Data in Brief 61 (2025) 111889

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

MSCardio seismocardiography dataset: Initial insights from remote monitoring of cardiovascular-induced chest vibrations via smartphones

Amirtahà Taebi[a,b,*], Mohammad Muntasir Rahman[a,b]

[a] Department of Bioengineering, Lehigh University, 7 Asa Drive, Suite 205, Bethlehem, PA 18015, USA
[b] Biomedical Engineering Program, Mississippi State University, 130 Creelman Street, Mississippi State, MS 39762, USA

ARTICLE INFO

Article history:
Received 12 March 2025
Revised 13 June 2025
Accepted 9 July 2025
Available online 17 July 2025

Dataset link: MSCardio Seismocardiography Dataset (Original data)

Keywords:
mHealth
Remote monitoring
Seismocardiography
Cardiovascular monitoring
MSCardio study
SCG dataset

ABSTRACT

The Mississippi State Remote Cardiovascular Monitoring (MSCardio) Study leverages smartphones to remotely collect seismocardiography (SCG) and gyrocardiography data from participants. The study employs an app that uses embedded sensors to record multi-dimensional vibrations, along with self-reported lifestyle and health data. Metadata is also captured to provide contextual information for analysis. This paper outlines the progress made during the first 250 days of the study, including the methodology, data collection process, and preliminary findings. Additionally, the SCG dataset collected during this period is publicly shared to support further research. These results highlight the potential of smartphone-based systems for scalable and accessible cardiovascular health monitoring.

* Corresponding author.
E-mail address: amt525@lehigh.edu (A. Taebi).
Social media: @mitatae (A. Taebi), @muntasir_6 (M.M. Rahman)

https://doi.org/10.1016/j.dib.2025.111889

Specifications Table

Subject	Health Sciences, Medical Sciences & Pharmacology
Specific subject area	Remote Cardiac Monitoring.
Type of data	Seismocardiogram (SCG) signals in CSV format and demographic data in JSON format.
Data collection	Data was collected remotely using smartphone devices. Participants were recruited via a study website and received an automated email with an informed consent form and app installation instructions. After providing consent, they downloaded the study app, completed a demographic questionnaire, and followed standardized procedures to record SCG signals. While lying down, they placed the smartphone on their chest and recorded chest vibrations. The collected data was then uploaded to a cloud storage.
Data source location	Biomedical Engineering Program, Mississippi State University, Mississippi State, MS 39762, USA.
Data accessibility	Repository name: MSCardio Seismocardiography Dataset [1] Data identification number: https://doi.org/10.5281/zenodo.15657893 Direct URL to data: https://github.com/TaebiLab/MSCardio
Related research article	None.

1. Value of the Data

- This dataset provides a unique collection of remotely recorded SCG signals from a diverse population, making it valuable for studying cardiovascular health in real-world settings. Unlike traditional SCG datasets collected in controlled environments, this large-scale dataset leverages smartphone sensors to integrate SCG signals with demographic data, providing valuable insights into physiological variations across different populations.
- The dataset can be used to study intra-subject variability over time, assess trends in cardiovascular health across different populations, and compare SCG signals from various smartphone models.
- Researchers can use this dataset to develop machine learning models for SCG signal analysis, noise reduction, feature extraction, and automated cardiovascular monitoring.
- This dataset supports the development of smartphone-based cardiac monitoring technologies.

2. Background

Cardiovascular diseases remain the leading cause of morbidity and mortality worldwide, accounting for an estimated 17.9 million deaths annually [2]. Early detection and continuous monitoring are critical to reducing the burden of these conditions. Traditional methods for cardiovascular monitoring often rely on resource-intensive clinical setups, which limit accessibility and scalability. With the rapid growth of mobile health (mHealth) technologies, there is an opportunity to address these challenges through cost-effective and user-friendly solutions that enable remote monitoring of cardiovascular health. Seismocardiography (SCG), a non-invasive technique that captures chest vibrations induced by cardiac activity and respiration, provides valuable insights into cardiac function and hemodynamics, making SCG a promising tool for cardiovascular assessment [3,4]. Advances in smartphone technology have made it feasible to collect SCG signals remotely using built-in sensors, offering an innovative approach to large-scale cardiovascular monitoring [5–7].

