35 with the respective LV volume variation and timing of the ECG signal. It can be seen in Fig. 35 that the relatively high amplitude portions of the SCG signal (i.e., SCG1 and SCG2 regions) were observed just after the end diastolic volume (EDV) and end systolic volume (ESV) times of the LV volume signal. This suggests that these SCG portions include components generated by valve opening and closure (i.e., similar to S1 and S2 heart sounds). Following previous studies (Crow et al., 1994; Amirtahà Taebi et al., 2019), several feature points were identified after the EDV: (1) Mitral valve

closure (MC) at the beginning of the sharp downfall of the SCG signal following QRS complex (2) Iso-volumic contraction (IC) at the nadir of following MC, (3) Aortic opening (AO) was at the peak following IC. (4) aortic valve closure (AC) at the beginning of the sharp downfall of the SCG after the ESV timing (5) mitral valve opening (MO) at the nadir of the down slope following AC. The ability of the model to predict the presence these feature points contributes to its validity.

Effect of chest soft tissue stiffness

The chest wall contains bones and soft tissue. In the current study, the soft tissue mainly consists of the skin, subcutaneous fat, and muscles such as the pectoralis major and intercostal muscles (ICM). ICM are known to have relatively higher stiffness and were treated as such. The rest of the soft tissue were assigned a lower stiffness as shown in Table 1. Since different stiffness values were reported (Chawla et al., 2006; Gefen & Dilmoney, 2007; Zigras, 2007), the effect of the stiffness on the simulated SCG signal was investigated. The simulated SCG waveform for different stiffness values are shown in Fig. 36.

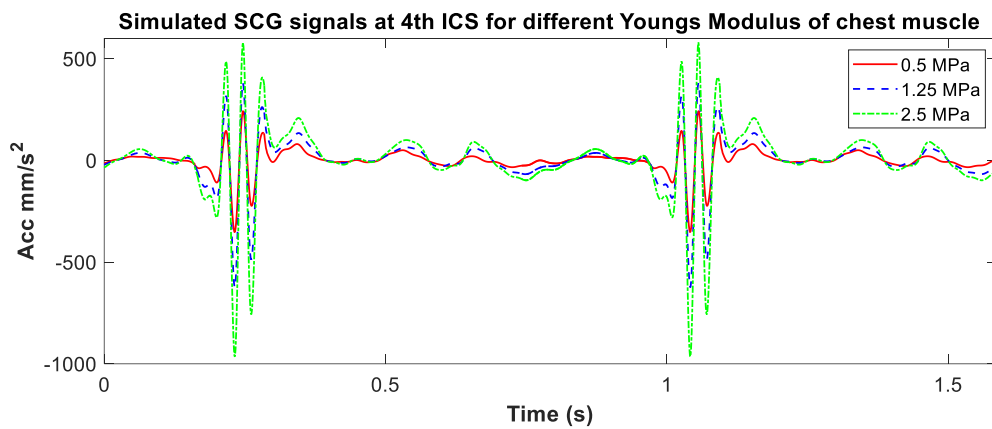


Figure 36- The effect of chest soft tissue stiffness on the SCG waveform at 4th ICS. Similar trends were observed at other chest surface locations. Higher stiffness appears to increase the waveform amplitude, which is consistent with simplified 1-D models.

As seen in Fig. 36, the acceleration magnitude increased with the muscle elasticity increase suggesting that the stiffer the muscle, the higher the SCG amplitude. This phenomenon can be explained by considering a simplified spring mass system with translational motion from one boundary as shown in Fig.37.



Figure 37- Simplified representation of a spring mass system with translational motion from a moving boundary

The acceleration of the mass can be represented by equation 20.

$$\ddot{x} = \frac{k}{m}(x - u) \quad (20)$$

where k , m denote the stiffness and mass of the system, respectively. The input motion (similar to the displacements at the chest tissue subsurface) is denoted by u while x , \ddot{x} denote the displacement and acceleration of the mass, respectively. Equation 20 suggests that the acceleration is proportional to the stiffness of the system which explains the trends observed in Fig. 37.

Surface distribution of SCG

Surface distribution of the acceleration (i.e., SCG) in the dorso-ventral direction was analyzed. Although the simulation can provide the instantaneous accelerations during the whole cardiac cycle, the results are shown for the timing of SCG 1 and SCG 2 peaks (i.e., high amplitude regions that correspond to heart sounds S1 and S2) as examples. Here, the timing of the SCG1 and SCG2 peaks were selected relative to the simulated SCG signal at 4th intercostal space (ICS).

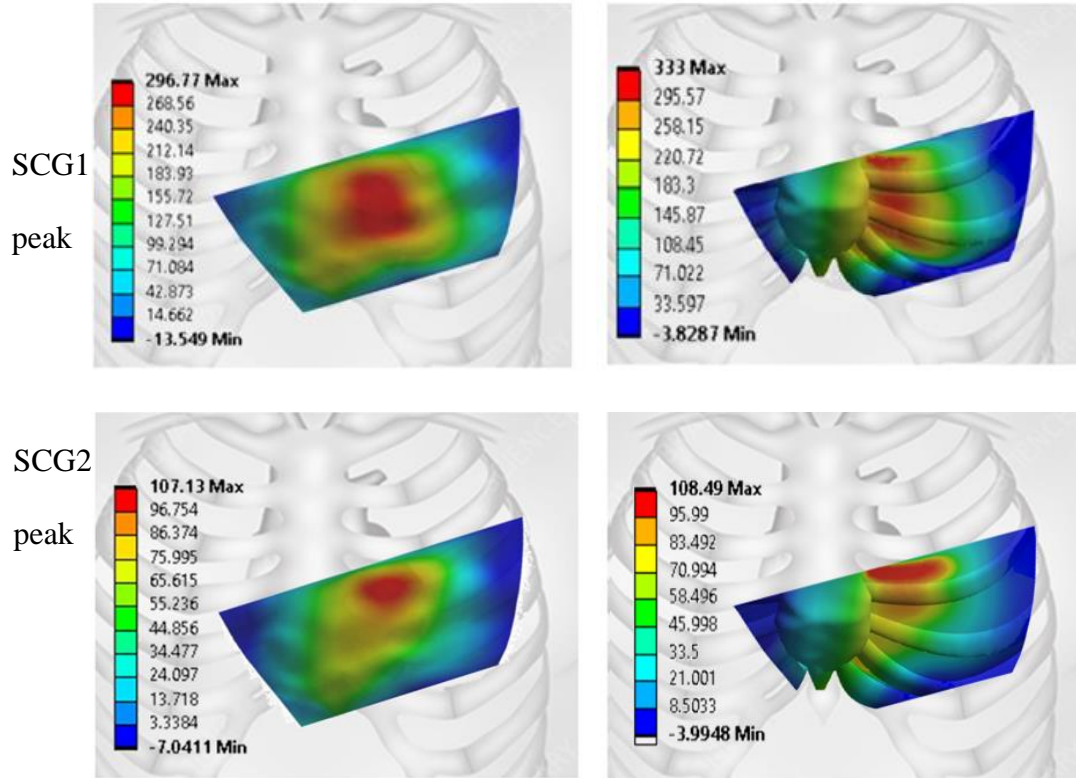


Figure 38- Surface acceleration distribution in dorso-ventral direction at: (a) & (c) the chest surface and (b) & (d) underneath the skin and pectoralis major. Sub-plots (a) & (b) show results at the SCG 1 peak timing while (c) and (d) are at SCG 2 peak timing. Timing was determined from the SCG signal at 4th intercostal space. The calculated surface SCG amplitude was comparable to measured values. The locations of the maximum SCG1 and SCG2 were around the 4th and 3rd ICS, respectively. The surface acceleration was less loud and more uniform than the subsurface SCG. There was less than 15% change in surface SCG amplitude in the 3cm diameter area surrounding the maximum amplitude location. This suggests that a sensor positioning error of 3 cm may be tolerable; all the measurements are in mm/s²

Fig. 38 (a) and (c) show the SCG distribution on the chest surface and Fig. 38 (b) and (d) show that acceleration beneath the chest muscle (also the surface of the rib cage and ICMs). It can be seen that maximum SCG1 amplitudes are located around 4th ICS near left sternal border. This is similar to the heart sound S1 caused by the mitral and tricuspid valve closure (Felner, 1990a). The location of maximum SCG2 occurred at the 3rd ICS near left sternal border, which is similar to the aortic and pulmonary

valve sounds near the heart base (Felner, 1990b). This Fig.38 (b) suggests that the SCG amplitude tended to decrease over the sternum possibly due to the increased bony content. Results also showed that the region of high SCG amplitudes tended to slightly spread towards the xiphoid process. This may be because the decreased bone content at the siphoned may have provided more freedom to the chest surface movements.

In addition, the surface acceleration on the rib surfaces (Fig. 38 (b) and (d)) showed high amplitude regions similar to those observed at the chest surface (Fig. 38 (a) and (c)). It can also be seen that higher accelerations occurred at the ICMs compared with the ribs, possibly due to the fixed boundary conditions at the medial and lateral boundaries of the ribs. Interestingly, more homogenous distribution was seen over the chest surface, suggesting that the chest muscle diffused the sub-surface non-homogeneous accelerations as they propagated to the chest surface.

Results also showed that there was less than 15% change in surface SCG amplitude in the 3cm diameter area surrounding the maximum amplitude location. This suggests that a sensor positioning error of 3 cm may be tolerable. In addition, the simulated results were compared with experimental measurements of SCG distribution over chest surface recorded on a healthy subject using 36 accelerometers (Model: 356A32, PCB Piezotronics, Depew, NY). Here, 32 sensors were placed on the intercostal spaces and 4 sensors were placed on right clavicle, left clavicle, mid sternum and xiphoid locations as seen in Fig. 39. Sensors placed in the intercostal spaces were approximately 3 cm apart (center to center) while the sensors closest to the sternum were placed at sternal borders.

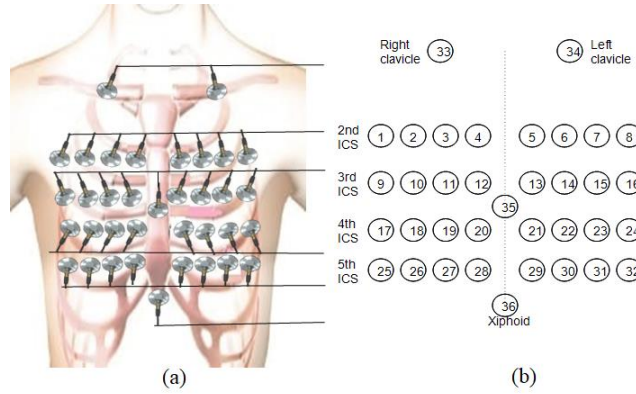


Figure 39-(a) Array of sensors attached to the chest surface (b) sensor locations with index. (Refer with SCG amplitude maps shown in Fig. 17)

Fig. 40 shows the numerically-modeled surface distribution of SCG at the time of different feature points of the SCG. Similarly, Fig. 41 shows the surface distribution of SCG measured on a healthy subject during inspiration breath hold. For comparison purposes, the computational domain is shown using a black dashed line in Fig.41.

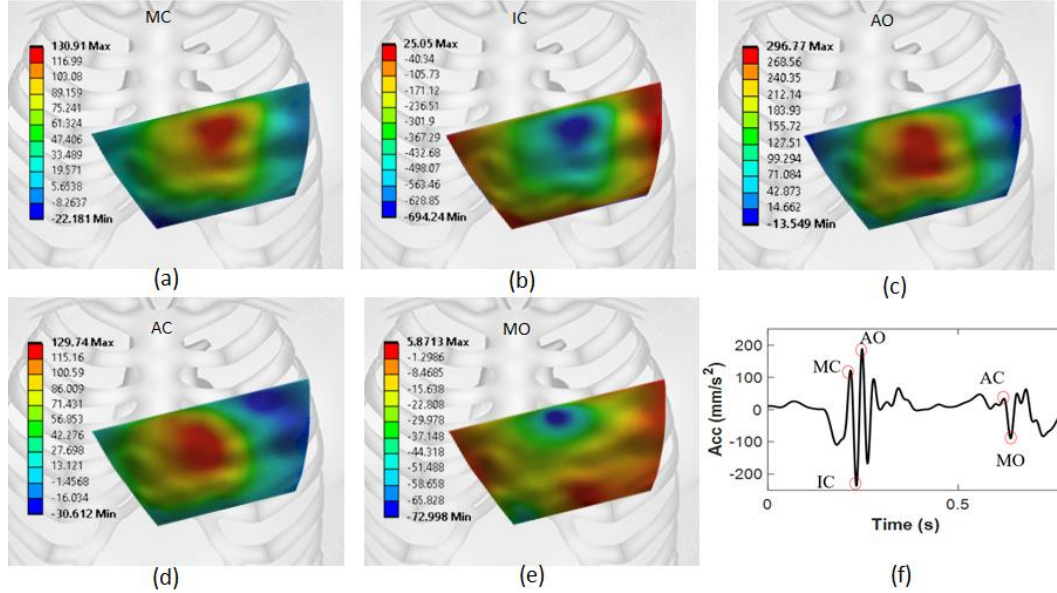


Figure 40- Computational surface SCG amplitude maps at the timing of feature points: (a) MC (b) IC (c) AO (d) AC (e) MO and (f) identified feature points on the SCG signal measured at 5th ICS; all the measurements are in mm/s^2

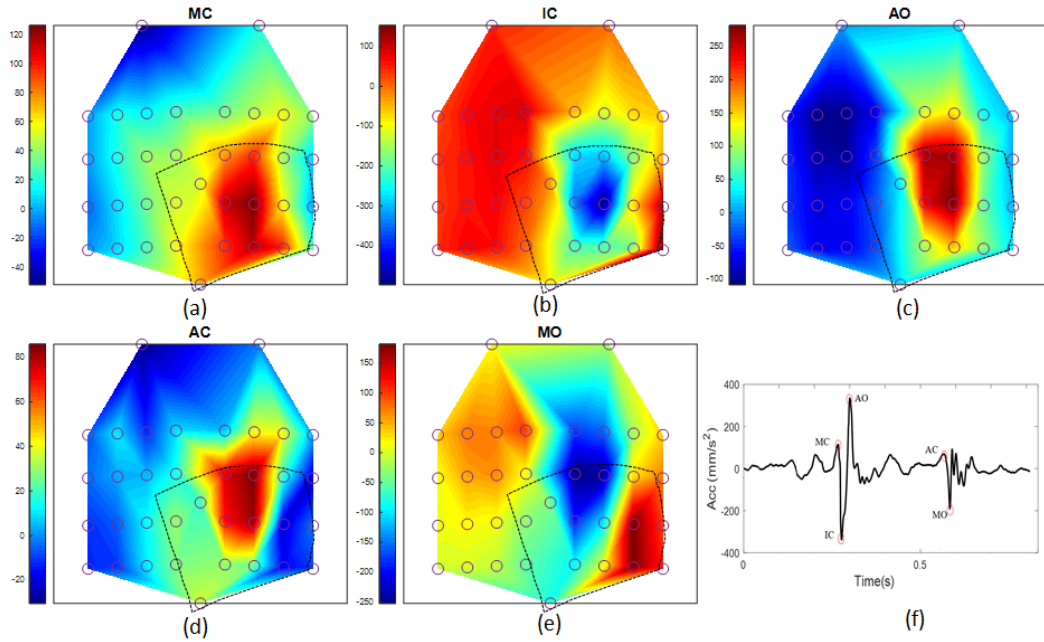


Figure 41- Experimental surface SCG amplitude maps at the timing of feature points: (a) MC (b) IC (c) AO (d) AC (e) MO and (f) identified feature points on the SCG signal measured at 5th ICS. (Small circles on Figures (a) to (e) represent the sensor locations as shown in Fig. 39 (b). The dashed line encloses the region to be compared with the computational model); all the measurements are in mm/s²

The numerically modeled surface SCG magnitude distribution results were comparable to the experimental measurements and with similar trends. At the timing of MC, AO and AC feature points, high positive surface accelerations were concentrated in the region between 3rd and 5th ICS near left sternal border. At the timing of IC and MO feature points, high negative accelerations were concentrated in the same region (3rd and 5th ICS near left sternal border). Differences between the experimental and simulated results may be attributed to the inter-subject variations, different resolutions of the computational and experimental measurements, and limitations of the computational model (i.e., limited geometrical domain, fixed boundary conditions at rib ends, simplified geometry and mechanical properties).

Effect of heart position (due to respiratory motion)

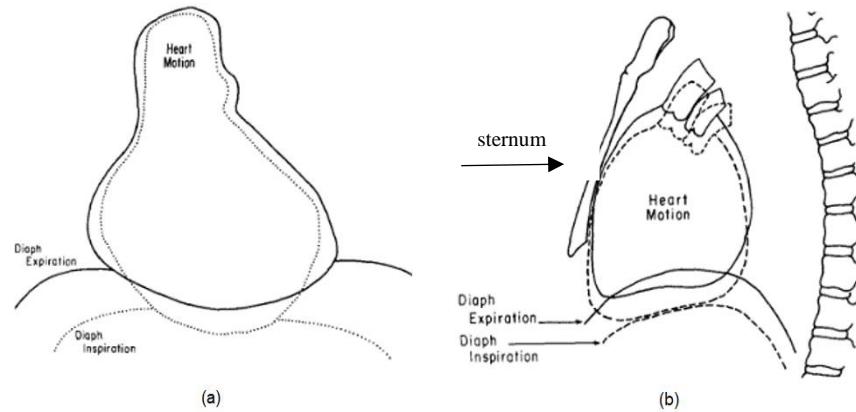


Figure 42- The heart location during respiration in (a) posterior-anterior projection and (b) lateral projection (Bogren, Lantz, Miller, & Mason, 1977)

The SCG signal is known to be affected by respiration. Several studies have investigated the SCG morphological variability due to respiration (Azad, Gamage, Sandler, Raval, & Mansy, 2019; Peshala T Gamage et al., 2018; Peshala T. Gamage, Azad, Taebi, Sandler, & Mansy, 2020; R. H. Sandler et al., 2019) and the possible causes of cardio-pulmonary interactions such as intra thoracic pressure variations and respiratory motion of the heart. To study the effect of heart position (due to respiration) on the SCG signal morphology, the relative position between the heart and ribcage was varied to approximate the end-expiration and end-inspiration states. Here, the original position of the model geometry (derived from MRI imaging) was at end inspiration. To model the end expiratory state, the rib cage was be moved (w.r.t the heart) to approximate end inspiration following a previous study (Bogren et al., 1977) that measured the 3D respiratory movement of the heart using cineangiography. Fig. 42 shows the overall cardiac motion reported in the previous study (Bogren et al., 1977) based on 39 cineangiographies.