The scarcity of open-access SCG databases remains a significant obstacle to advancing research in SCG signal processing and interpretation. While limited, a few publicly available SCG datasets have contributed to progress in this field. The CEBS database, available on PhysioNet, provides synchronized electrocardiogram (ECG), breathing, and SCG signals collected from healthy individuals in a controlled setting [8–10]. The SCG-RHC wearable database, also hosted on PhysioNet, contains wearable SCG recordings acquired during right heart catheterization, offering valuable insights into hemodynamic changes in clinical settings [11]. Additionally, the Multichannel Seismocardiography dataset, available on Mendeley Data, was collected using a multichannel SCG system equipped with 16 three-axis sensors to study the nature and propaga-

tion of cardiac-induced vibrations [12,13]. Moreover, the SCG Signal Processing Dataset, hosted on IEEE DataPort, includes SCG recordings from subjects under controlled conditions, supporting the development and validation of SCG signal processing algorithms [14].

While these datasets provide high-quality SCG signals, they were collected in controlled laboratory or clinical environments using dedicated sensor systems. In contrast, the Mississippi State Remote Cardiovascular Monitoring (MSCardio) study was designed to evaluate the feasibility of remote SCG data collection using widely available smartphone sensors. Our approach enables scalable, real-world cardiovascular monitoring without requiring specialized equipment or controlled experimental conditions. By allowing participants to record SCG data in everyday settings, this study provides a unique dataset that reflects real-world variability in sensor placement, environmental factors, and user interactions. This paper reports the progress of the study in the first 250 days, focusing on the methodologies used, the demographics of the study population, the types of data collected, and preliminary findings. By addressing the technical and logistical challenges encountered in this study, we hope to contribute to the growing body of knowledge in remote cardiovascular monitoring and provide a foundation for future innovations in this field.

3. Data Description

3.1. Study population

During the first 250 days of the MSCardio study (April 26, 2024 - January 1, 2025), a total of 253 online interest forms were submitted through the study webpage. However, this number does not represent unique individuals, as some participants may have submitted multiple forms or entered their contact details, e.g., email address and phone number, separately. Ultimately, 123 subjects successfully enrolled in the study via the app. Among them, 108 subjects (46 males, 61 females, and one individual who did not provide gender information), ranging in age from 18 to 62 years, uploaded a total of 515 recordings to the study's cloud storage. However, 10 participants submitted the same recordings multiple times, bringing the total number of unique recordings to 502 after filtering out redundant recordings. Fig. 1 illustrates the age and race distribution of the study population, while the distribution of recordings per participant is shown in Fig. 2. This data shows that the majority of the participants (74.1 %) were in the range of 18–24 years old. Additionally, 77.8 %, 8.3 %, and 8.3 % of the participants identified themselves as White, Asian, and Black or African American, respectively.

Participants used a variety of iOS and Android devices, with iOS smartphones comprising the majority (87.9 %). A histogram displaying the distribution of platforms and phone models is presented in Fig. 3. All participants were required to report their height and weight in centimeters and kilograms, respectively. However, some entries contained implausible values, possibly due to typographical errors or confusion between units (e.g., misreporting weight in pounds instead of kilograms or height in feet instead of centimeters).

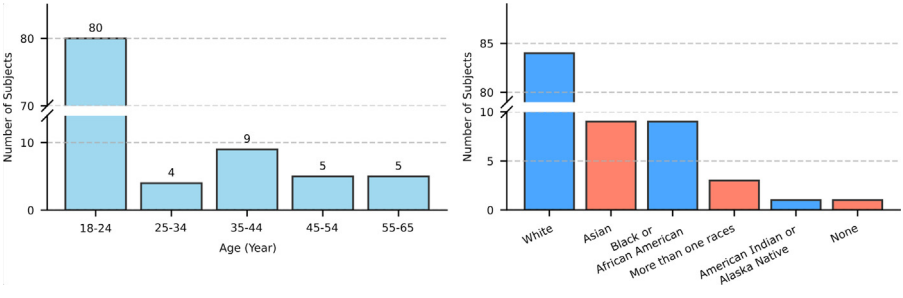

Fig. 1. The distribution of age (left panel) and race (right panel) of the study population.