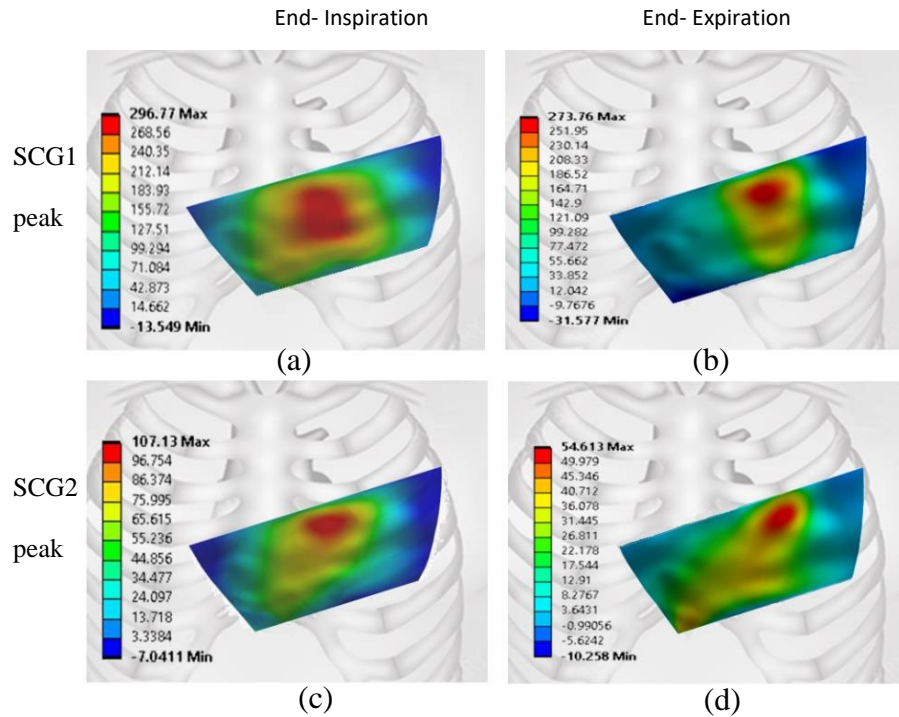


Figure 43- Surface acceleration distribution in dorso-ventral direction at: (a) &(c) End inspiration state; and (b) & (d) end expiration during both SCG 1 and SCG 2 peak timing. The SCG peak tended to be louder at end inspiration. The location of the SCG1 peak moved superiorly at end expiration by about 3cm, which is comparable with the corresponding upward shift of the heart location; all the measurements are in mm/s^2

The surface acceleration distribution of the chest surface at SCG1 peak and SCG2 peak for end inspiratory and expiratory states are shown in Fig. 43. A change in the regions of maximum amplitudes can be observed among the two states. At the timing of SCG1 peak, the region of maximum amplitudes was centered around the left sternal border of the 4th ICS for the end-inspiration and this region has moved laterally and upward towards 3rd ICS for the end expiration state. Similarly, at the timing of SCG 2 peak, the maximum amplitude region moved slightly laterally and superiorly with expiration. This phenomenon can be due to the lateral and upward movement of the heart with expiration state.

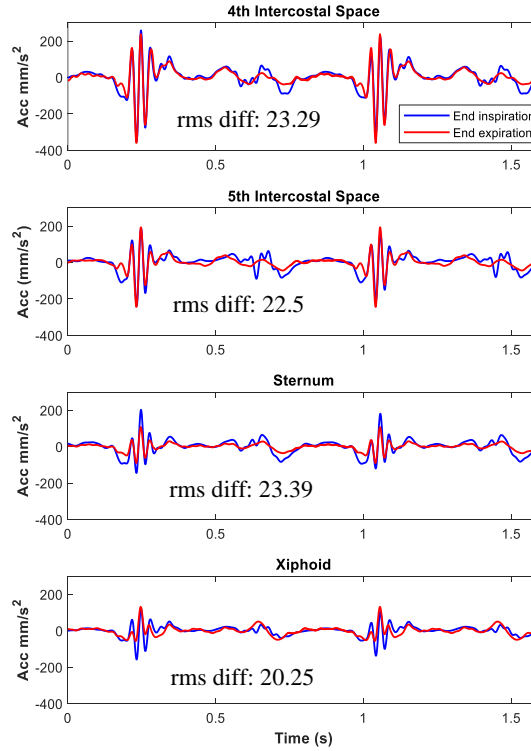


Figure 44- Simulated SCG signals at different chest locations for the end inspiration and end expiration states. The waveform variability (calculated as the rms of the waveform difference) was comparable with experimental results.

To assess the effect of heart position (due to respiratory motion) on the SCG morphology, the simulated SCG was calculated on the same surface location for the end inspiration and end expiration simulations. Slight morphological variations were observed between end-inspiration and end-expiration states. To quantify the dissimilarity between the SCG morphologies, the morphological variability of two SCG waveforms were measured using root mean square (rms) of the difference between two waveforms. Fig. 44 shows the SCG waveforms at end inspiration and end expiration states with their respective rms differences. These differences were comparable with experimental breath-hold (i.e., measured during end-inspiration and end-expiration) SCG data acquired in a previous study (Azad et al., 2019) that reported a rms difference of (mean=28.24, SD=8.909 mm/s²) for the measurements at 4th ICM near left ICB on 5

healthy subjects. These results suggest that the change in relative location of the heart and rib cage during respiration is among the possible reason for the differences in SCG morphology with respiration observed in previous studies (Azad et al., 2019; Peshala T. Gamage et al., 2020).

Acceleration over the heart wall

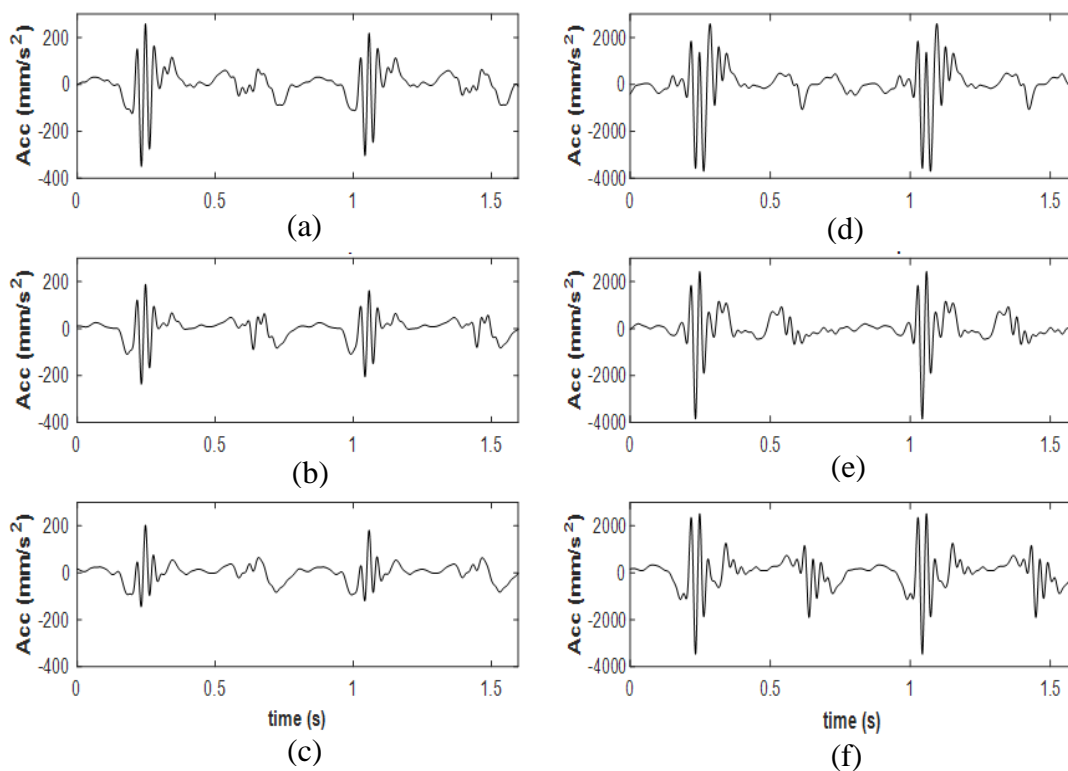


Figure 45- Calculated SCG at 4th ICS, 5th ICS, and sternum (a,b,c, respectively) and corresponding acceleration on the pericardium (d,e,f). The latter locations were determined by projecting the surface locations on the pericardium (The surface point locations are shown in Fig. 34). There was some similarity between surface and subsurface waveforms although the surface amplitudes were about 10 times lower than those at the pericardium, which is consistent with a previous study (Bombardini et al. 2007).

Acceleration in the dorso-ventral direction at heart wall (i.e., pericardium boundary) was studied. Here, the accelerations were recorded at several locations on the chest surface and their corresponding projections on the heart wall (in dorso-ventral direction). These results are shown in Fig. 45. It was observed that the accelerations

measured on heart wall were approximately 10 times larger than the accelerations measured on chest surface which is consistent with literature (Bombardini et al., 2007). Although, the waveforms measured on the heart wall had morphological differences with the corresponding SCG signals at chest surface, they demonstrated SCG like waveforms with similar high amplitude peaks observed in SCG1 and SCG2 regions. There was a trend towards more oscillatory cycles at the heart surface around the SCG1 and SCG2 timings. These results suggested that an invasive acceleration measurement on the heart surface can provide a SCG like high amplitude waveform with similar features to the SCG measured at chest surface. Fig. 46 shows an example of a similar signal measured with an accelerometer on a dog heart wall in a previous study (Ozawa, Smith, & Craige, 1983a) which demonstrated a SCG like waveform.

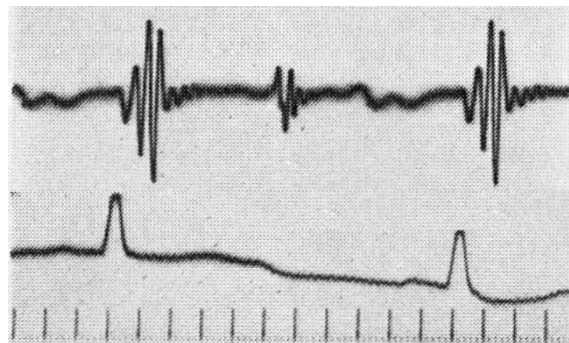


Figure 46- SCG measured at pericardium of a dog (measured using an accelerometer attached at left ventricular wall near interventricular septum) with corresponding ECG (Ozawa, Smith, and Craige. "Origin of the third heart sound. II. Studies in dogs." *Circulation* 67.2 (1983): 399-404). This measured acceleration shows characteristics similar to MRI-extracted signals, especially that of Fig 19 f.

Summary and Conclusions

Computational modeling of cardiac vibration propagation from the heart wall boundary to the chest surface was performed. The model geometry was derived from ECG gated short-axis cine MRI imaging of a healthy subject. The region of interest included a region of the chest above heart ventricles that contained ribs, intercostal

muscles, chest tissue and lung. Displacements at the heart wall were tracked by applying a 3D optical flow-based motion tracking algorithm to the voxel volumes produced by short-axis cine MRI imaging. The accuracy of the tracking was validated, and the tracked displacements were imposed as the boundary conditions at the heart wall in the computational model.

Simulated acceleration signals (i.e. SCG) at the chest surface showed similar morphological features observed in previous experimental SCG studies (Akhbardeh et al., 2009; Crow et al., 1994; Amirtahà Taebi et al., 2019). These feature points were further validated by analyzing their timing with respect to the left ventricle (LV) volume variation and ECG R-peak. The surface SCG distribution results showed similar trends to experimental results. The increased stiffness of the chest muscle tissue resulted in increased SCG amplitudes. The effect the heart displacement (due to respiration) on SCG was investigated by changing the relative location of the heart and rib cage to mimic end-inspiratory and end-expiratory breath hold states. Results showed that the morphological differences of SCG measured at these two states are comparable to the differences observed in experimental measurements suggesting that respiratory motion of the heart can be a contributing factor to the different SCG morphologies with respiration described in previous studies (Azad et al., 2019; Peshala T. Gamage et al., 2020). The acceleration measured over heart wall (i.e., pericardium) produced high amplitude waveforms with morphological features similar to the surface SCG signals. These results provide a strong link between surface measurements and myocardial movements. The study findings should enhance our understanding of SCG spatial distribution and genesis, which may lead to increased SCG clinical utility. With more high resolution medical imaging, the proposed methodology of motion tracking and

computational modeling can be extended to include complex blood flow and valve movements, which would provide more in-depth information about SCG genesis and relation between SCG feature points and cardiac events.

CHAPTER 4 – PREDICTION OF HEART FAILURE PATIENT READMISSION USING SCG

This chapter presents the application of machine learning (ML) utilizing SCG features for predicting HF patient readmission. Heart failure (HF) is one of the major global healthcare concerns. In USA, HF affected adults are estimated as 6.5 million from 2011 to 2014 (*National Health and Nutrition Examination Survey*) and this number is projected to increase by more than 40% between 2015 and 2035 (Nelson, Whitsel, Khavjou, Phelps, & Leib, 2016). Economic burden associated with HF is also extensive which will reach \$70 billion by 2030 (Heidenreich et al., 2013). Substantial amount of the expense is due to hospital readmission which has become a vital area to improve overall healthcare system especially for HF patients. A recent study shows that more than 20% of HF patients are readmitted within 30 days and up to 50% by 6 months (O'Connor, 2017).

These high rates of HF readmission urge to administer the readmission events properly and more efficiently. One approach for better management is to develop a predictive model which can reliably anticipate the readmission of a HF patient. Previous studies (Aggarwal & Gupta, 2014; Annema, Luttik, & Jaarsma, 2009) suggested that main reason behind high readmission rate is worsening HF. Therefore, features which are indicative of different heart conditions should come in handy when predicting HF readmission. Seismocardiography (SCG) is a noninvasive way of evaluating cardiac mechanical processes which possesses potential to detect and monitor cardiac conditions (Amirtahà Taebi et al., 2019). SCG based measurements have gained lots of attention among the researchers since the discovery due to its noninvasive nature and effectiveness in diagnosis of cardiovascular diseases. Arrhythmias such atrial and

ventricular fibrillation which occur due to heart's irregular electrical impulses can be detected using ECG features. But cardiac dysfunctions like heart failure reduced ejection fraction (HFrEF) and heart failure preserved ejection fraction (HFprEF) are caused due to heart wall thickening and diagnostic tests such as ECG, X-ray aren't effective detecting these heart conditions. SCG signal features can be used as a potential way to detect these conditions as SCG originates from cardiac mechanical processes such as myocardial wall motions (Sahoo et al., 2018). Advantage of using SCG signal over acoustic cardiac signals (e.g. phonocardiographic (PCG) signal) lies in its ability to capture cardiac information below audible range which may provide important features related to myocardial movements.

Methodology

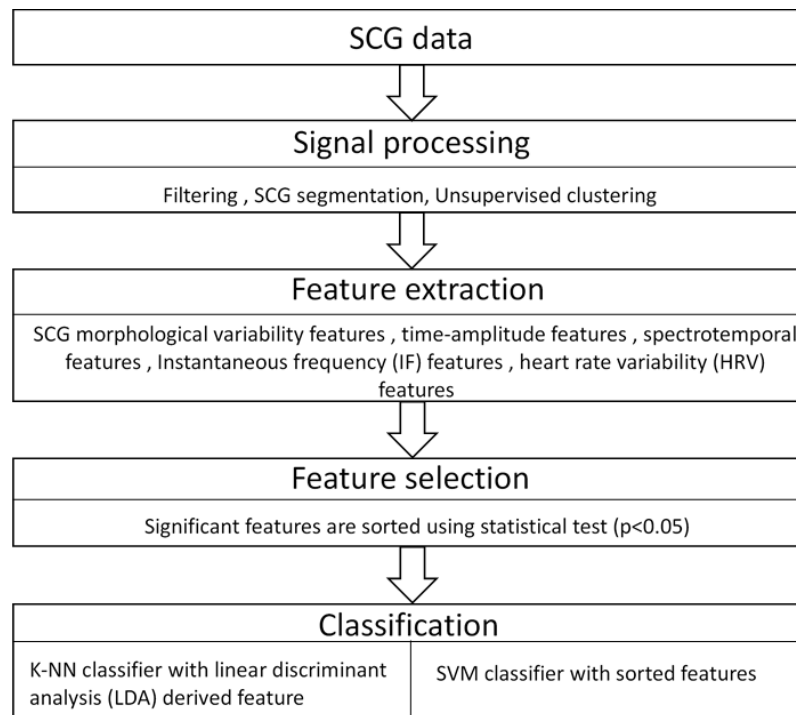


Figure 47-Methodology

Fig. 47 summarizes the methodology employed in the study which is further discussed in the following sections.

Study protocol

After Institutional Review Board (IRB) approval, HF patients were enrolled at their discharge from hospital. For the study, only the HF patients with reduced ejection fraction ($< 45\%$) were considered. These patients were longitudinally studied within a 6-month period from their initial discharge. The ongoing study has collected data from 41 patients and among them 8 patients were readmitted while 13 patients were not re-admitted during the 6-months period after initial discharge and rest are in the process of data collection. The patients with pacemakers (but who were not actively paced during the measurements) were included in the study. Table 3 shows the distribution of

data recording sessions categorized in time for re-admitted and not re-admitted HF patients. The longitudinal distribution of data acquisition is shown in Fig. 48.

Table 3- Data recording sessions longitudinally categorized in time for re-admitted and not re-admitted patients from data gathered in a 6 months duration after initial discharge.

Readmitted (n=13)				Not Readmitted (n=8)					
Days>62 (2 months before re- admission)	62>days>31 (Within 1-2 months before re- admission)	31>days>2 (Within 2 days-1 month before readmission)	Readmission (Within 2 days before or after re- admission)	Discharge (At the day of initial discharge)	days<31 (Within 1 month after discharge)	31<days<62 (Within 1-2 months after discharge)	62<days<93 (Within 2-3 months after discharge)	93<days (3 months after discharge)	All
6	7	12	6	8	11	7	3	4	63
Total=31				Total=32					

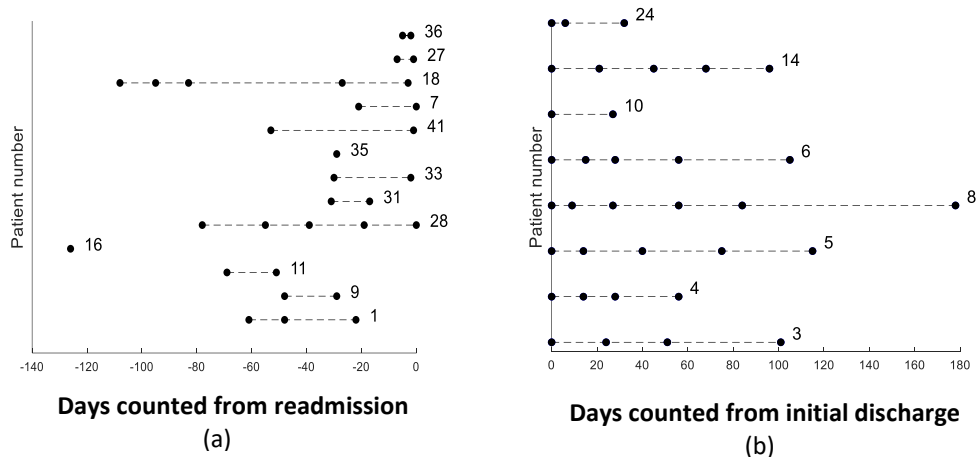


Figure 48- Data session distribution of (a) readmitted and (b) non readmitted patients. Numbers represent the numbers assigned

Data acquisition

Data acquisition was done when patients were sitting on a 45-degree inclined exam table with their legs extended. Seismocardiographic signal was acquired using an accelerometer (Model: 356A32, PCB Piezotronics, Depew, NY) which was affixed on the chest surface at the 4th intercostal space near the left lower sternal border using double-sided medical grade tape. The accelerometer signal was amplified using a signal

conditioner (Model: 482C, PCB Piezotronics, Depew, NY) with a gain of 100-fold. The dorso-ventral component of the acceleration is focused in the current study. ECG signal was acquired simultaneously using a IX-B3G bio-potential recorder (iWorx Systems, Inc., Dover, NH). In each measurements session data were acquired approximately for 2 minutes at a sampling rate of 10 kHz. Fig. 49 shows a diagram of the data acquisition setup.

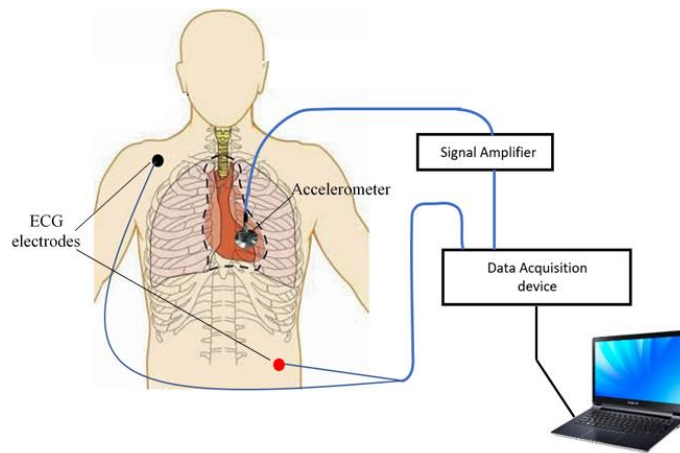


Figure 49- Data acquisition setup.

Pre-processing of SCG data

Raw SCG and ECG signals were forward-backward filtered to reduce background noise and baseline wandering due to respiration using a 4th order Chebyshev 2 type band-pass filter with a cut-off of 0.5-50 Hz (Peshala T. Gamage et al., 2020). After filtering, SCG signal was segmented into SCG events (also called heartbeats) using the R peaks locations of the ECG signal, which were detected using Pan Tomkins algorithm (Tompkins, 1985). The start location of each SCG event was selected 0.1 seconds before the R peak of the corresponding ECG and the end location at 0.1 seconds before the proceeding R peak (i.e., see Fig. 6).

Clustering of SCG morphology for accurate feature extraction

SCG is known to have high morphological variability which may impede the accuracy of feature extraction (Azad et al., 2019; Peshala T. Gamage et al., 2020; R. H. Sandler et al., 2019). Previous studies (Peshala T Gamage et al., 2018; Peshala T. Gamage et al., 2020) have reported clusters with different morphological patterns with correlations to respiratory phases and suggested that SCG morphology can be optimally divided into two clusters which help to reduce the morphological variability of SCG allowing for more accurate extraction of SCG morphological features. Hence, prior to extracting features from the segmented SCG events, unsupervised clustering was used to reduce the morphological variability of the SCG signal. The current study employed k-medoid clustering with dynamic time warping (DTW) to cluster the SCG events into two groups as suggested in previous studies (Azad et al., 2019; Peshala T. Gamage et al., 2020). After clustering, a representative SCG event was calculated by averaging the 15% of the events closest to the medoid of each cluster. Here, the averaging was used to further reduce the noise of the waveforms and 15% of events allowed averaging of (10 ± 3) number of events in each cluster for each patient. When averaging, the length of the events was truncated to the length of the shortest event in the group. However, since the closest events to the medoid had similar lengths (i.e., similar heart rate), the truncated parts of the events were insignificant (± 15 ms) and had minimum effect on the SCG waveform features. Also, due to the similar heart rate and similar morphologies of these events, averaging had minimal effect on the SCG waveform features which is otherwise expected in the case of significant misalignments between SCG features (i.e., SCG peaks) (H. A. M. Sandler et al., 2015). Fig. 50 shows an example of SCG events closest to each medoid and their respective averages.

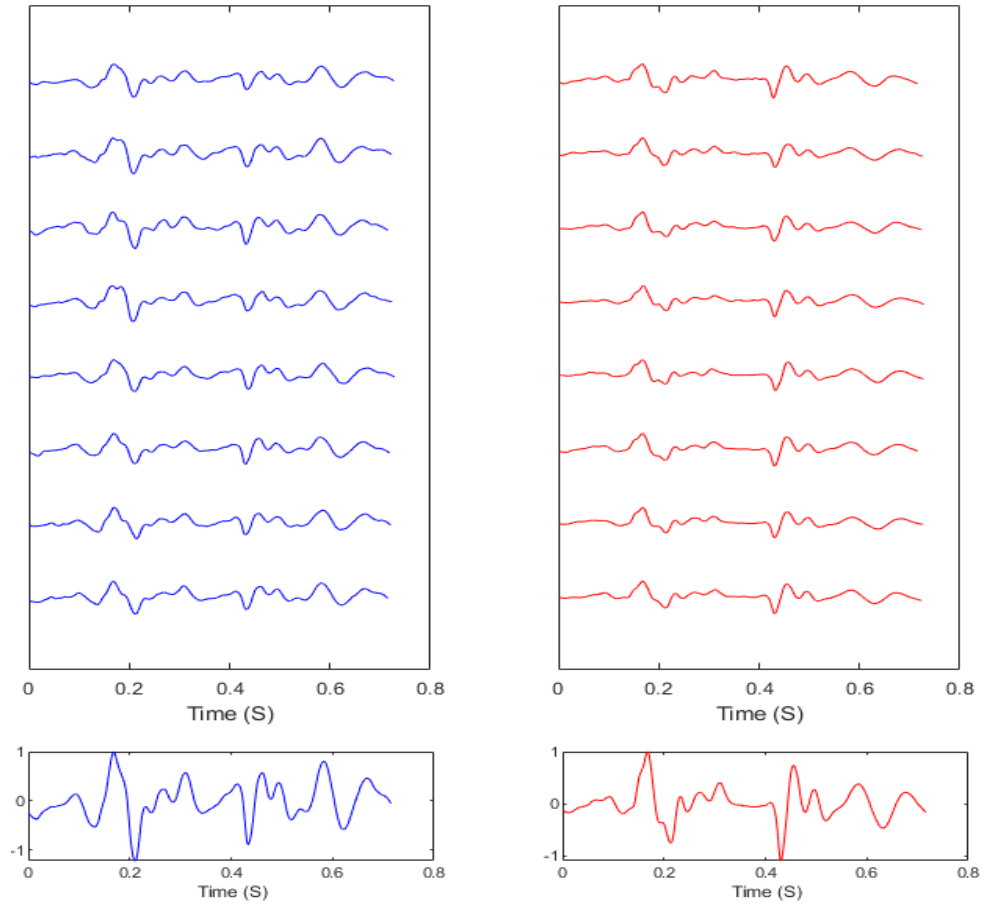


Figure 50- Events closest to the medoid with their average waveforms (from same session)

SCG Features

Overall, 103 total features were calculated. These contained 3 SCG morphological variability features, 9 HRV features and 92 SCG features derived from the average SCG waveforms of two clusters. These 92 SCG features were formed by the average and the difference of 46 SCG features calculated for two SCG clusters as denoted in equations 21 and 22, respectively.

$$Feature_{ave} = \frac{Feature_{cluster1} + Feature_{cluster2}}{2} \quad (21)$$

$$Feature_{diff} = \frac{abs(Feature_{cluster1} - Feature_{cluster2})}{\min(Feature_{cluster1}, Feature_{cluster2})} \quad (22)$$

The description of the features are presented in Table 4.

Table 4- Feature description

<i>Feature type</i>	<i>Feature name</i>	<i>Description</i>
Waveform variability	Variability before clustering	Average dissimilarity between SCG events with their medoid before clustering
	Intra cluster variability	Average dissimilarity between SCG events with their own medoid
	Inter cluster variability	Average dissimilarity of SCG events with the other cluster medoid
Time-Amplitude	SCG1 amp	rms amplitude of a 125 ms window centered around SCG1 location
	SCG2 amp	rms amplitude of a 125 ms window centered around SCG2 location
	SCG3 amp	rms amplitude of a 125 ms window centered around SCG3 location
	SCG1-SCG2 amp	rms amplitude of the window starting 62.5 ms after SCG1 and ending 62.5 ms before SCG2 locations
	SCG1 amp aud	Audible (i.e., 20-50 Hz) rms amplitude of a 125 ms window centered around SCG1 location
	SCG2 amp aud	Audible rms amplitude of a 125 ms window centered around SCG2 location
	SCG3 amp aud	Audible rms amplitude of a 125 ms window centered around SCG3 location
	SCG1-SCG2 amp aud	Audible rms amplitude of the window starting 62.5 ms after SCG1 and ending 62.5 ms before SCG2 locations
	SCG_PEP	ECG Q –SCG1 time interval
	SCG_LVEP	SCG1-SCG2 time interval
	TPR	Turning Point Ratio (TPR) (see Eqn 41)
	pk-pk	Peak to peak amplitude
	rms_amp	Root mean square (rms) amplitude
	En	Spectral energy between frequency range (0.5-50 Hz)
	ER (f1-f2/f3-f4)	Spectral energy ratio between frequency range (f1-f2 Hz) and (f3-f4 Hz) $ER(0.5-5/0.5-15), ER(0.5-20/20-50), ER(0.5-10/0.5-50), En(10-20/0.5-50), ER(20-30/0.5-50), ER(30-40/0.5-50), ER(40-50/0.5-50)$ were calculated. (see Eqn 26)
	F_avg (f1,f2)	$F_{ave}(0.5,50), F_{ave}(0.5-5), F_{ave}(5-15), F_{ave}(0.5,10), F_{ave}(10,20), F_{ave}(20,30), F_{ave}(30,40), F_{ave}(40,50)$ were calculated. (see Eqn 27)
	F_peak	Frequency at FFT peak
	FFT_peak_amp	Peak FFT amplitude
	FFT_sharpness	$F_{peak}/(F1-F2)$, where F1 and F2 are the frequencies at the amplitude $(FFT_{peak_amp}/\sqrt{2})$ and $(F1>F2)$
	SpEn	Spectral Entropy (see Eqn 30)
Spectro-Temporal using polynomial chirplet transform (PCT)	PCT_ER_SCG1	PCT Energy ratio between frequency range (0.5-5Hz) and (0.5-15 Hz) in the 125 ms window centered around SCG1 location
	PCT_ER_SCG2	PCT Energy ratio between frequency range (0.5-5Hz) and (0.5-15 Hz) in the 125 ms window centered around SCG2 location
	PCT_ER_SCG3	PCT Energy ratio between frequency range (0.5-5Hz) and (0.5-15 Hz) in the 125 ms window centered around SCG3 location
	PCT_ER_SCG1-SCG2	PCT Energy ratio between frequency range (0.5-5Hz) and (0.5-15 Hz) in the 125 ms window starting 62.5 ms after SCG1 and ending 62.5 ms before SCG2 locations
Instantaneous frequency (IF)	F_SCG1	maximum instantaneous frequency near SCG1
	F_SCG2	maximum instantaneous frequency near SCG2
	T_SCG1_max_freq	Timing at F _{sy}

Heart Rate Variability (HRV)	T_SCG2_max_freq	Timing at Fdia
	Frms	Root mean square amplitude of the instantaneous frequency signal
	SDNN	Standard deviation of ECG R-R intervals
	PNN50	Percentage of successive RR intervals that differ by more than 50 ms
	VLF	Spectral power of all RR intervals in 0.003–0.04 Hz
	LF	Spectral power of all RR intervals in 0.04–0.15 Hz
	HF	Spectral power of all RR intervals in 0.15–0.4 Hz
	VHF	Total spectral power of all normal sinus RR intervals 0.4–1.0 Hz
	TP	Total spectral power of all normal sinus RR intervals 0–0.4 Hz
	LF/HF ratio	The ratio of LF to HF
	Ave_HR	Average of mean heart rates of cluster1 and cluster2

SCG waveform variability features

SCG waveform variability features describe the morphological similarity (or dissimilarity) of the SCG events. Here, three different variability features were calculated. One feature before clustering and two after clustering the SCG events in to two groups such that they have optimum similarity within groups (Peshala T. Gamage et al., 2020). The variability before clustering represents the intra session beat to beat morphological variability of the SCG signal. The intra cluster variability is a measure of the similarity within clusters while inter cluster variability is a measure of similarity between the two clusters. These features are denoted in equations 23, 24 and 25. Similar variability calculations can be found in the previous studies (Azad et al., 2019; Peshala T. Gamage et al., 2020).

$$\text{Variability before clustering} = \frac{1}{n_1+n_2} \left[\sum_{i=1}^{n_1+n_2} \frac{dtw(C, X_i)}{l_i} \right] \quad (23)$$

$$\text{Intra cluster variability} = \frac{1}{n_1+n_2} \left[\sum_{i=1}^{n_1} \frac{dtw(C_1, X_{i1})}{l_i} + \sum_{i=1}^{n_2} \frac{dtw(C_2, X_{i2})}{l_i} \right] \quad (24)$$

$$\text{Inter cluster variability} = \frac{1}{n_1+n_2} \left[\sum_{i=1}^{n_1} \frac{dtw(C_1, X_{i2})}{l_i} + \sum_{i=1}^{n_2} \frac{dtw(C_2, X_{i1})}{l_i} \right] \quad (25)$$

where, X_{i1} , X_{i2} are the i^{th} SCG event belonging to cluster 1 and cluster 2, respectively while C_1 and C_2 are the respective cluster medoid event. C is the medoid event before clustering and n_1, n_2 are the total number of events belong to cluster 1 and

2, respectively. The function *dtw* calculates the DTW distance (i.e., morphological dissimilarity) between two events and the DTW distance is divided by DTW warping path length l_i to normalize the calculated DTW by signal length. Also, SCG event amplitude is normalized by its peak to peak amplitude prior to variability calculations. This type of normalization in signal length and amplitude allows to compare the morphological variability features between different subjects and sessions.

Spectral features

Previous studies have focused on the study of the changes in the intensity and frequency of sub-audible precordial vibrations on human and animals (Foulger, Smith Jr, & Fleming, 1947, 1948) and found that anoxia or other deleterious agents in blood can lower the average frequency of vibrations (based on intensity) in sub-audible frequencies (Johnston & Overy, 1951). Moreover, the study (Foulger et al., 1947) suggested that these sub-audible frequencies show a shift towards lower frequencies in patients with serious heart conditions in resting state. The analysis of sub-audible spectrum may provide important information about myocardium function (Johnston & Overy, 1951). In the current study, to extract the possible energy and frequency shifts in the spectral domain, several features which calculate the spectral energy ratios between different frequency intervals in both audible and sub-audible domains were calculated. The average frequencies of different frequency bands were calculated. Equations 26 and 27 denote the calculation of spectral energy ratios between different frequency bands of $(f_1-f_2)/(f_3-f_4)$ and average frequency of a defined frequency interval (f_1-f_2) .

$$ER = \frac{\int_{f_1}^{f_2} \text{abs}(FFT(f))^2 df}{\int_{f_3}^{f_4} \text{abs}(FFT(f))^2 df} \quad (26)$$

$$F_{avg} = \frac{\sum_{f_1}^{f_2} f_i \cdot FFT(f_i)}{\sum_{f_1}^{f_2} FFT(f_i)} \quad (27)$$

where f, FFT denote the frequency and fast fourier transformation function, respectively. Information about considered different frequency bands are shown in Table 4. In addition to these, 4 more spectral features, namely, FFT peak amplitude (FFT_peak_amp) , frequency at peak FFT amplitude (F_peak) , FFT sharpness factor (FFT_sharpness) and spectral entropy (SpEn) were defined. Description of these features are given in Table 4. While FFT_peak_amp, F_peak, FFT_sharpness are identified as potential features for identifying normal and unhealthy heart conditions in an early study (Koiwa et al., 1991) which measured cardiac related vibrations on precordium, SpEn is used as a feature in a recent study for classification of heart conditions using precordial accelerations (Lahdenoja et al., 2017). The following procedure (Shen, Hung, & Lee, 1998) is used for calculating SpEn for a signal $x(n)$. Here, $x(n)$ is a SCG event in time domain where n denotes a sample point in time. The power spectrum of $x(n)$ can be denoted as shown in equation 28.

$$S(m) = |X(m)|^2 \quad (28)$$

In equation 28, $X(m)$ is the discrete Fourier transform of $x(n)$ where m denoted a frequency point in spectral domain. Then, the probability distribution $P(m)$ is defined as,

$$P(m) = S(m) / \sum_{i=1}^N S(i) \quad (29)$$

and the spectral entropy (SpEn) is defined as,

$$SpEn = - \sum_{m=1}^N P(m) \log_2 P(m) \quad (30)$$

where N is the total frequency points.

Spectro-temporal features

Spectro-temporal features were extracted by analyzing the time frequency distribution (TFD) of the SCG signal. TFD allows the study of a signal in both time and frequency domain simultaneously which helps the more comprehensive analysis of time varying frequency (or instantaneous frequency (IF)) in the SCG signal. Although, there are no previous studies which used TFD to extract spectro-temporal features of SCG to classify cardiac conditions, TFD analysis has proven to be an effective method of extracting spectro-temporal features to classify cardiac conditions using ECG (Rad et al., 2017; Zhao, Särkkä, & Rad, 2018). In TFD literature, several well-known forms of methods such as short time Fourier transform (STFT), wavelet Transform (WT) and chirplet transform (CT) are available. The study (Amirtaha Taebi & Mansy, 2017b) provided an extensive comparative analysis of different methods for obtaining the TFD of SCG signal and concluded that PCT is more suited for estimating IF of SCG signals due to its consistently in accurate IF estimations. Hence, PCT is used to obtain the TFD of SCG signals in the current study.

Polynomial Chirplet Transform (PCT) is an improved version of the chirplet transform (CT) developed to provide better performance when used with signals containing non-linear instantaneous Frequency (IF) trajectory (Z. Peng et al., 2011). Equation 31 describes the CT transformation if a signal $x(t)$,

$$\bar{X}_{CT}(t_0, \omega, \alpha, \sigma) = \int_{-\infty}^{+\infty} \bar{z}(t) \omega_{(\sigma)}(t - t_0) e^{-j\omega t} dt \quad (31)$$

where t_0, ω, α denote time, frequency and chirp rate, respectively. The analytical signal of $x(t)$ is defined as $z(t)$ (i.e., using Hilbert transform H , $(z(t) = x(t) + jH[x(t)])$). The function $\omega_{(\sigma)}$ is a symmetric, non-negative and normalized window which usually considered as a gaussian window as denoted in equation 32.

$$\omega_{(\sigma)}(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-0.5\left(\frac{t}{\sigma}\right)^2\right) \quad (32)$$

The normalized function analytical signal $z(t)$ is defined as,

$$\bar{z}(t) = z(t)\varphi_{\alpha}^R(t)\varphi_{\alpha}^M(t, t_o) \quad (33)$$

where $\varphi_{\alpha}^R(t)$ and $\varphi_{\alpha}^M(t, t_o)$ denote the frequency rotating operator and frequency shift operator, respectively. The operator $\varphi_{\alpha}^R(t)$ rotates $z(t)$ by an angle of $\tan^{-1}(-\alpha)$ while the operator $\varphi_{\alpha}^M(t, t_o)$ shifts the frequency of $z(t)$ from ω to $\omega + \alpha t_o$ as indicated in equations 34 and 35.

$$\varphi_{\alpha}^R(t) = \exp\left(-\frac{j\alpha t^2}{2}\right) \quad (34)$$

$$\varphi_{\alpha}^M(t, t_o) = \exp(j\alpha t_o t) \quad (35)$$

In contrast to conventional CT, PCT is improved by defining aforementioned higher order non-linear operators in polynomial form. The modified operators are shown in equations 36 and 37.

$$\varphi_{\alpha_1, \alpha_2, \dots, \alpha_n}^R(t) = \exp\left(-j \sum_{k=2}^{n+1} \frac{1}{k} \alpha_{k-1} t^k\right) \quad (36)$$

$$\varphi_{\alpha_1, \alpha_2, \dots, \alpha_n}^M(t, t_o) = \exp\left(-j \sum_{k=2}^{n+1} \alpha_{k-1} t_o^{k-1} t\right) \quad (37)$$

In cooperating the modified higher order non-linear operators PCT transformation is denoted in equation 38. In detail description of the PCT algorithm is presented in the study (Z. Peng et al., 2011).

$$\begin{aligned} \bar{X}_{PCT}(t_o, \omega, \alpha_1, \dots, \alpha_n, \sigma) = & \int_{-\infty}^{+\infty} z(t) \varphi_{\alpha_1, \alpha_2, \dots, \alpha_n}^R(t) \varphi_{\alpha_1, \alpha_2, \dots, \alpha_n}^M(t, t_o) \dots \\ & \dots \times \omega_{(\sigma)}(t - t_o) e^{-j\omega t} dt \end{aligned} \quad (38)$$

Fig. 50 shows an example of the PCT TFD distribution for a SCG heart cycle acquired from a patient. The high energy regions in SCG1 and SCG 2 shown in Fig. 50

correspond to S1, S2 heart sounds of phonocardiography (PCG) which is primarily generated due to the valve movements of the heart [ref]. The fluctuations from the heart surface can travel through the thorax to the chest surface in the form of two different type of waves: namely, the slower, high amplitude shear waves and faster low amplitude pressure waves. Here, the pressure waves can travel through the thorax with speed in order of 1000 m/s and shear wave with a speed in the order of 10 m/s (Bombardini et al., 2007; Foulger et al., 1947). In solid mediums, shear waves travel by deforming the medium with a much higher amplitude than the pressure waves (Bombardini et al., 2007). Unlike a microphone, (used in PCG) which is designed to measure audible frequencies (>20 Hz), the current accelerometer (Model: 356A32, PCB Piezotronics, Depew, NY) is sensitive to both audible and infrasonic (< 20 Hz) frequencies. Hence, the higher energy content in SCG1 and SCG 2 regions observed in the infrasonic frequency range (i.e., <20 Hz) can be explained by the fact that the accelerometer is sensitive to high amplitude, low frequency shear waves transmitted through the thorax. In contrast to PCG which detects audible (i.e., >20 Hz) S1 and S2 sounds, SCG allows the analysis of these sounds in the infrasonic range that may help provide additional diagnostic features.