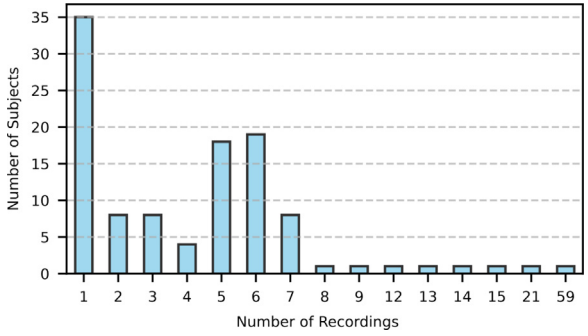

Fig. 2. Histogram of the number of recordings per participants.

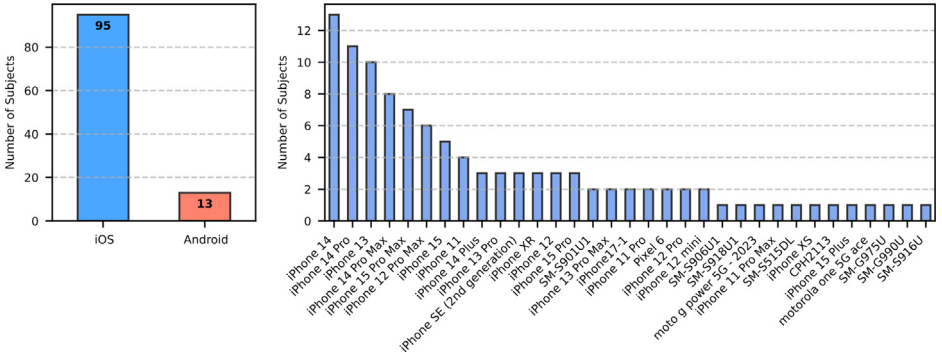

Fig. 3. The distribution of phone platforms and models.

3.2. Signal data

The app utilizes smartphone-embedded sensors to record axial vibrations in three directions (SCG), as illustrated in Fig. 4, with these signals publicly available through this paper. Additionally, the MSCardio study records other key data, including rotational velocity in three directions (GCG), gravity, and geolocation information (latitude, longitude, and altitude), also illustrated in Fig. 4. This additional dataset will be shared publicly in the future. The signals are recorded at the maximum sampling frequency supported by the participant's device, which may vary across different smartphone models. This flexibility ensures the app adapts to diverse hardware capabilities. To maintain temporal accuracy, the app ensures synchronization among the recorded signals, preserving their interrelationships for reliable analysis. Since participants can use any model of iPhone or Android device, differences in sensor quality, orientation, and maximum sampling frequency across devices should be carefully managed during data preprocessing to minimize variability and ensure consistent data quality.

The app supports recording both calibrated and raw (uncalibrated) SCG signals. Both versions are included in the dataset. By default, the SCG signals are calibrated and processed by the phone's operating system. However, upon inspecting the dataset, we identified missing data in some recordings. Specifically, calibrated SCG data were absent in recordings 3 and 4 of Subject 23, and in all five recordings of Subject 39. Additionally, all two recordings from Subject 34 lacked the uncalibrated SCG data. The uncalibrated SCG signals include the low frequency chest vibrations related to respiration, and therefore, can provide insight about SCG variations during a respiration cycle as documented in previous studies [15,16].

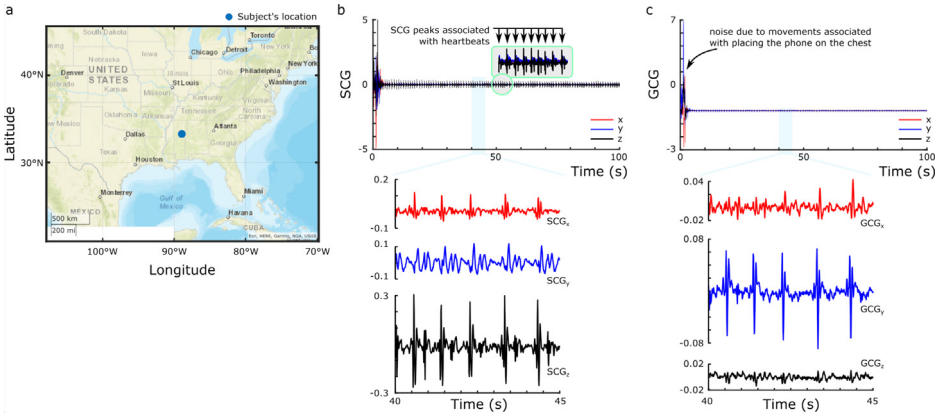

Fig. 4. Remote data collection in the MSCardio study using participants' smartphones. (a) Participant location data. (b) Sample tri-axial SCG signals. (c) Sample tri-axial GCG signals. This paper publicly releases the SCG dataset, while location and GCG data will be made available in future publications.