As seen in Fig. 51, in addition to SCG 1 and SCG 2 regions, another region with high energy can be observed after SCG 2 region. This high energy region was identified as SCG 3 which corresponds to the third heart sound S3 described in previous literature (Siejko, Thakur, Maile, Patangay, & Olivari, 2013). S3 is described as a low-frequency brief vibration detected in the phonocardiography (PCG) signal, occurring in early diastole at the end of the rapid diastolic filling period of the right or left ventricle (Ozawa et al., 1983a; Ozawa, Smith, & Craige, 1983b).

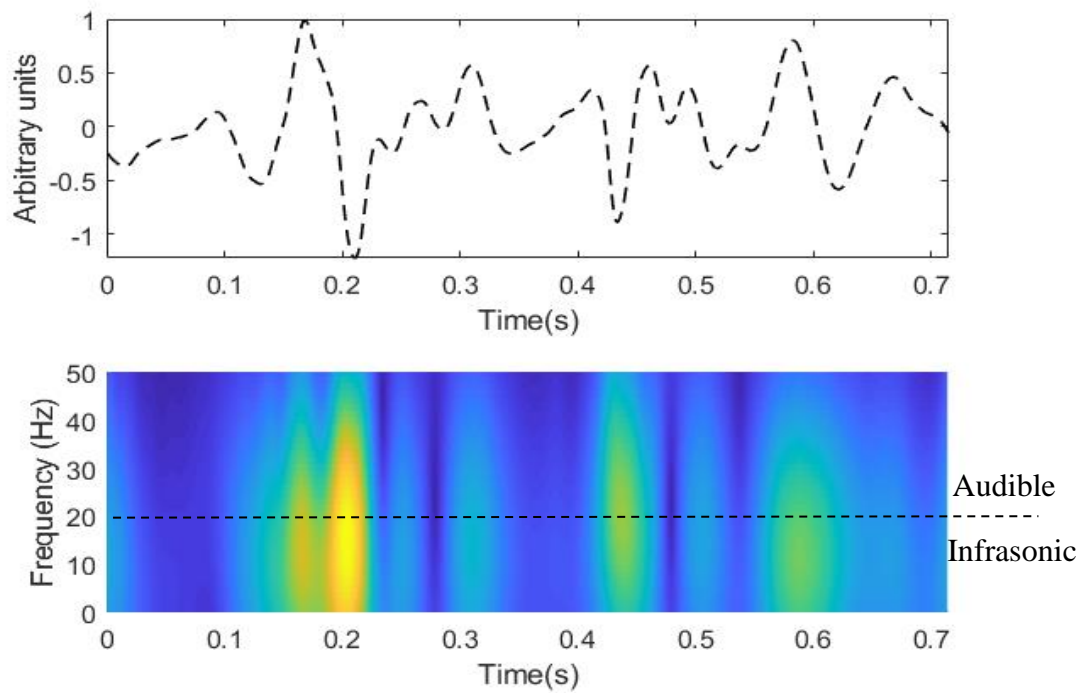


Figure 51- Top: SCG waveform acquired from a patient, bottom: Time frequency distribution of SCG generated by PCT with identified high energy regions

S3 is regarded as a poor prognostic sign found in patients with failing heart (Ozawa et al., 1983b). Several studies (Drazner, Rame, Stevenson, & Dries, 2001; Patel, Bushnell, & Sobotka, 1993) analyzed the correlation between s3 and cardiac pathologies. Hence, the current study also included the time-frequency features in SCG 3 region. A

summary of spectro-temporal features extracted from the TFD distribution of SCG using PCT is shown in Table 4.

Locating the timing of SCG1, SCG2 and SCG3 and feature derivation

Timing of the SCG1, SCG2 and SCG3 were determined by locating the high energy regions in corresponding frequency bands of the PCT distribution. This was done by calculating an energy signal for a defined frequency band and locating its peaks. Equation 17 represents the energy signal derived from PCT distribution for the frequency range $f1$ to $f2$.

$$E_{PCT}(t) = \int_{f_1}^{f_2} \bar{X}_{PCT}(f, t) df \quad (39)$$

Since SCG1 and SCG2 peaks are defined corresponding to S1 and S2 heart sounds, their timings were located by finding the peaks of the energy signal in the audible frequency band (20Hz-50Hz). As heart sound S3 is known to be distributed over both audible and infrasonic frequency ranges with higher energy in infrasonic range (Ozawa et al., 1983b; Siejko et al., 2013), SCG3 which corresponds to heart sound S3 was located by finding the peaks of the energy signal in the frequency band (0.5-50 Hz). Here, SCG3 search window was refined to the region 100-200 ms after the detected SCG2 location considering the timing of S3 described in previous literature (Ozawa et al., 1983b; Siejko et al., 2013). A similar procedure for detecting heart sounds S1, S2, S3 using accelerometer-based measurements is presented in the previous study (Siejko et al., 2013). Fig. 52 shows the detected SCG1, SCG2 and SCG3 for two different SCG waveforms.

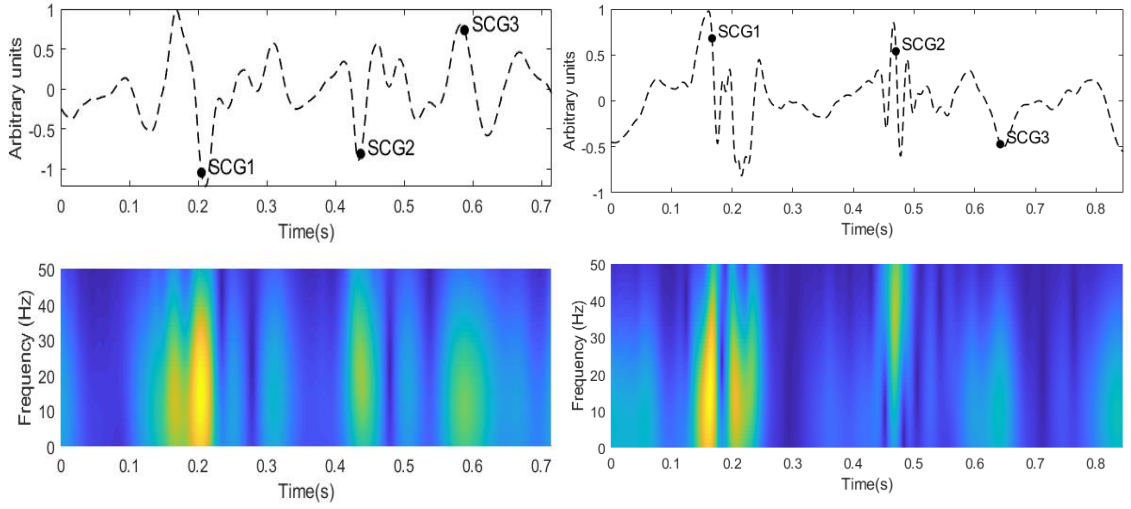


Figure 52- Examples of two SCG waveforms with SCG1, SCG2 and SCG3 locations

After finding the locations of SCG1, SCG2 and SCG2, energy ratio between frequency range (0.5-5Hz) and (0.5-15 Hz) in the 125 ms window centered around each location were calculated as features. These frequency ranges were chosen similar to the energy ratio feature described earlier which analyzes the low frequency shift in sub-audible frequencies A description of features calculated using PCT energy distribution is shown in Table 4 under spectro-temporal features. Equation 18 describes the calculation of the energy ratio features using PCT distribution.

$$PCT_ER = \frac{\int_{t_1}^{t_2} \int_{0.5}^5 \bar{X}_{PCT}(f,t) df dt}{\int_{t_1}^{t_2} \int_{0.5}^{15} \bar{X}_{PCT}(f,t) df dt} \quad (40)$$

where t_1 , t_2 denote the start time and end time of the considered time window (as stated in Table 4). Note that if any SCG 3 peak is not detected (i.e., absence of 3 rd. heart sound), SCG 3 location was defined at 150 ms after the SCG2 location for feature analysis purpose.

Time-amplitude features

Time-amplitude features included the rms amplitudes in defined time window of 125 ms in the SCG waveform centered at the locations of SCG1, SCG2 and SCG3. Two-time interval features, namely; SCG pre ejection period (SCG_PEP) and SCG left ventricular ejection period (SCG_LVEP) were calculated. Similar amplitude and time interval-based features have been previously studied (Siejko et al., 2013; Amirtahà Taebi et al., 2019). Peak to peak amplitude and rms amplitude of the SCG event were also included in time-amplitude features. A description of time-amplitude features is available in Table 4. In addition to these, another feature called turning point ratio (TPR) was calculated. TPR derived from precordial acceleration and ECG signals has proven to be an accurate feature when classifying between normal subjects and HF patients with atrial fibrillation (Dash, Chon, Lu, & Raeder, 2009; Lahdenoja et al., 2017). TPR is defined in equation 41,

$$TPR = RD(x)/N \quad (41)$$

where RD is a function which counts the total number of consecutive increasing and decreasing runs (i.e. count if $(x(n_2) - x(n_1)) \times (x(n_3) - x(n_2)) < 0$, provided $n_1 < n_2 < n_3$) in signal $x(n)$. N is the total number of points in x .

Instantaneous frequency SCG features

The instantaneous frequency (IF) is a transient parameter that relates to the average of the frequencies present in a signal as the signal evolves in time (Boashash, 1992). In the current study the instantaneous frequency of the SCG heart cycles were calculated to extract features. Here, the instantaneous frequency ' f_{ins} ' was calculated using equation 42.

$$f_{ins}(t) = \frac{\int_{0.5}^{50} f \cdot PCT(t, f) df}{\int_{0.5}^{50} PCT(t, f) df} \quad (42)$$

where, f is the frequency and $PCT(t, f)$ is the power in the time frequency distribution using PCT.

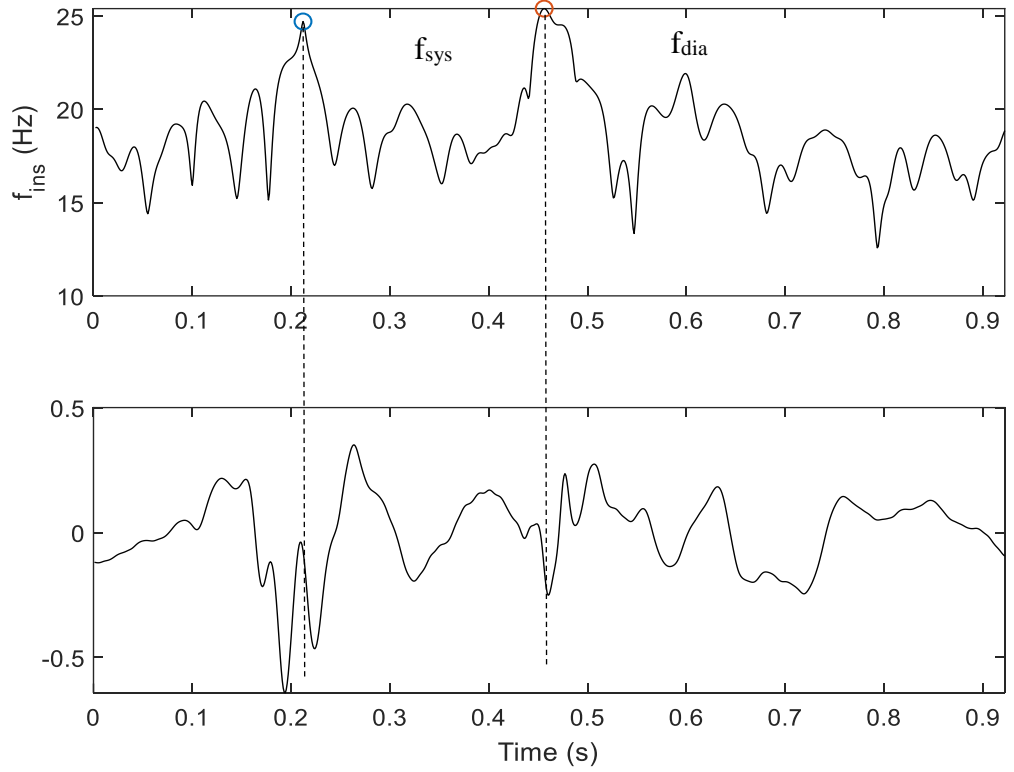


Figure 53- (top): Instantaneous frequency signal with identified features (bottom): SCG heart cycle

Fig. 53 shows an example of the IF frequency signal and the corresponding SCG heart cycle. As shown in Fig. 53, Two high frequency peaks around SCG1 and SCG2 regions are observed. Under the hypothesis that these high frequencies are possibly related to valve opening and closing at the beginning of systole and diastole, these frequencies were named as f_{sys} and f_{dia} . These frequencies were detected by selecting the maximum IF frequency in a 150 ms window centered around SCG1 and SCG2

locations, respectively. The timing at these frequencies were also considered as features. In addition, the rms amplitude of the IF signal was considered as a feature.

Heart rate variability features

In addition to the SCG based features, eight common Heart Rate Variability (HRV) features were also analyzed. The changes in HRV features have shown high correlations when classification between normal and HF subjects (Isler, 2016; Liu et al., 2014; Shahbazi & Asl, 2015) as well as when analyzed with mortality rate (Kleiger, Miller, Bigger Jr, & Moss, 1987). The current study is focused on evaluating and comparing the performance of these features when predicting the re-admission of HF patients with the potential SCG features. HRV features are described in Table 4.

Feature Analysis and Selection

The features can be analyzed longitudinally (counted as days before re-admission or after discharge) or group wise (i.e., features in re-admitted (R) vs not readmitted (nR) groups). While the longitudinal analysis can help understand useful trends of how different features change in time after HF patients are discharged from the hospital, accuracy of such analysis is limited by irregular time intervals between the data sessions acquired from different patients. Fig. 54 shows an example of longitudinal analysis of a feature trajectories of each patient. In Fig. 54, each point represents a feature (i.e., SCG variability before clustering) extracted from a data session in time. Feature points are also connected using a dashed line to link the features of the same patient

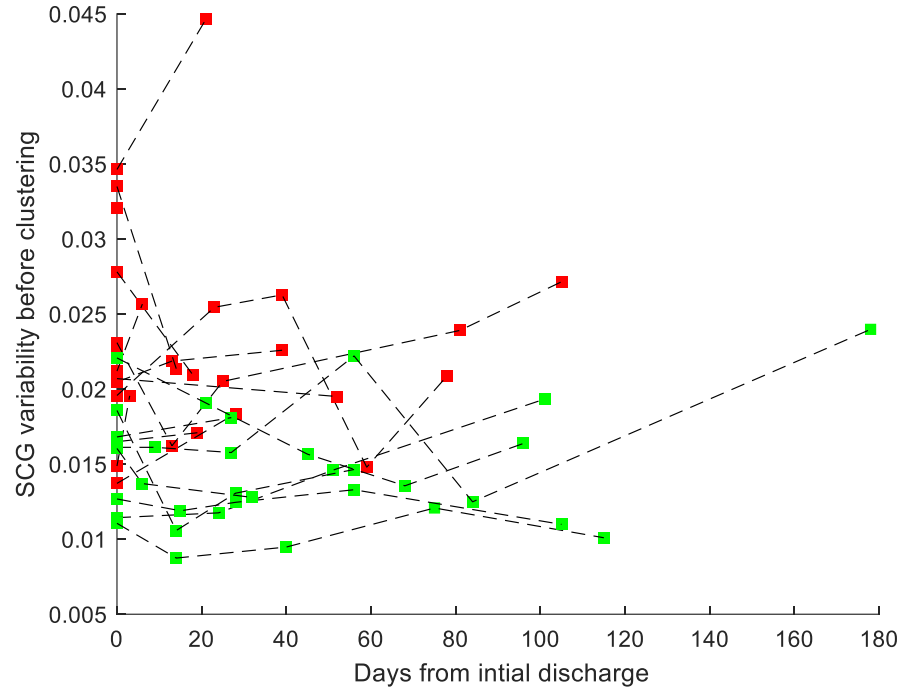


Figure 54- An example of longitudinal variation of a feature in each patients. Red: re-admitted, Green: non readmitted. Data points belongs to each patient is linked with a dashed line.

In Fig. 54, a generalized trend of the trajectories for each patient towards readmission or after discharge is not prominent as the feature values are fluctuating (within a range for majority of subjects). However, feature values between re-admitted and non-re-admitted HF patients are distributed in different value ranges (i.e., 0.075-0.225 for non-re-admitted and 0.015-0.045 for re-admitted). Hence, when analyzing features, the attention was given to the differences between the feature values of re-admitted and non-readmitted groups. Here, following two groups were defined for analyzing the feature value differences for feature extraction (or sorting).

1. **R**: Readmitted (data sessions within 60 days of re-admission of readmitted HF patients)
2. **nR**: not Readmitted (data sessions within 60 days from initial discharge of non-readmitted HF patients)

These groups were formed under the hypothesis that, a subject is likely to show bad cardiac mechanical activity within a 2 month period before the re-admission and a subject is likely to show relatively better cardiac mechanical activity within 2 months after initial discharge (also provided that the subject is not re-admitted within 6 months after readmission). Such classification can help clinicians to predict the likelihood of a HF patient being re-admitted within 2 months after initial discharge based on the SCG data recordings acquired when patients come for their follow up checkups after initial discharge. Data sessions in R and nR groups are highlighted in Fig. 55.

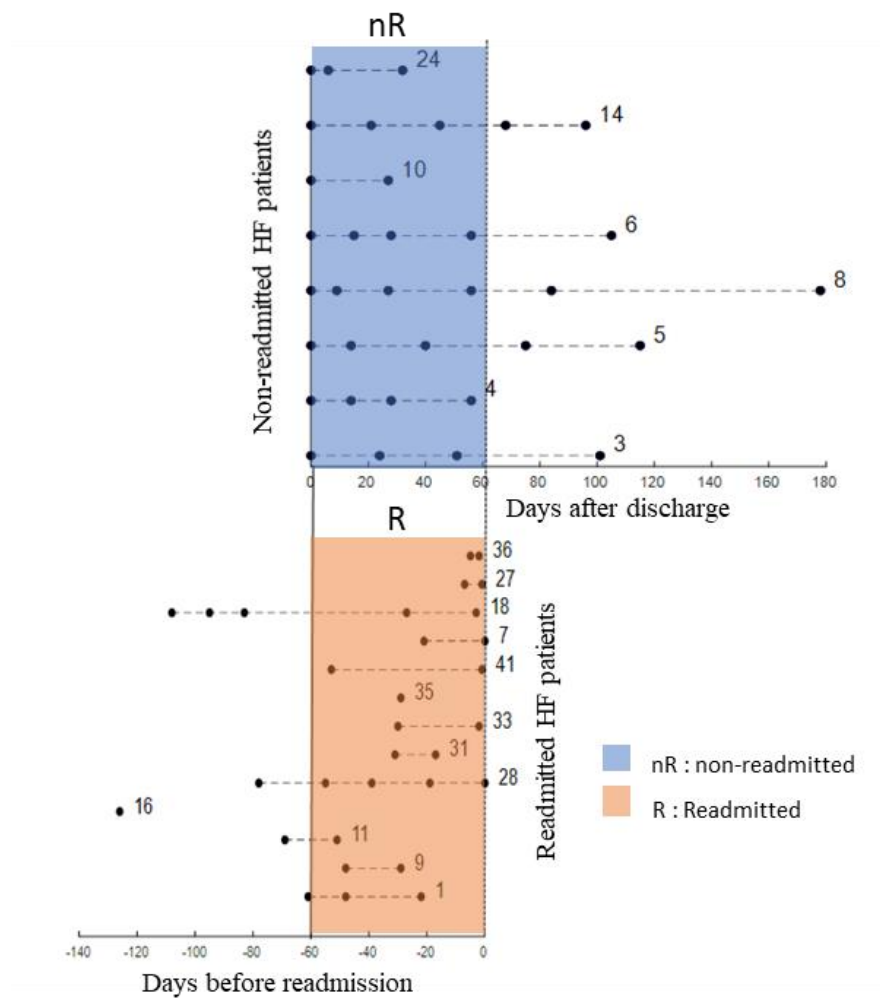


Figure 55- Highlighted regions defining R and nR groups for feature analysis.

Feature extraction was done by sorting the features which have significant differences between nR and R groups using statistical measures. Provided that the statistical analysis assumes the features belong to two independent distributions (i.e., R and nR), independent statistical test (i.e., unpaired t-test) was used to sort features with significance differences between these groups.

Fig. 56 shows an example of analyzing the difference between the feature points lumped into groups R and nR . In addition, features extracted from another group ‘H’ which contains data acquired from healthy subjects with no previous cardiac health issues are presented. Group H was introduced for comparison purposes and validation of the observed differences between R and nR (i.e., nR distribution is expected to be closer to H).

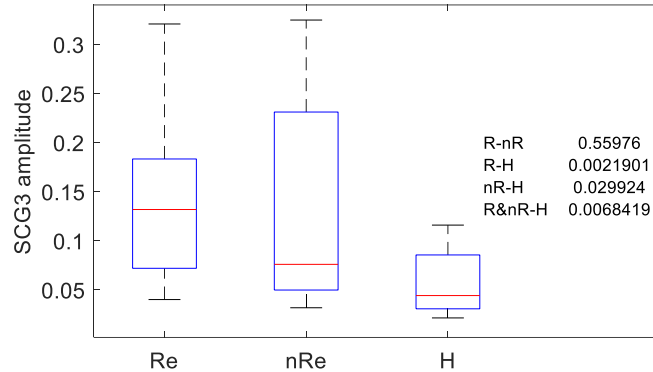


Figure 56- An Example of analyzing the differences of a feature value (SCG3 amplitude) distributions between R and Re groups.. Features from a healthy group ‘H’ is also included for comparison purposes. Distributions are represented by box and whisker plots and the p values between the groups are presented. SCG3 amplitude shows significant differences ($P < 0.05$) between H and the HF patient groups (i.e., R and nR) while no significance difference is observed between R and nR. Such observation for SCG3 amplitude is consistent with the previous study (Siejko et al., 2013)

Minimum redundancy maximum relevance algorithm

In addition, Minimum Redundancy Maximum Relevance (MrMR) algorithm (H. Peng, Long, & Ding, 2005) was also used for sorting good features for classification

of R and nR groups. MrMR algorithm selects a subset of the input feature vector containing features which have the most correlation with a class (i.e., R or nR) while having the least correlation between the features. Here, the correlation between a feature and the class is the ‘relevance’ while the correlation between a feature and other features is the ‘redundancy’. MrMR ranks the features based on minimal-redundancy-maximal-relevance criteria.

In MrMR algorithm, mutual information I between two variables X, Y is defined as,

$$I(x, y) = \sum_{i,j} P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i) \cdot P(y_j)} \quad (43)$$

The objective of MrMR algorithm is to find an optimal set S of features which maximizes the relevance D between the feature variable x and class c and minimizes the redundancy R between the features x and y .

$$\text{Maximize, } D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i, c) \quad (44)$$

$$\text{Minimize, } R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \quad (45)$$

Above criteria is achieved by optimizing the following equation,

$$\text{Maximize, } \varphi = D - R \quad (46)$$

The algorithm delivers the ranked the features sorted in descending order based on their φ value (H. Peng et al., 2005).

Principal component analysis for visualizing the features

As a starting point, before inputting the features into machine learning (ML) classifiers, Principal component analysis (PCA) was used to visualize whether the variability of the extracted features can potentially help differentiate between the classes R and nR. PCA is a method used to visualize strong patterns in a dataset by using the inherent variability of the dataset. In Machine learning, PCA is often uses to

visualize the features and to reduce the dimensionality of the feature vector by combining features. PCA transforms data (i.e., features) into a new coordinate system by an orthogonal linear transformation such that the first coordinate of the orthogonal coordinate system (i.e., first principle component) is generated by a scalar projection of the maximum variance of the original data set. Similarly, the second orthogonal coordinate is generated by a projection of the second largest variance of the dataset and so on (e.g., see Fig. 57).

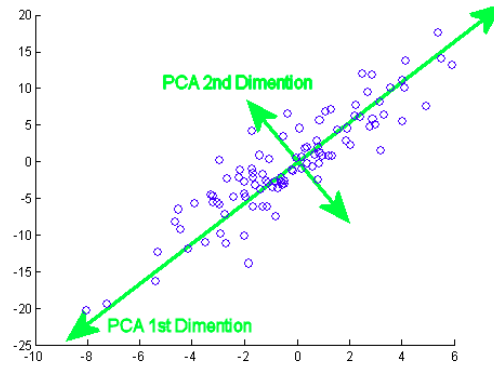


Figure 57- Example of PCA components generated using 2D dataset (i.e., 2 features)
(ref: <https://towardsdatascience.com>)

When PCA is calculated, the relationships between the original features and the principle components can be visualized using a biplot (e.g., see Fig. 59). Here, as the principle components are a linear combination of the original features, the original features are represented as vectors in the PCA coordinate system.

PCA calculation steps can be explained as following.

Let x be the feature matrix which contains m features and n observations,

$$x = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix} \quad (47)$$

Normalize each column of x (i.e., normalize each feature to get the variability centered around the mean) and get X .

$$X = \begin{pmatrix} \frac{x_{11}-\bar{x}_1}{\sigma_{x_1}} & \dots & \frac{x_{1m}-\bar{x}_m}{\sigma_{x_m}} \\ \vdots & \ddots & \vdots \\ \frac{x_{n1}-\bar{x}_1}{\sigma_{x_1}} & \dots & \frac{x_{nm}-\bar{x}_m}{\sigma_{x_m}} \end{pmatrix} \quad (48)$$

where,

$$\bar{x}_l = \frac{1}{n} \sum_{j=1}^n x_{ji} \quad (49)$$

$$\sigma_{x_i} = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ji} - \bar{x}_l)^2} \quad (50)$$

Compute the covariance matrix of X .

$$Cov = \frac{1}{n} X^T X \quad (51)$$

Note that Cov is a square matrix of size $m \times m$.

Calculate the eigenvalues of Cov . The eigenvalues are derived from the following equation where V is the eigen vector.

$$Cov.V = \lambda.V \quad (52)$$

Simplifying the above equation, the following can be derived.

$$\text{Det}(Cov - \lambda I) = 0 \quad (53)$$

By solving above equation, m number of eigenvalues can be found. Sort the eigen values in descending order.

$$\lambda_1, \lambda_2, \dots, \lambda_m$$

where λ_1 is the largest and λ_m is the smallest. The corresponding eigenvectors determines the direction of the principle components while eigenvalues determines its amplitude.

let V_1 ($1 \times m$) matrix be the corresponding eigenvector of λ_1 . Then the features can be transformed in to the first principle component by the product $V_1 X$.

Linear discriminant analysis

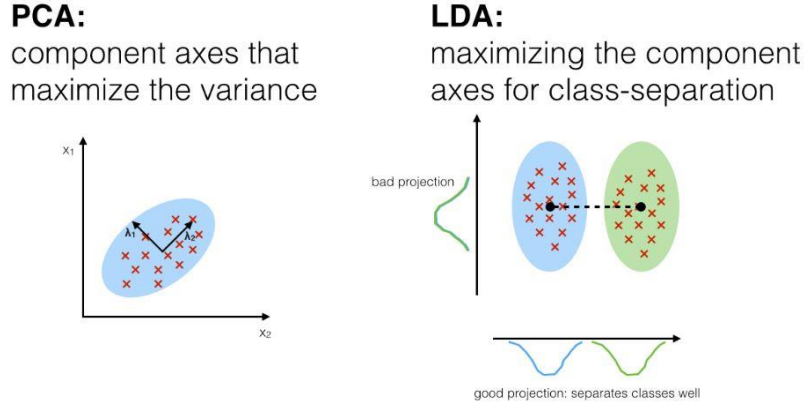


Figure 58- Simplified illustration of PCA and LDA component axes
(https://sebastianraschka.com/Articles/2014_python_lda.html)

Linear Discriminant Analysis (LDA) is a dimensionality reduction technique which is commonly used in ML applications. LDA projects the feature data onto a lower dimension such that the class separability is optimum (Fisher, 1936). In contrast to PCA which finds the axes in the direction of maximum variance, LDA finds the axes which maximize the separation between classes (e.g., see Fig. 58). LDA calculations steps can be explained as following.

Let X be the feature matrix which contains n observations (i.e., feature vectors),

$$X = \begin{pmatrix} x_{1,j} \\ \vdots \\ x_{i,j} \\ \vdots \\ x_{n,j} \end{pmatrix} \quad (54)$$

where $j \in C$ represents the corresponding class (i.e., $C = 1, 2$) of feature vector x_i .

Calculate the between-class scatter matrix S_b and within-class scatter matrix S_w .

$$S_b = \sum_{j=1}^g N_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T \quad (55)$$

$$S_w = \sum_{j=1}^g \sum_{i=1}^{N_j} (x_{i,j} - \bar{x}_j)(x_{i,j} - \bar{x}_j)^T \quad (56)$$

Where $x_{i,j}$ is an n dimensional feature point (i.e., vector) i from class j and N_j is the number of feature points from class j and g is the total number of classes.

Then the LDA components are found by solving the generalized eigenvalue problem for the matrix $S_w^{-1}S_B$ solving the following equation to find the eigen values λ and eigen vectors v .

$$(S_w^{-1}S_B)v = \lambda v \quad (57)$$

While the direction of the LDA components are represented by the eigenvectors, their amplitude is represented by the eigenvalues. Hence, the LDA components in the direction of best class separability can be found by sorting the eigen values with highest amplitudes. However, in LDA, it can be observed that at most only $g - 1$ number of eigenvalues will be non-zero. This is due to the fact that only g number of points are used when calculating S_b (Fisher, 1936).

After finding the $g - 1$ eigenvectors with nonzero eigenvalues, the feature vector can be projected into LDA space as following.

$$X_{LDA} = X \times W \quad (58)$$

where X_{LDA} is the transformed feature vector in LDA space and W is the $(g - 1) \times n$ dimensional transform vector which consists of $g - 1$ eigenvectors correspond to non-zero eigenvalues.

Classification Algorithms

Although there are many different ML algorithms available, the current study only consider two well-known classifiers, namely, linear Support Vector Machine (SVM) and K- nearest neighbor (K-NN). These ML models are considered due to their relative simplicity (i.e., compared to neural nets in deep learning applications) and the

limited amount of data available for training the models. Also, the study is not focused on optimizing the model accuracy through optimizing the hyperparameters of ML models, since, such optimization may provide a too generalized solution to the available small dataset. In addition, such classifiers (i.e., linear SVM and K-NN) can be used to provide a simple interpretation (or visualization) of the separation of different classes (i.e., decision boundary) in feature space. Similar ML models were used in previous literature (ref) for classifying HF using physiological features, owing to aforementioned reasons.

Here, SVM was used with the sorted features based on the statistical significance test while K-NN algorithm was used with the LDA derived feature calculated from the sorted sorted features.

Support vector machine

Support Vector Machine (SVM) is considered as a very efficient ML method that was successfully utilized in many healthcare related ML applications (Chow, Zhong, Blackmon, Stolz, & Dowell, 2008; Venkatesan, Karthigaikumar, Paul, Satheskumaran, & Kumar, 2018). SVM focuses on defining a hyperplane (or decision boundary) such that the margin between two classes are maximized (Cortes & Vapnik, 1995).

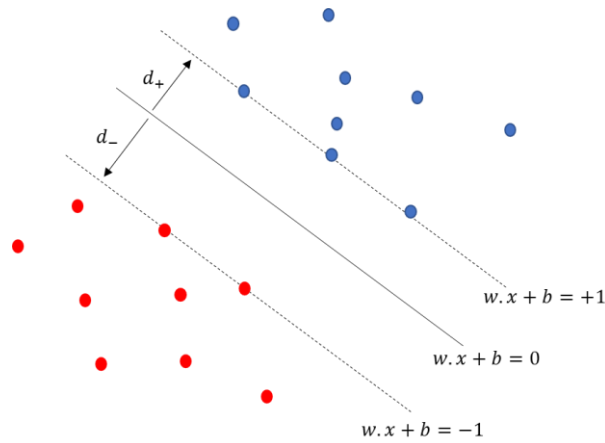


Figure 59- Simplified illustration of SVM hyperplanes

Provided that the data is linearly separable, a hyperplane margin which is bounded by the planes, $w \cdot x_i + b = y_i$ where $y_i = +1$ or -1 can be defined. Here, $y_i = +1$ and $y_i = -1$ defines the class of the data point. The data points on the margin boundary planes are defined as the support vectors. The separation hyperplane $w \cdot x + b = 0$ is defined as the median between the boundary hyperplanes where w, x and b are the weight vector, feature vector and the bias, respectively. Then, the margin of separation between the separation hyperplane and each classes can be defined as d_+ and d_- . It can be shown that d_+ or $d_- = \frac{1}{\|w\|}$. Hence, the distance between two boundary hyperplanes (i.e., margin) becomes $\frac{2}{\|w\|}$.

Since, the goal of SVM is to maximize the margin of separation between the classes, the following formulation is formed.

$$\text{maximize } \frac{2}{\|w\|} \rightarrow \text{minimize } \|w\| \text{ subjected to the condition } y_i(w \cdot x_i) - b = 1.$$

Above is a constrained optimization problem which is solved using the Lagrangian multiplier method (Cortes & Vapnik, 1995). The determined values of optimum w and b is used to define the separation hyperplane.

K- nearest neighbor classification

K- Nearest Neighbor (K-NN) classifier is a supervised classification algorithm which predicts (or classify) a new observation point by finding a predefined number of K training samples which are closest in distance to the new observation point. The new point will be classified into the group which majority of K samples are included. While the distance can be measured by any metric, Euclidean distance is the most common choice. In K-NN classifier, the parameter K can be a predefined constant or a parameter which varies based on the local density of points. The latter is also known as radius-based K-NN classifier where a point is classified into a group based on the groups of the training points in a region defined by a radius around the testing point.

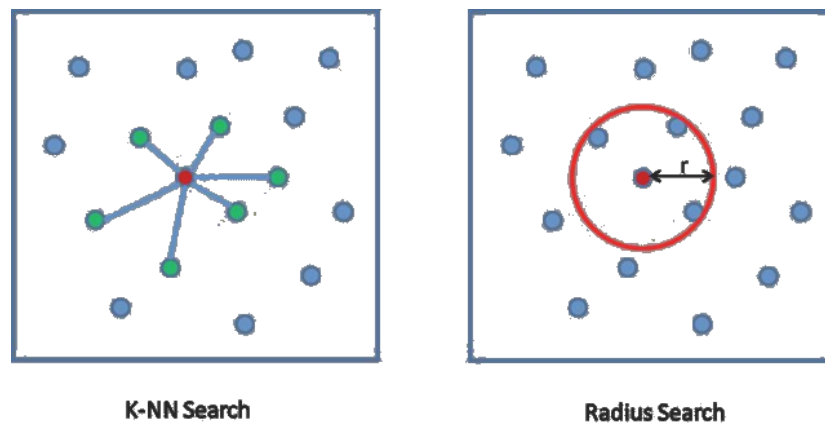


Figure 60- K-NN classifier illustration (Sun, Shaohui. "Automatic 3D building detection and modeling from airborne LiDAR Point clouds." (2013).)

Results and Discussion

As discussed in feature analysis and selection section , features were sorted by analyzing the statistical significance of the difference of the features between following two groups.

1. **R**: Readmitted (data sessions within 60 days of re-admission of readmitted HF patients)

2. **nR**: not Readmitted (data sessions within 60 days from initial discharge of non-readmitted HF patients)

These groups were formed under the hypothesis that, a subject is likely to show bad cardiac mechanical activity within a 2 month period before the re-admission and a subject is likely to show relatively better cardiac mechanical activity within 2 months after initial discharge (also provided that the subject is not re-admitted within 6 months after readmission).

Feature sorting using statistical analysis

Statistical significance of the feature difference between R and nR groups were evaluated using unpaired t-test. A feature is sorted as a good feature if the P value is less than 0.05. When calculating the P values, a third group H, which consisted of features derived from recordings on 10 healthy subjects without any reported heart disease was considered. Here, H was used to validate the observed differences between R and nR distributions under the hypothesis that if a feature shows a significant difference (P value <0.05) between R and nR, for that feature to be valid, feature distribution of nR should be closer to H than R (i.e., P value between R-H $<$ P value between nR-H). Hence, the conditions for sorting a feature are summarized as following.

- Small p value ($p < 0.05$) between R and nR
- Small p value ($p < 0.05$) between R and H
- p value between R and H $<$ p value between nR and H (i.e., nR distribution is closer to H than R)

Based on aforementioned criteria 14 features were sorted. The p values of these features between three groups are shown in Table 5.