3.3. Self-reported data

In addition to sensor-based data, the study collects self-reported information from participants. Upon enrolling in the app, participants complete a comprehensive survey covering demographic and lifestyle factors, such as age, height, weight, sex, race, education level, physical activity, coffee consumption, sugary beverage consumption, alcohol consumption, smoking habits, sleeping behavior, and history of heart disease. After each recording session, participants are prompted to answer a brief set of questions about the time and context of the recording. These questions include adherence to the protocol, perceived stress level, symptoms experienced, and the time elapsed since last consuming alcohol, coffee, sugary beverages, or smoking. The self-reported data is integrated with the recorded signals to enable correlation analyses, such as assessing how lifestyle factors or stress levels may influence cardiovascular signals. However, self-reported data introduces potential challenges, including inaccurate or incomplete responses due to recall bias or participant fatigue. To address these issues, the app is designed to provide clear instructions and a user-friendly interface to encourage accurate and consistent reporting. While this data has already been collected, it will not be published in this paper as additional validation and quality control measures are needed to ensure its accuracy and reliability. However, we plan to release the dataset in future publications once these steps have been completed.

3.4. Metadata

Each recording contains metadata, including the device model (e.g., iPhone Pro Max), recording time (UTC), time zone, and platform (e.g., iOS or Android). These metadata provide valuable insights into potential variations in signal quality and sampling frequency across different devices and platforms, aiding in standardizing the analysis pipeline. The recording time and time zone data also enable the investigation of temporal patterns, such as circadian influences on cardiovascular signals.

3.5. File structure

The dataset is systematically organized in a hierarchical directory structure to facilitate easy access and analysis. The root directory contains two main folders:

I. ***info/*:** This folder contains metadata and general information about the dataset.
   - *all subject data.csv*: It includes details about the smartphone devices and platforms (iOS or Android) used by each subject during data collection. Additionally, it provides demographic information (gender, race, age, height, and weight) for all subjects, consolidated in a CSV file.

II. ***MSCardio/*:** This folder contains the recorded data, structured by subject and individual recordings. Each subject has a dedicated folder named Subject XXXX/, where XXXX represents the subject ID. Within each subject folder, multiple recordings are stored in separate subfolders named Recording XXX/, where XXX denotes the recording number. Additionally, each subject folder includes general metadata.
   - ***Subject XXXX/*:**
     - *general_metadata.json*: A JSON file that stores demographic and device-related information, including gender, race, age, height, weight, smartphone model, and platform (iOS or Android).
     - *Recording XXX/*:
       - *scg.csv*: A CSV file storing SCG data collected from smartphone sensors.
       - *Uncalibrated_scg.*csv: A CSV file containing raw (uncalibrated) SCG data.
       - *recording_metadata.json*: A JSON file that records the UTC timestamp of data collection, along with the participant's local time zone at the time of recording.

4. Experimental Design, Materials and Methods

4.1. Study design

The MSCardio study is a comprehensive initiative that aims to collect a large dataset of SCG signal from a diverse population. The study was approved by the Institutional Review Board (IRB) of Mississippi State University and began on April 26, 2024. The study incorporates observational, exploratory, cross-sectional, and longitudinal approaches, providing a flexible framework for participants to engage either once or multiple times. This design facilitates the study of cardiovascular signal patterns across a wide range of demographic and lifestyle variables, as well as over time.

The primary objectives of the study are to generate a large and diverse public dataset of SCG signal for investigating these signals under various conditions, analyze inter- and intra-subject variability in these signals over time, and assess the feasibility of smartphone-based remote data collection. Key outcome measures include, but are not limited to, signal quality, participant adherence, data completeness, and the analysis of SCG and GCG signal patterns across diverse demographic groups and varying conditions. To achieve these objectives, the study includes adults aged 18 years or older with access to an Android or iPhone smartphone. Individuals are asked not to participate if they are under 18, lack smartphone access or the ability to operate one, or have medical conditions that would prevent safe participation, such as severe skin conditions. Recruitment is carried out through social media platforms, email newsletters, flyers, and in-person interactions to maximize diversity in the participant pool.