Table 5- Sorted features based on unpaired t-test ($p < 0.05$)

Feature	P values			
	R-nR	R-H	nR-H	(R&nR)-H
Frms_{ave}	3.61E-13	5.55E-11	0.480887	0.070465
Variability before clustering	2.04E-06	1.07E-04	0.13784	0.008234
Inter cluster variability	7.86E-06	0.000595	0.471128	0.031405
Intra cluster variability	2.55E-05	0.00054	0.206867	0.015878
SCG2 amp aud_{ave}	0.00094	0.001365	0.167365	0.019987
Frms_{dif}	0.000953	1.88E-02	0.687118	0.121227
SCG2 amp_{ave}	0.001481	8.08E-05	0.126767	0.003049
SCG1 amp_{dif}	0.006894	0.112748	0.775736	0.333798
SCG2 amp_{dif}	0.00815	0.000311	0.002401	6.08E-06
LF/HF	0.008849	0.205675	0.112346	5.55E-01
F_avg(10,20)_{ave}	0.010607	0.014029	0.610074	0.127506
SCG2 amp aud_{dif}	0.011014	0.179785	0.380613	0.482817
PCT_ER_SCG1-SCG2_{ave}	0.030356	0.001671	0.097688	0.013009
PCT_ER_SCG1_{ave}	0.048541	0.000878	0.025976	0.003967

PCA analysis was performed on the sorted features. As PCA components represents the axes in the directions of highest variability of these features, PCA can be used to visualize the variability of the features in low dimensional space (i.e 2D or 3D) which will provide insights on how the variability in the feature data separates different classes (i.e., Re, nRe, H) . Fig. 61 and 62 show biplots which represent the sorted features as vectors in the space of dominant PCA components (i.e., PCA components are the axes of the plot). Here, the direction and the length of each feature vector represents how each vector contributes to the principle component.

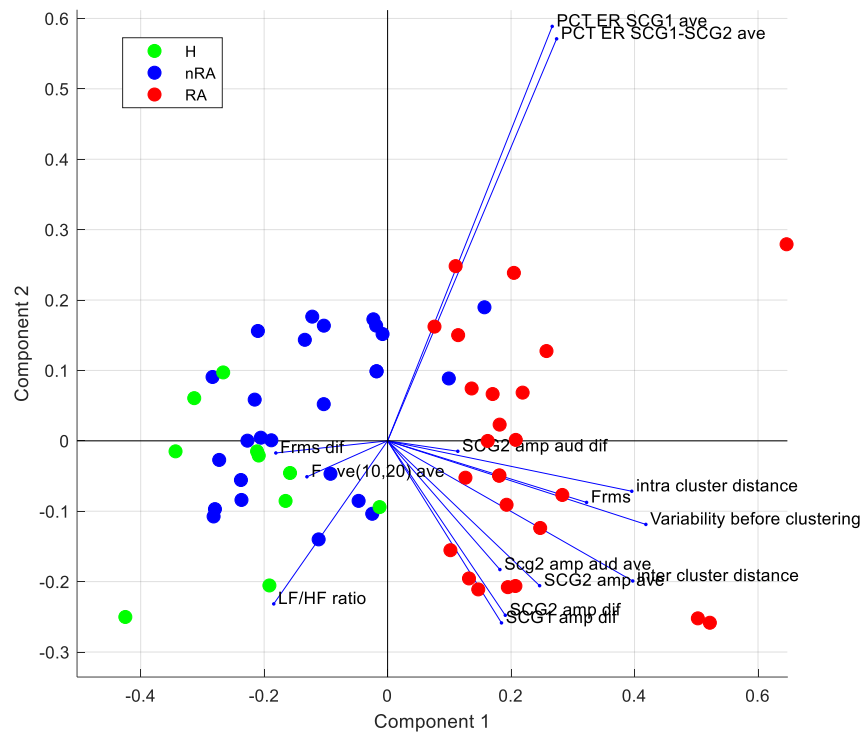


Figure 61- 2D biplot with first 2 PCA components representing PCA analysis of sorted features and class distribution

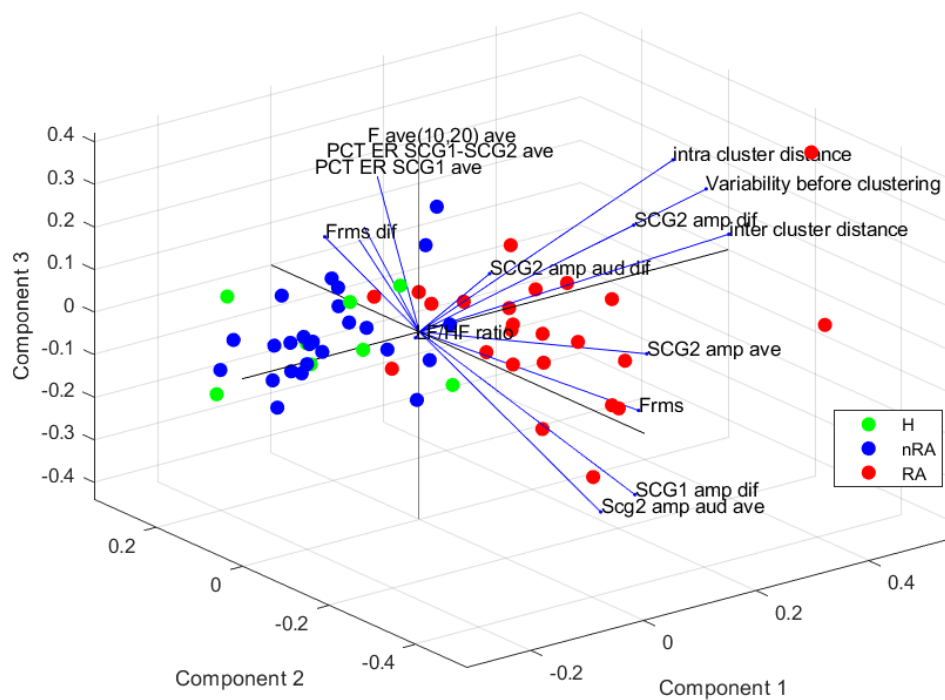


Figure 62- 3D biplot with first 3 PCA components representing PCA analysis of sorted features and class distribution

The results presented in Fig. 58 and 59 provide evidence that the inherent variability of the sorted features can help separate the classes R and nR. Also, in these biplots, it can be seen that the observations of group H and nR are clustered in the same region indicating that the variability of the features of healthy group 'H' is closer to that of nR group. This verify the hypothesis that features of a HF patient are likely to reach (or be closer) to those of a healthy subject if the HF patient is not readmitted (within 6 months after intial discharge).

Feature derivation using LDA

In the previous section, the best subset of features were selected based on a statistical measure (using unpaired t-test). However, having mulptiple features may complicate the interpretation of the feature differences in different classes (i.e., maximum 3 features can be visulaized together in 3D) and increase the calculation cost when used with ML classification. In such cases, the derivation of new simplified features using a feature transformation mechanism which minimizes the within-class scatter while maximizing the between class scatter is helpful. LDA is a proven feature reduction method for classification purposes (Balakrishnama & Ganapathiraju, 1998) and some studies (Mandelkow, de Zwart, & Duyn, 2016; Shahbazi & Asl, 2015) have shown superior accuracy in ML cassification when used with LDA dirived features. The current study utilized linear discreminant analysis (LDA) to transform the sorted features in to a new feature (i.e., number of non zero LDA components equal to $c-1$ where c is the number of classes) . Fig. 63 shows the box plots of the sorted features and the LDA derived feature.

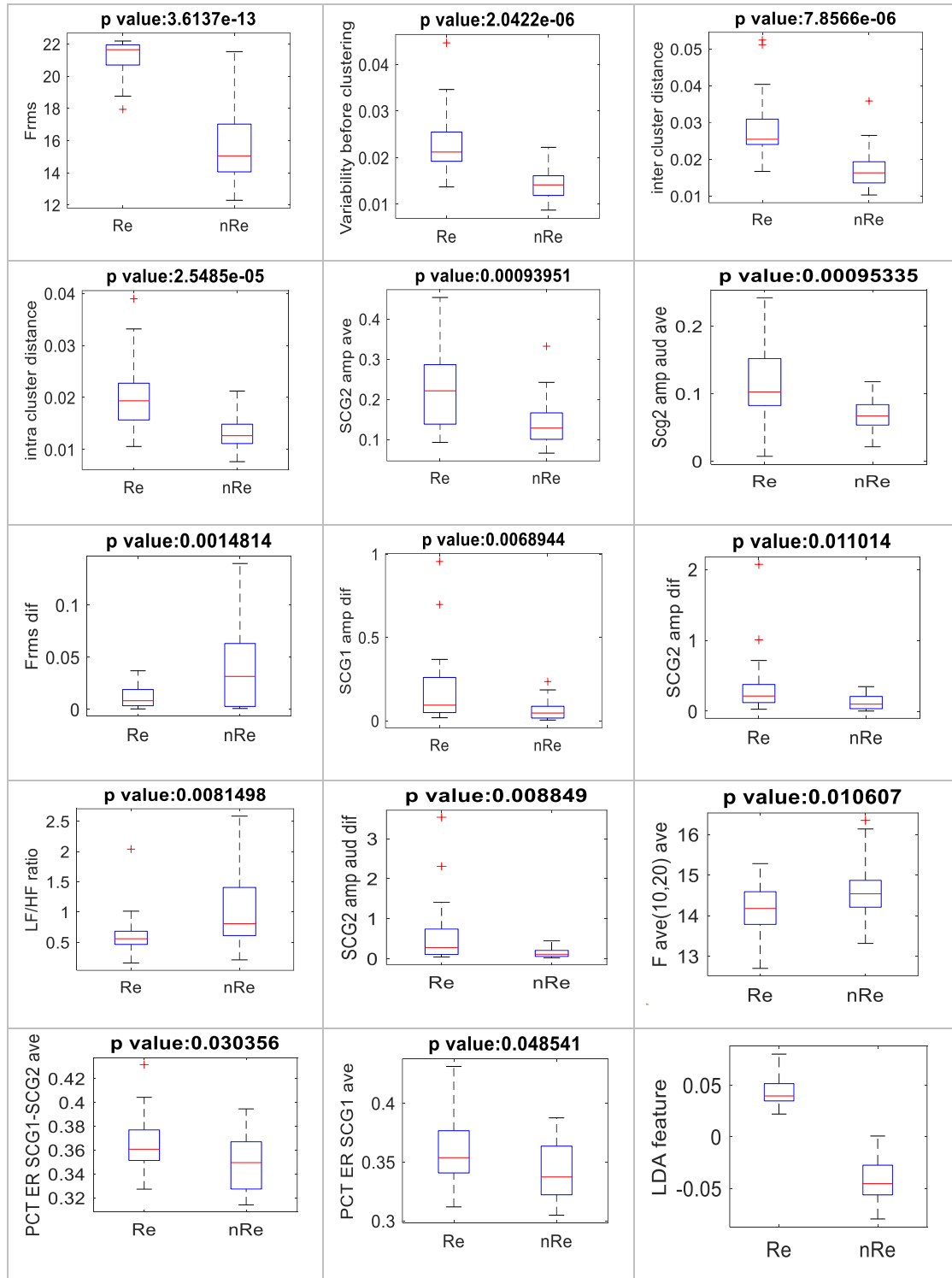


Figure 63- Box plots of sorted features and LDA derived feature

In Fig. 60, certain overlaps between the classes can be observed in the sorted features. The LDA derived feature shows total separability between the classes. This

separation between the two classes can be also seen in Fig. 64 which shows the feature space plot of the new feature created using LDA.

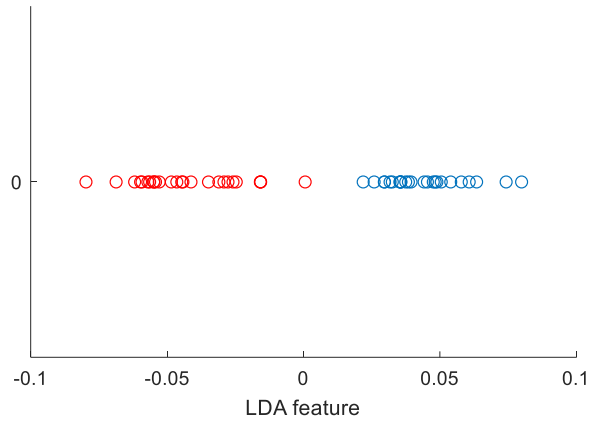


Figure 64- LDA derived feature (red-R , blue nR)

Classification using machine learning (ML)

Two ML models were used to classify the HF groups; readmitted (R) and non-readmitted (nR).

1. Linear SVM model with 14 sorted features
2. K-NN model with LDA derived feature.

ML models were trained using features derived from 25 and 26 number of data sessions in R and nR groups, respectively. The cross validation accuracies using leave one out and 10-fold cross validation is shown in Table 6.

Table 6- Validation accuracies of ML classification

ML model	Leave one out accuracy%	10-fold cross fold accuracy%
SVM	92.16	94.12
K-NN (k=1)	100	100

The results showed that K-NN classifier with derived LDA feature delivered 100% validation accuracy. This result suggests that LDA derived feature has the potential to deliver a high accuracy than all 14 sorted features used for classification. A similar result (i.e., 100% accuracy) were observed in a previous study where generalized discriminant analysis (GDA) (i.e., GDA is an extension of LDA where LDA is performed after transforming features in to a new space using kernel functions) was used to transform multiple sorted features in to a new feature to classify congestive heart failure (CHF) patient risk levels (Shahbazi & Asl, 2015).

It should be noted that current results are delivered ~25 observations in each class and the ML models may provide generalized solutions to the small number of data. However, under the assumption that the new data will come from similar distributions, these results show the potential capability of both ML algorithms to classify HF data in to R and nR with high accuracy.

SVM accuracy results were also analyzed by varying the number of input features to the ML classifier. Here the number of features were sorted using MrMR algorithm. Fig. 65 and Fig. 66 show the variation of the training accuracy with leave one out and 10 fold cross validation accuracies, respectively. The results show the the training accuracy reaches up to ~96 % with the increase of number of features (beyond 5) while the validation accuracies remains on a lower value. This phenomena is expected due to the generalization of the ML optimum solution to the training data. It can be observed that the validation accuracies reaches its maximum with (4-6) number of features. The fluctuations in the 10-fold cross validation accuracy is probably caused by the differences in the randomly generated folds in different runs, and with large data populations a smoother curve is expected.

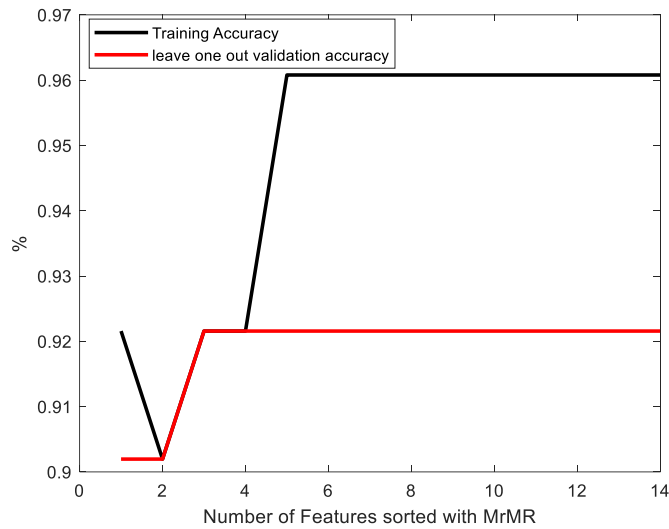


Figure 65- Variation of the training accuracy and the leave one out validation accuracy vs. number of features used (sorted using MrMR algorithm)

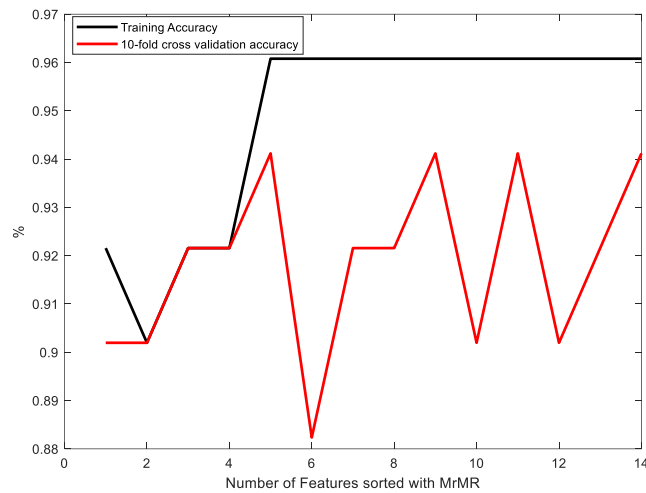


Figure 66- Variation of the training accuracy and the 10-fold cross validation accuracy vs. number of features used (sorted using MrMR algorithm)

Figs 64 and 65 represent the visualization of SVM hyperplane (or decision boundary) when only 3 and 2 sorted features were used, respectively. When 3 features were used SVM delivered a 92.16% leave one out validation accuracy and when 2 features were used 90.20% leave one out accuracy was observed. In both cases the training accuracy was 92.16%.

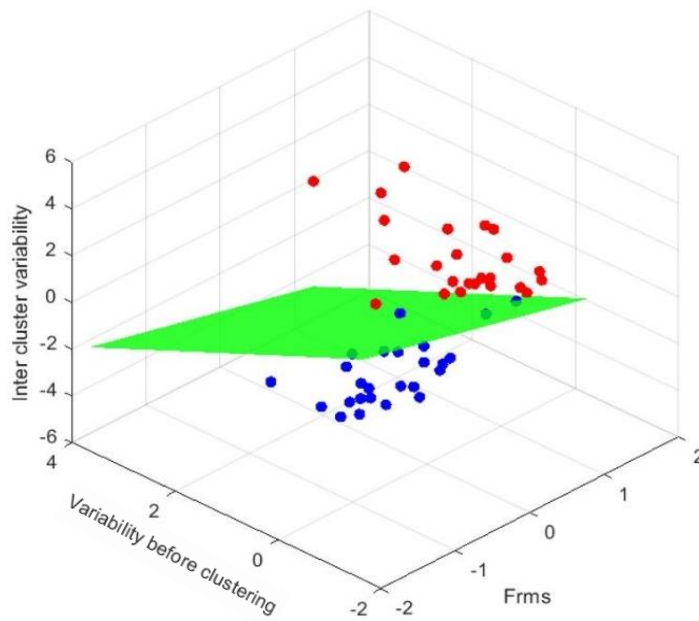


Figure 67- Representation of SVM hyperplane with first 3 sorted features. Observation points of classes R and nR are represented by red and blue circles respectively.

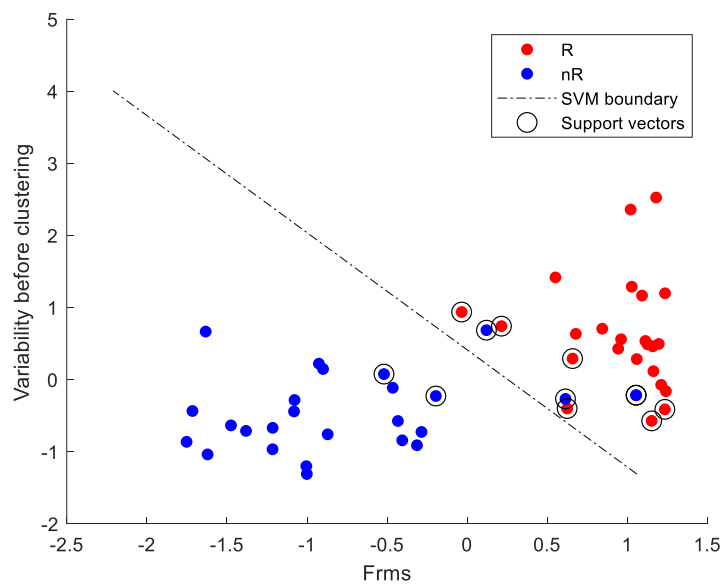


Figure 68- Representation of SVM hyperplane with first 2 sorted features. Observation points of classes R and nR are represented by red and blue circles respectively.

HF patients longitudinal trajectory prediction using ML

Trained ML models were used to predict the longitudinal class trajectory (i.e., R or nR) of each patient. Here, feature vectors generated from all data sessions from each patient were left out (i.e., one patient at a time) and used as testing data. The ML model predicted the class (i.e., R or nR) of each data session.

Class trajectory of each patient using SVM and K-NN classifiers are shown in Fig. 69 and 70, respectively.

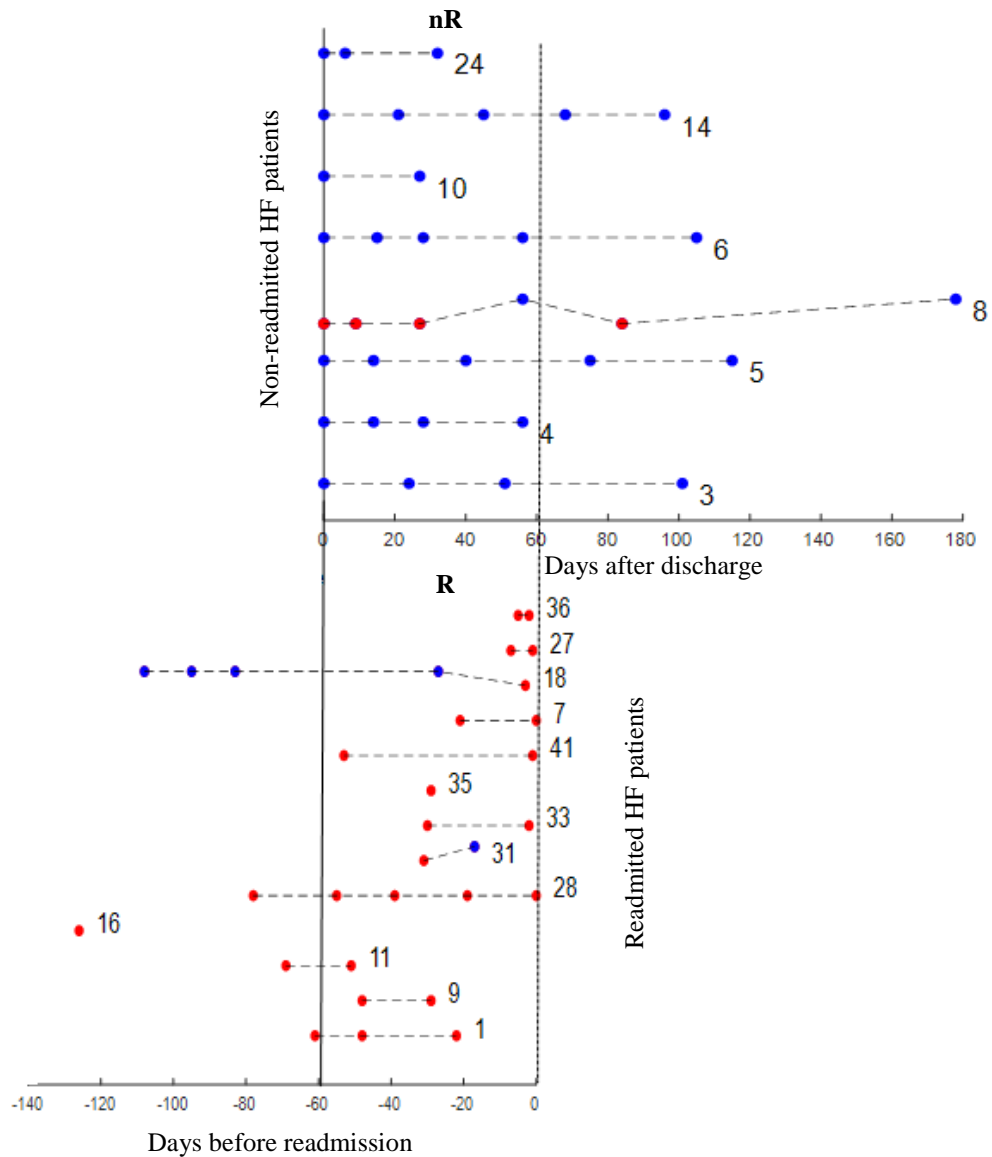


Figure 69- Class trajectory of each patient predicted using SVM. (blue-nR, red-R)

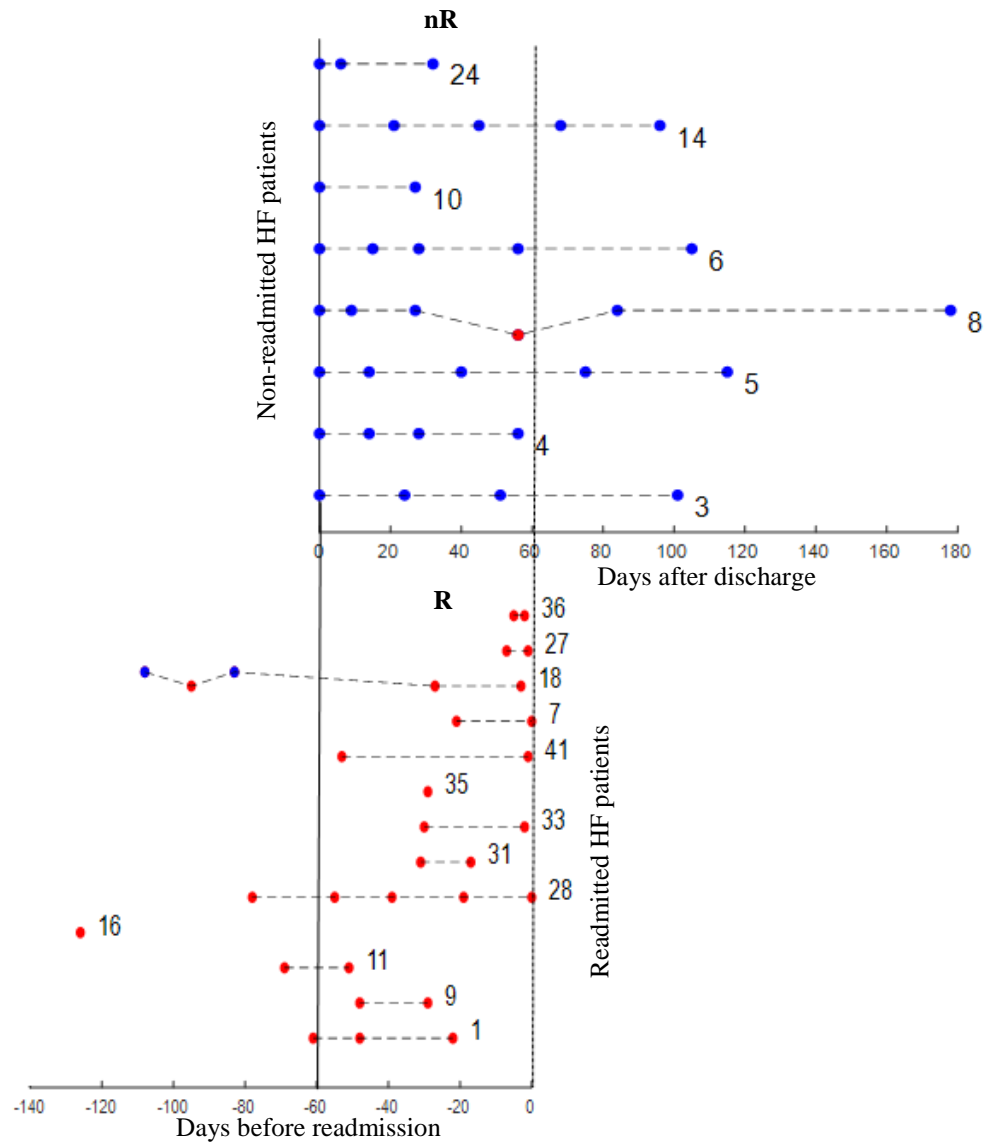


Figure 70- Class trajectory of each patient predicted using K-NN. (blue-nR, red-R)

In these Fig. 69 and 70, class nR is indicated in blue dots while class R is indicated in red dots. The numbers represent the numbers assigned to HF patients in the study. Here, it can be noted that for readmitted HF patients, days are counted relative to the day of readmission while for non readmitted HF patients, days are counted relative to the day of initial discharge. Two graphs (i.e., graphs containing the trajectories of readmitted HF patients and non readmitted HF patients) are shifted to align the groups R and nR. Here, the vertical lines in Fig.s 69 and 70 are used to visually

partition the data sessions within R and nR groups. As mentioned earlier, when predicting a trajectory of a patient, all the data sessions (including the data sessions in R and nR range) were excluded when training the ML algorithm. Hence, the predicted trajectory can be named as ‘leave one patient out trajectory prediction’.

In Fig. 69, all the HF patients except patient numbers 8,18 and 31 show a constant trajectory indicating that the patients likely to be readmitted (i.e., similar to class R) for readmitted HF patients and to be not readmitted (i.e., similar to class nR) for non readmitted HF patients. The first 3 data sessions of HF patient number 8 suggest that the patient is likely to be readmitted while the 4th data session suggests the HF condition got better. However, his/her HF condition again fluctuates after 60 days from discharge. HF patient number 18 shows that initially the patient is in a better condition (i.e., first 4 data sessions are classified as nR) before he/she reaches readmission. HF patient number 31 shows that the HF condition got better 15 days before the readmission. However, his/her HF condition near readmission couldn't be evaluated due to the unavailability of data, proximate or at the day of readmission.

As shown in Fig. 70, in contrast to SVM model, K-NN model correctly classified all the data sessions except the 4th data session of HF patient number 8. In addition, the trajectory of HF patient number 18 shows that the HF condition was fluctuating between R and nR before 90 days from readmission.

These probable misclassifications may be due to a complex heart conditions which cannot be accurately captured by SCG features. Also, it should be noted that HF patient readmission is a specific decision taken by a clinician, evaluating the clinical data of the patient, while the current ML model prediction is purely based on the physiological features extracted from SCG.

In the ideal application of the developed ML models, ML algorithms are trained using the available HF patient data and the future HF patient data will be used to make the predictions. To mimic a similar scenario, a set of most recently acquired HF patient data (i.e., Patient set 33,35,36,41 and patient set 10,14,24 for readmitted and non-readmitted groups, respectively) were excluded from training the machine learning model and were used to make the predictions. These results are shown in Fig. 71. As shown in the results, ML models (i.e., SVM and K-NN) were able to deliver 100% accurate predictions.

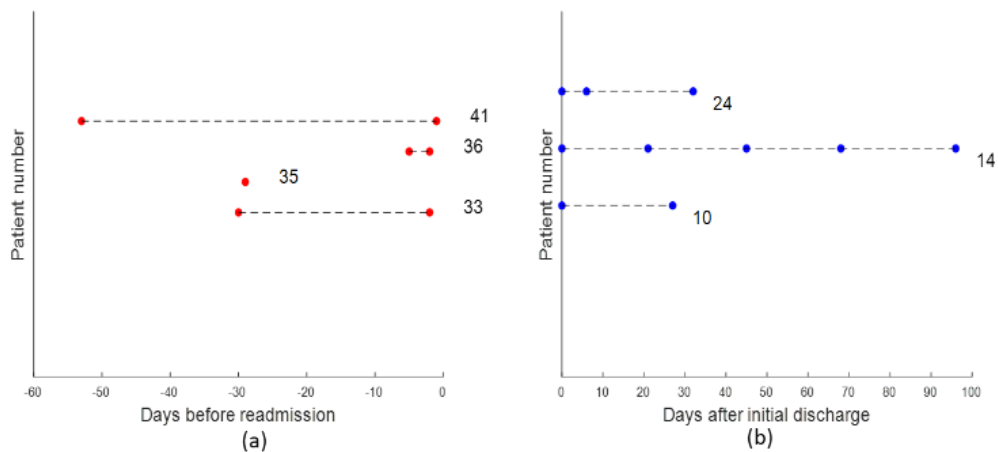


Figure 71- Class trajectory predictions of (a) readmitted and (b) non-readmitted HF patients when patients are divided into training and testing sets. (i.e., Patient set 33,35,36,41 and patient set 10,14,24 are used as the testing set for readmitted and non-readmitted groups,respectively)

The ML models were also tested with a new set of SCG data acquired from HF patient who have not completed the 6 months of monitoring. According to the predictions all the data sessions in these patients were predicted as nR class (using both SVM and K-NN models). These results are shown in Fig. 72 wghich indicate that all these HF patients are not likely to be readmitted within 6 months after discharge.

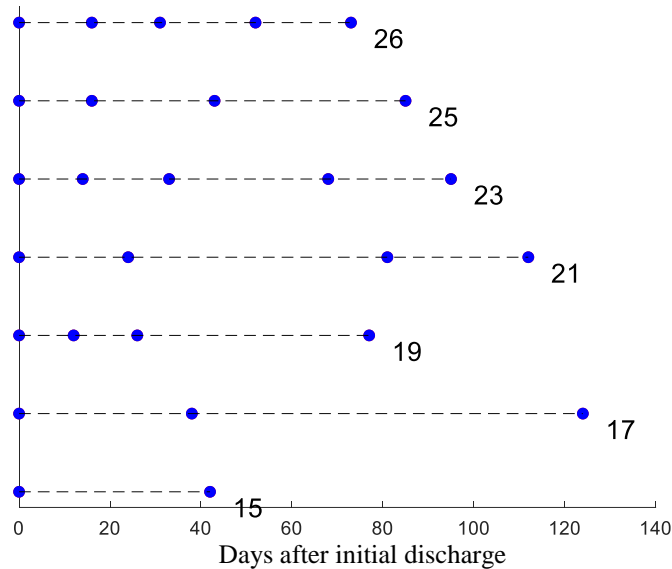


Figure 72- ML predictions for the HF patients who have not completed the 6 months of study

Use of clinical data for predicting HF readmission

The ability of collected clinical data information to predict the HF readmission was also evaluated. Here, two criteria were analyzed.

1. Meta-Analysis Global Group in Chronic (MAGGIC)-HF Risk Score
2. Thoracic impedance

Meta-Analysis Global Group in Chronic (MAGGIC)-HF Risk Score

The MAGGIC Risk score is predicted using 13 variables: age, sex, body mass index, systolic blood pressure, EF, creatinine, smoking status, diabetes mellitus, chronic obstructive pulmonary disease, NYHA class, β -blocker use, and angiotensin-converting enzyme inhibitor use. The study (Pocock et al., 2013) used MAGGIC score for predicting survival in heart failure, and the score was calculated using 13 variables with a weighted integer, having totals ranging from 0-50 points (where more points = higher risk patient).

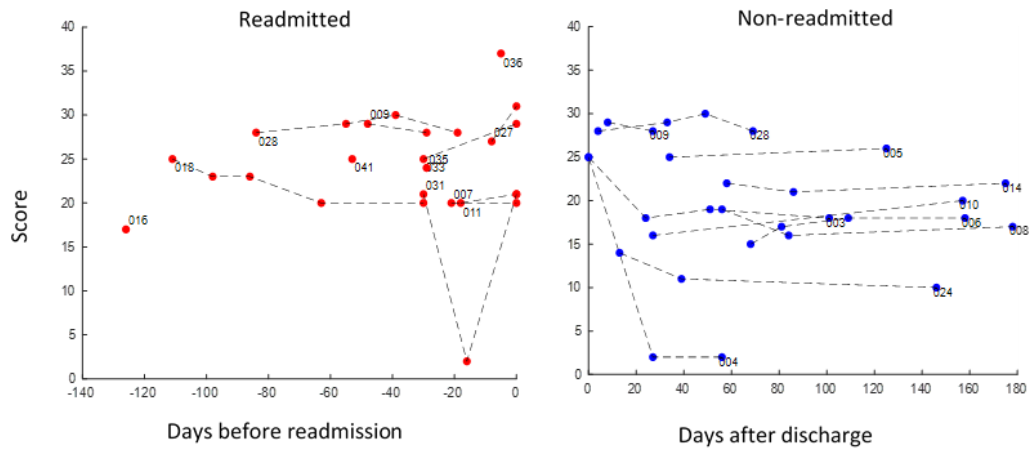


Figure 73- plots of longitudinal variation of MAGGIC score in readmitted and non-readmitted HF patients.

Fig. 73 shows the longitudinal variation of MAGGIC score in both readmitted and non-readmitted HF patients. As seen in the results, MAGGIC scores in both groups vary in similar ranges (2-35) without any significance differences between readmitted and non-readmitted groups.

Thoracic impedance

Thoracic impedance (TI) is simply a measurement of the resistance of a electric current as it travels from the top to the bottom of the thorax. Hence, TI is an indirect measure of the water content in the lungs (i.e., high water content in lungs will result in low TI and vise versa). Excess fluid buildup in the thorax (i.e., low TI) is considered as a symptom of heart failure. In the current study, TI was measured using ZOE® (FDA approved) system and the measured TI (Z_0) values of the HF patients are shown in Fig. 74.

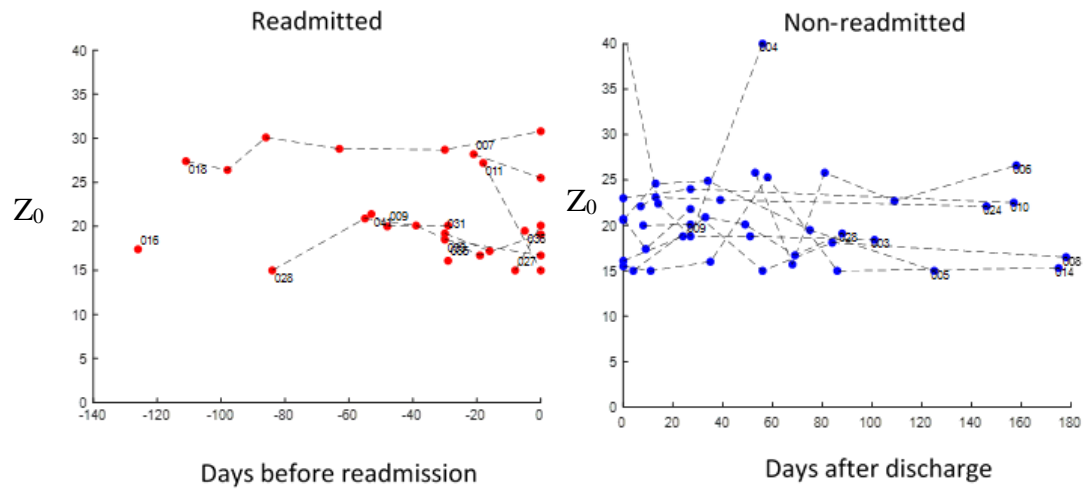


Figure 74- Thoracic impedance (Z_0) values of HF patients

As shown in Fig. 74, no clear differences between the TI were found between the readmitted and non-readmitted HF patients suggesting that TI alone is not able to serve as an accurate measure to predict HF patient readmission.

Conclusion

In general, the presented ML models demonstrated the capability of predicting a HF patient is likely to be readmitted or not (within 60 days) with a high accuracy (> 92%) based on the features extracted from SCG data. The clinical data were also assessed to identify the differences between the readmitted and non readmitted groups, and the results showed clinical data alone could not make accurate predictions. The use of ML classifiers with SCG derived features combined with clinical information will allow doctors to pre-identify the HF patients who need extra care after the initial discharge, which will help save money and reduce mortality.