4.2. Data collection protocol

The recruitment materials direct prospective participants to an online website with detailed information (Fig. 5). If interested, participants provide their email address or phone number (currently limited to U.S. numbers). They then receive an email containing the study's informed consent form, app download and protocol instructions, and a unique code or QR code to enroll. Enrollment in the study and use of the app signifies their informed consent, a step clearly communicated during the process. Participants are asked to collect data at a time and location of their choice. They begin by downloading the app, using the unique code or QR code to enroll,

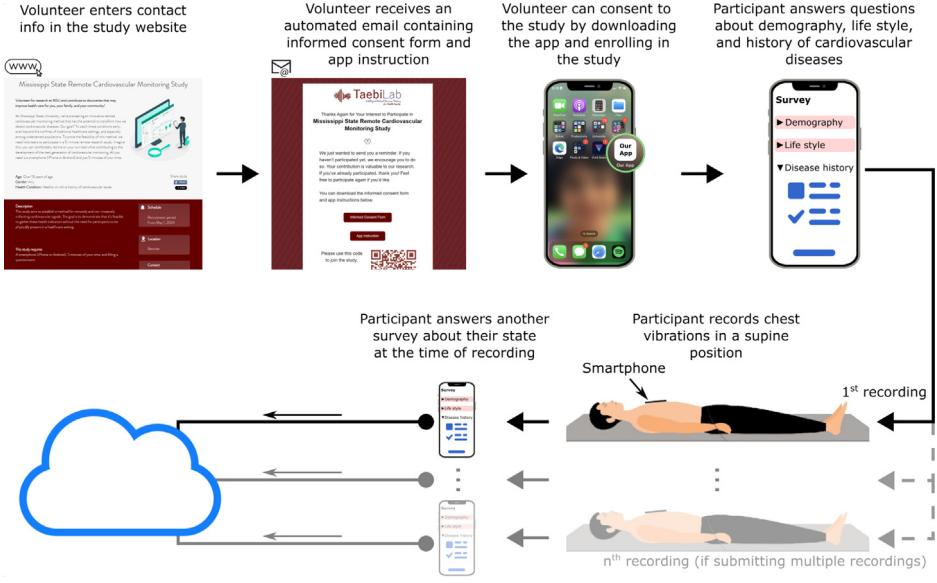

Fig. 5. Data collection protocol of the MSCardio study.

and completing an initial questionnaire within the app. This questionnaire collects demographic information, lifestyle habits, and self-reported cardiovascular health history to provide context for the collected signals.

4.3. Method

Participants are asked to follow the following protocol to record SCG signals. Participants prepare by removing clothing from their upper body and lying on their back on a bed or floor. They position their smartphone at the midpoint between their nipples, ensuring the screen faces upward and the top of the phone is aligned toward their head. A picture of ideal smartphone position is provided in both the informed consent form and app download instruction documents. Using the app, participants start a recording and remain still while breathing normally for approximately two minutes. After completing the recording, they stop the session, name the file, and answer a few additional questions about their lifestyle and adherence to the study protocol. Finally, they are prompted to upload the recording to the study's cloud storage. While participants have access to their recordings within the app, only the research group has access to the cloud storage, which stores the recordings uploaded by different participants. This protocol enables participants to contribute data conveniently while following standardized instructions that ensure consistency and reliability in the collected signals.

Limitations

Several potential technical challenges may arise during the course of the study. One such issue is when a participant switches phones and re-enrolls in the study using the new device, leading to the creation of two distinct records for the same individual. This could pose challenges when dividing data into training and test datasets, as it may inadvertently result in data leakage if a subject's data appears in both sets. This issue could become more significant if many participants switch devices during the study. Another potential challenge arises when a participant

records data for multiple individuals, such as family members, using a single phone. In this case, the recordings could be misclassified as longitudinal data for one person, distorting the analysis of intra-subject variability and affecting the accuracy of conclusions about individual cardiovascular signals. Additionally, the accuracy of self-reported data presents a challenge, as there is no way to independently verify health information, such as a participant's or their family's history of cardiovascular disease. This could lead to biases in interpreting correlations between lifestyle factors and cardiovascular signals. Variability in smartphone sensors is another concern, as different models may have varying sensor qualities, potentially introducing inconsistencies in the recorded data. Environmental factors, such as background noise,participant movement during recordings, or variations in phone position on the chest in longitudinal data, could also affect data quality, complicating the analysis and interpretation of signals.