LIST OF REFERENCES

- Aggarwal, S., & Gupta, V. (2014). Factors associated with 30-day readmissions in congestive heart failure patients: analysis of 2,536,439 admissions. *Journal of the American College of Cardiology*, 63(12 Supplement), A731. doi:10.1016/s0735-1097(14)60731-2
- Akhbardeh, A., Tavakolian, K., Gurev, V., Lee, T., New, W., Kaminska, B., & Trayanova, N. (2009). *Comparative analysis of three different modalities for characterization of the seismocardiogram*. Paper presented at the Conference Proceedings.
- Al-Mayah, A., Moseley, J., & Brock, K. (2007). Contact surface and material nonlinearity modeling of human lungs. *Physics in Medicine & Biology*, 53(1), 305.
- Angelone, A., & Coulter JR, N. A. (1964). Respiratory sinus arrhythmia: a frequency dependent phenomenon. *Journal of applied physiology*, 19(3), 479-482.
- Annema, C., Luttik, M.-L., & Jaarsma, T. (2009). Reasons for readmission in heart failure: perspectives of patients, caregivers, cardiologists, and heart failure nurses. *Heart & Lung*, 38(5), 427-434.
- Azad, M. K., Gamage, P. T., Sandler, R. H., Raval, N., & Mansy, H. A. (2019). *Seismocardiographic Signal Variability During Regular Breathing and Breath Hold in Healthy Adults*. Paper presented at the 2019 IEEE Signal Processing in Medicine and Biology Symposium (SPMB).
- Balakrishnama, S., & Ganapathiraju, A. (1998). *Linear discriminant analysis-a brief tutorial*. Paper presented at the Institute for Signal and information Processing.
- Baron, J. (2004). *Experience with 3D optical flow on gated MRI cardiac datasets*. Paper presented at the First Canadian Conference on Computer and Robot Vision, 2004. Proceedings.
- Boashash, B. (1992). Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals. *Proceedings of the IEEE*, 80(4), 520-538.
- Bogren, H., Lantz, B., Miller, R., & Mason, D. (1977). Effect of Respiration on Cardiac Motion Determined by Cineangiography: Implications concerning Three-Dimensional Heart Reconstruction Using Computer Tomography. *Acta Radiologica. Diagnosis*, 18(6), 609-620.
- Bombardini, T., Gemignani, V., Bianchini, E., Venneri, L., Petersen, C., Pasanisi, E., . . . Giannoni, M. (2007). Cardiac reflections and natural vibrations: Force-frequency relation recording system in the stress echo lab. *Cardiovascular ultrasound*, 5(1), 42.
- Bozhenko, B. (1961). Seismocardiography--a new method in the study of functional conditions of the heart. *Terapevticheskii arkhiv*, 33, 55.
- Chawla, A., Mukherjee, S., & Karthikeyan, B. (2006). Mechanical properties of soft tissues in the human chest, abdomen and upper extremities. *Institution of Engineers, J. Mechanical Engineering*.
- Chow, R., Zhong, W., Blackmon, M., Stolz, R., & Dowell, M. (2008). *An efficient SVM-GA feature selection model for large healthcare databases*. Paper presented at the Proceedings of the 10th annual conference on Genetic and evolutionary computation.

- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- Crow, R. S., Hannan, P., Jacobs, D., Hedquist, L., & Salerno, D. M. (1994). Relationship between seismocardiogram and echocardiogram for events in the cardiac cycle. *American journal of noninvasive cardiology*, 8, 39-46.
- Dash, S., Chon, K., Lu, S., & Raeder, E. (2009). Automatic real time detection of atrial fibrillation. *Ann Biomed Eng*, 37(9), 1701-1709.
- Di Rienzo, M., Meriggi, P., Rizzo, F., Vaini, E., Faini, A., Merati, G., . . . Castiglioni, P. (2011). *A wearable system for the seismocardiogram assessment in daily life conditions*. Paper presented at the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
- Drazner, M. H., Rame, J. E., Stevenson, L. W., & Dries, D. L. (2001). Prognostic importance of elevated jugular venous pressure and a third heart sound in patients with heart failure. *New England Journal of Medicine*, 345(8), 574-581.
- Felner, J. M. (1990a). The First Heart Sound. In H. W. D. H. J. W. Walker Hk (Ed.), *Clinical Methods: The History, Physical, and Laboratory Examinations*. (3 ed.). Boston: Butterworths.
- Felner, J. M. (1990b). The Second Heart Sound. In H. W. D. H. J. W. Walker Hk (Ed.), *Clinical Methods: The History, Physical, and Laboratory Examinations*. Boston: Butterworths.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2), 179-188.
- Foulger, J. H., Smith Jr, P. E., & Fleming, A. J. (1947). Changes in cardiac vibrational intensity in response to physiologic stress. *American heart journal*, 34(4), 507-539.
- Foulger, J. H., Smith Jr, P. E., & Fleming, A. J. (1948). Cardiac vibrational intensity and cardiac output. *American heart journal*, 35(6), 953-961.
- Gamage, P. T., Azad, M. K., Taebi, A., Sandler, R. H., & Mansy, H. A. (2018). *Clustering Seismocardiographic Events using Unsupervised Machine Learning*. Paper presented at the 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB).
- Gamage, P. T., Azad, M. K., Taebi, A., Sandler, R. H., & Mansy, H. A. (2020). Clustering of SCG Events Using Unsupervised Machine Learning. In I. Obeid, I. Selesnick, & J. Picone (Eds.), *Signal Processing in Medicine and Biology: Emerging Trends in Research and Applications* (pp. 205-233). Cham: Springer International Publishing.
- Gefen, A., & Dilmoney, B. (2007). Mechanics of the normal woman's breast. *Technology and Health Care*, 15(4), 259-271.
- Giorgis, L., Hernandez, A., Amblard, A., Senhadji, L., Cazeau, S., Jauvert, G., & Donal, E. (2008). *Analysis of cardiac micro-acceleration signals for the estimation of systolic and diastolic time intervals in cardiac resynchronization therapy*. Paper presented at the Computers in Cardiology, 2008.
- Gurev, V., Tavakolian, K., Constantino, J., Kaminska, B., Blaber, A. P., & Trayanova, N. A. (2012). Mechanisms underlying isovolumic contraction and ejection peaks in seismocardiogram morphology. *Journal of medical and biological engineering*, 32(2), 103.

- Hamzah, M., Subit, D., Boruah, S., Forman, J., Crandall, J., Ito, D., . . . Yasuki, T. (2013). *An inverse finite element approach for estimating the fiber orientations in intercostal muscles*. Paper presented at the IRCOBI conference.
- Heidenreich, P. A., Albert, N. M., Allen, L. A., Bluemke, D. A., Butler, J., Fonarow, G. C., . . . Maddox, T. M. (2013). Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. *Circulation: Heart Failure*, 6(3), 606-619.
- Inan, O. T., Migeotte, P.-F., Park, K.-S., Etemadi, M., Tavakolian, K., Casanella, R., . . . Prisk, G. K. (2015). Ballistocardiography and seismocardiography: a review of recent advances. *IEEE J. Biomedical and Health Informatics*, 19(4), 1414-1427.
- Ingels Jr, N. B., Daughters 2nd, G., Stinson, E., & Alderman, E. (1975). Measurement of midwall myocardial dynamics in intact man by radiography of surgically implanted markers. *Circulation*, 52(5), 859-867.
- Isler, Y. (2016). Discrimination of systolic and diastolic dysfunctions using multi-layer perceptron in heart rate variability analysis. *Comput Biol Med*, 76, 113-119.
- Johnston, F. D., & OVERY, D. C. (1951). Vibrations of low frequency over the precordium. *Circulation*, 3(4), 579-588.
- Khosrow-Khavar, F., Tavakolian, K., Blaber, A. P., Zanetti, J. M., Fazel-Rezai, R., & Menon, C. (2015). Automatic annotation of seismocardiogram with high-frequency precordial accelerations. *IEEE journal of biomedical and health informatics*, 19(4), 1428-1434.
- Kido, T., Kido, T., Nakamura, M., Watanabe, K., Schmidt, M., Forman, C., & Mochizuki, T. (2016). Compressed sensing real-time cine cardiovascular magnetic resonance: accurate assessment of left ventricular function in a single-breath-hold. *Journal of Cardiovascular Magnetic Resonance*, 18(1), 50.
- Kleiger, R. E., Miller, J. P., Bigger Jr, J. T., & Moss, A. J. (1987). Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *The American journal of cardiology*, 59(4), 256-262.
- KOIWA, Y., OHYAMA, T., TAKAGI, T., KIKUCHI, J.-I., HONDA, H., SHIMIZU, Y., . . . TAKISHIMA, T. (1991). Vibration analysis of the human left ventricular posterobasal wall at the moment of the first heart sound emission. *The Tohoku journal of experimental medicine*, 164(2), 125-144.
- Korzeniowska-Kubacka, I., Kuśmierczyk-Droszcz, B., Bilińska, M., Dobraszkiewicz-Wasilewska, B., Mazurek, K., & Piotrowicz, R. (2006). Seismocardiography-a non-invasive method of assessing systolic and diastolic left ventricular function in ischaemic heart disease. *Cardiology Journal*, 13(4), 319-325.
- Krishnan, K., Mansy, H., Berson, A., Mentz, R. J., & Sandler, R. H. (2018). Siesmocardiographic Changes with HF Status Change: Observations from a Pilot Study. *Journal of Cardiac Failure*, 24(8), S54.
- Lahdenoja, O., Hurnanen, T., Iftikhar, Z., Nieminen, S., Knuutila, T., Saraste, A., . . . Pänkäälä, M. (2017). Atrial fibrillation detection via accelerometer and gyroscope of a smartphone. *IEEE journal of biomedical and health informatics*, 22(1), 108-118.
- Laurin, L.-P., Jobin, V., & Bellemare, F. (2012). Sternum length and rib cage dimensions compared with bodily proportions in adults with cystic fibrosis. *Canadian respiratory journal*, 19(3), 196-200.

- Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern Recognition*, 38(11), 1857-1874.
- Liu, G., Wang, L., Wang, Q., Zhou, G., Wang, Y., & Jiang, Q. (2014). A new approach to detect congestive heart failure using short-term heart rate variability measures. *PloS one*, 9(4), e93399.
- Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision.
- Mandelkow, H., de Zwart, J. A., & Duyn, J. H. (2016). Linear discriminant analysis achieves high classification accuracy for the BOLD fMRI response to naturalistic movie stimuli. *Frontiers in human neuroscience*, 10, 128.
- Morillo, D. S., Ojeda, J. L. R., Foix, L. F. C., & Jiménez, A. L. (2010). An accelerometer-based device for sleep apnea screening. *IEEE Transactions on Information Technology in Biomedicine*, 14(2), 491-499.
- Mounsey, P. (1957). Praecordial ballistocardiography. *British heart journal*, 19(2), 259.
- National Health and Nutrition Examination Survey. 2011.
- Nelson, S., Whitsel, L., Khavjou, O., Phelps, D., & Leib, A. (2016). Projections of cardiovascular disease prevalence and costs: RTI International. Retrieved from <http://www.heart.org/idc/groups....>
- O'Connor, C. M. (2017). High heart failure readmission rates: is it the health system's fault? : JACC: Heart Failure.
- Ozawa, Y., Smith, D., & Craige, E. (1983a). Origin of the third heart sound. I. Studies in dogs. *Circulation*, 67(2), 393-398.
- Ozawa, Y., Smith, D., & Craige, E. (1983b). Origin of the third heart sound. II. Studies in human subjects. *Circulation*, 67(2), 399-404.
- Paparrizos, J., & Gravano, L. (2015). *k-shape: Efficient and accurate clustering of time series*. Paper presented at the Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data.
- Paparrizos, J., & Gravano, L. (2017). Fast and accurate time-series clustering. *ACM Transactions on Database Systems (TODS)*, 42(2), 8.
- Patel, R., Bushnell, D., & Sobotka, P. (1993). Implications of an audible third heart sound in evaluating cardiac function. *Western journal of medicine*, 158(6), 606.
- Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8), 1226-1238.
- Peng, Z., Meng, G., Chu, F., Lang, Z., Zhang, W., & Yang, Y. (2011). Polynomial chirplet transform with application to instantaneous frequency estimation. *IEEE Transactions on Instrumentation and Measurement*, 60(9), 3222-3229.
- Peshala T Gamage, M. K. A., Richard A Sandler, Hansen A Mansy,. (2019). *Modeling Seismocardiographic Signal using Finite Element Modeling and Medical Image Processing*. Paper presented at the IEEE SPMB Philadelphia.
- Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., & Keogh, E. (2014). *Dynamic time warping averaging of time series allows faster and more accurate classification*. Paper presented at the Data Mining (ICDM), 2014 IEEE International Conference on.
- Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3), 678-693.

- Pocock, S. J., Ariti, C. A., McMurray, J. J., Maggioni, A., Køber, L., Squire, I. B., . . . Whalley, G. A. (2013). Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *European heart journal*, 34(19), 1404-1413.
- Rad, A. B., Eftestøl, T., Engan, K., Irusta, U., Kvaløy, J. T., Kramer-Johansen, J., . . . Katsaggelos, A. K. (2017). ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. *IEEE Transactions on Biomedical Engineering*, 64(10), 2411-2418.
- Reinvoio, T., Hannula, M., Sorvoja, H., Alasaarela, E., & Myllyla, R. (2006). *Measurement of respiratory rate with high-resolution accelerometer and EMFit pressure sensor*. Paper presented at the Proceedings of the 2006 IEEE Sensors Applications Symposium, 2006.
- Rokach, L., & Maimon, O. (2005). Clustering methods *Data mining and knowledge discovery handbook* (pp. 321-352): Springer.
- Sahoo, P., Thakkar, H., Lin, W.-Y., Chang, P.-C., & Lee, M.-Y. (2018). On the design of an efficient cardiac health monitoring system through combined analysis of ecg and scg signals. *Sensors*, 18(2), 379.
- Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43-49.
- Sandler, H. A. M., Robert, A. B., William, H. W., Thomas, J. R., Zoujun, D., Ying, P., & Richard, H. (2015). Pneumothorax effects on pulmonary acoustic transmission. doi:10.1152/jappphysiol.00148.2015
- Sandler, R. H., Azad, K., Rahman, B., Taebi, A., Gamage, P., Raval, N., . . . Mansy, H. A. (2019). Minimizing Seismocardiography Variability by Accounting for Respiratory Effects. *Journal of Cardiac Failure*, 25(8), S172.
- Shahbazi, F., & Asl, B. M. (2015). Generalized discriminant analysis for congestive heart failure risk assessment based on long-term heart rate variability. *Computer methods and programs in biomedicine*, 122(2), 191-198.
- Shen, J.-l., Hung, J.-w., & Lee, L.-s. (1998). *Robust entropy-based endpoint detection for speech recognition in noisy environments*. Paper presented at the Fifth international conference on spoken language processing.
- Siejkó, K. Z., Thakur, P. H., Maile, K., Patangay, A., & OLIVARI, M. T. (2013). Feasibility of heart sounds measurements from an accelerometer within an ICD pulse generator. *Pacing and Clinical Electrophysiology*, 36(3), 334-346.
- Silva, D. F., & Batista, G. E. (2016). *Speeding up all-pairwise dynamic time warping matrix calculation*. Paper presented at the Proceedings of the 2016 SIAM International Conference on Data Mining.
- Sørensen, K., Schmidt, S. E., Jensen, A. S., Søgaaard, P., & Struijk, J. J. (2018). Definition of Fiducial Points in the Normal Seismocardiogram. *Scientific reports*, 8(1), 15455.
- Sundaram, S., & Feng, C. (1977). Finite element analysis of the human thorax. *J Biomech*, 10(8), 505-516.
- Taebi, A. (2018). Characterization, Classification, and Genesis of Seismocardiographic Signals.
- Taebi, A., & Mansy, H. A. (2017a). *Grouping similar seismocardiographic signals using respiratory information*. Paper presented at the Signal Processing in Medicine and Biology Symposium (SPMB), 2017 IEEE.

- Taebi, A., & Mansy, H. A. (2017b). Time-frequency distribution of seismocardiographic signals: A comparative study. *Bioengineering*, 4(2), 32.
- Taebi, A., Solar, B. E., Bomar, A. J., Sandler, R. H., & Mansy, H. A. (2019). Recent Advances in Seismocardiography. *Vibration*, 2(1), 64-86.
- Taebi, A., Solar, B. E., & Mansy, H. A. (2018). An Adaptive Feature Extraction Algorithm for Classification of Seismocardiographic Signals. *arXiv preprint arXiv:1803.10343*.
- Tavakolian, K., Portacio, G., Tamddondoust, N. R., Jahns, G., Ngai, B., Dumont, G. A., & Blaber, A. P. (2012). *Myocardial contractility: A seismocardiography approach*. Paper presented at the Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE.
- Tobon-Gomez, C., De Craene, M., Mcleod, K., Tautz, L., Shi, W., Hennemuth, A., . . . Kapetanakis, S. (2013). Benchmarking framework for myocardial tracking and deformation algorithms: An open access database. *Medical image analysis*, 17(6), 632-648.
- Tompkins, J. P. W. J. (1985). A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3), 230 - 236. doi:10.1109/TBME.1985.325532
- Venkatesan, C., Karthigaikumar, P., Paul, A., Satheeskumaran, S., & Kumar, R. (2018). ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications. *IEEE Access*, 6, 9767-9773.
- Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17(4), 395-416.
- Wick, C. A., Su, J.-J., McClellan, J. H., Brand, O., Bhatti, P. T., Buice, A. L., . . . Tridandapani, S. (2012). A system for seismocardiography-based identification of quiescent heart phases: implications for cardiac imaging. *IEEE Transactions on Information Technology in Biomedicine*, 16(5), 869-877.
- Wilson, R. A., Bamrah, V. S., Lindsay Jr, J., Schwaiger, M., & Morganroth, J. (1993). Diagnostic accuracy of seismocardiography compared with electrocardiography for the anatomic and physiologic diagnosis of coronary artery disease during exercise testing. *The American journal of cardiology*, 71(7), 536-545.
- Yu, C.-M., Wang, L., Chau, E., Chan, R. H.-W., Kong, S.-L., Tang, M.-O., . . . Lau, C.-P. (2005). Intrathoracic impedance monitoring in patients with heart failure: correlation with fluid status and feasibility of early warning preceding hospitalization. *Circulation*, 112(6), 841-848.
- Zakeri, V., Akhbardeh, A., Alamdari, N., Fazel-Rezai, R., Paukkunen, M., & Tavakolian, K. (2017). Analyzing seismocardiogram cycles to identify the respiratory phases. *IEEE Transactions on Biomedical Engineering*, 64(8), 1786-1792.
- Zanetti, J., Poliac, M., & Crow, R. (1991). *Seismocardiography: waveform identification and noise analysis*. Paper presented at the [1991] Proceedings Computers in Cardiology.
- Zanetti, J. M., & Tavakolian, K. (2013). *Seismocardiography: Past, present and future*. Paper presented at the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

- Zelnik-Manor, L., & Perona, P. (2005). *Self-tuning spectral clustering*. Paper presented at the Advances in neural information processing systems.
- Zeng, Y., Yager, D., & Fung, Y. (1987). Measurement of the mechanical properties of the human lung tissue. *Journal of biomechanical engineering*, 109(2), 169-174.
- Zhang, Z., Tang, P., Huo, L., & Zhou, Z. (2014). MODIS NDVI time series clustering under dynamic time warping. *International Journal of Wavelets, Multiresolution and Information Processing*, 12(05), 1461011.
- Zhao, Z., Särkkä, S., & Rad, A. B. (2018). *Spectro-temporal ECG analysis for atrial fibrillation detection*. Paper presented at the 2018 IEEE 28Th international workshop on machine learning for signal processing (MLSP).
- Zigras, T. C. (2007). *Biomechanics of human pericardium: a comparative study of fresh and fixed tissue*. McGill University Doctoral dissertation.