Despite these challenges, the MSCardio study demonstrates the feasibility of large-scale, remote cardiovascular data collection using widely available smartphones. The dataset shared through this study represents one of the largest publicly available SCG datasets, offering valuable insights into cardiovascular health monitoring in real-world settings. By leveraging a diverse participant pool and real-world conditions, this study paves the way for future advancements in mobile health technologies and data-driven cardiovascular research.

Ethics Statement

The study was approved by the institutional review board of Mississippi State University (# IRB-24–148). Privacy rights of human subjects have been observed and that informed consent was obtained for experimentation with human subjects.

Data Availability

MSCardio Seismocardiography Dataset (Original data) (Zenodo).

CRediT Author Statement

Amirtahà Taebi: Data curation, Software, Validation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Resources, Supervision, Project administration; **Mohammad Muntasir Rahman:** Data curation, Software, Validation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing.

Acknowledgements

The authors thank Ms. Pranavi Paudel and Ms. Hannah Henson for their assistance with subject recruitment.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Taebi, M.M. Rahman, MSCardio seismocardiography dataset (2025). (v1.0) [Data set]. Zenodo. 10.5281/zenodo.15657893.

[2] W. H. Organization, Cardiovascular diseases (CVDs), https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), accessed: 2025-03-03 (2021).

[3] A. Taebi, B.E. Solar, A.J. Bomar, R.H. Sandler, H.A. Mansy, Recent advances in seismocardiography, Vibration 2 (1) (2019) 64–86.

[4] J. Cook, M. Umar, F. Khalili, A. Taebi, Body acoustics for the non-invasive diagnosis of medical conditions, Bioengineering 9 (4) (2022) 149.

[5] W. Wu, M. Elgendi, R.R. Fletcher, H. Bomberg, U. Eichenberger, C. Guan, C. Menon, Detection of heart rate using smartphone gyroscope data: a scoping review, Front. Cardiovascul. Med. 10 (2023) 1329290.

[6] M.M. Rahman, B. Kakavand, W. Van Wurm, W.L. Holman, M.R. Movahed, A. Taebi, From video to vital signs: a new method for contactless multichannel seismocardiography, npj Cardiovascul. Health 2 (1) (2025) 1.

[7] M.M. Rahman, J. Cook, A. Taebi, Non-contact heart vibration measurement using computer vision-based seismocardiography, Sci. Rep. 13 (1) (2023) 11787.

[8] M.A. Garc´ıa-González, A. Argelagós-Palau, M. Fernández-Chimeno, J. Ramos-Castro, A comparison of heartbeat detectors for the seismocardiogram, in: Computing in Cardiology, IEEE, 2013, pp. 461–464. 2013.

[9] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.

[10] M. Garc´ıa-González, A. Argelagós, M. Fernández-Chimeno, J. Ramos-Castro, Differences in qrs locations due to ecg lead: relationship with breathing, in: XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013: MEDICON 2013, 25-28 September 2013, Seville, Spain, Springer, 2014, pp. 962–964.

[11] M. Chan, L. Klein, J. Fan, O. Inan, Scg-rhc: Wearable Seismocardiogram Signal and Right Heart Catheter Database, 2023.

[12] K. Munck, K. Sorensen, J.J. Struijk, S.E. Schmidt, Multichannel seismocardiography: an imaging modality for investigating heart vibrations, Physiol. Meas. 41 (11) (2020) 115001.

[13] K. Munck, K. Sorensen, J. Struijk, S. Schmidt, Data for: multichannel seismocardiography: an imaging modality for investigating heart vibrations, Mendel. Data 2 (2020).

[14] K. R. K M, G. N, Seismocardiography (scg) signal processing dataset (2024). doi:10.21227/ztd5-6f57. 10.21227/ztd5-6f57

[15] P.T. Gamage, M.K. Azad, A. Taebi, R.H. Sandler, H.A. Mansy, Clustering of SCG events using unsupervised machine learning, in: Signal Processing in Medicine and biology: Emerging trends in Research and Applications, 2020, pp. 205–233.

[16] A. Taebi, A.J. Bomar, R.H. Sandler, H.A. Mansy, Heart rate monitoring during different lung volume phases using seismocardiography, in: SoutheastCon 2018, IEEE, 2018, pp. 1–5.