

Research and Applications

Synthetic seismocardiogram generation using a transformer-based neural network

Mohammad Nikbakht ¹, Asim H. Gazi ¹, Jonathan Zia¹, Sungtae An², David J. Lin¹, Omer T. Inan¹, and Rishikesan Kamaleswaran ³

¹Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA, ²Department of Interactive Computing, Georgia Institute of Technology, Atlanta, Georgia, USA and ³Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, Georgia, USA

Corresponding Author: Mohammad Nikbakht, BE, MSc, PhD Candidate, Department of Electrical and Computer Engineering, Georgia Institute of Technology, North Avenue, Atlanta, GA 30332, USA; mohnikbakht@gatech.edu

Received 2 December 2022; Revised 8 February 2023; Editorial Decision 21 March 2023; Accepted 4 April 2023

ABSTRACT

Objective: To design and validate a novel deep generative model for seismocardiogram (SCG) dataset augmentation. SCG is a noninvasively acquired cardiomechanical signal used in a wide range of cardiovascular monitoring tasks; however, these approaches are limited due to the scarcity of SCG data.

Methods: A deep generative model based on transformer neural networks is proposed to enable SCG dataset augmentation with control over features such as aortic opening (AO), aortic closing (AC), and participant-specific morphology. We compared the generated SCG beats to real human beats using various distribution distance metrics, notably Sliced-Wasserstein Distance (SWD). The benefits of dataset augmentation using the proposed model for other machine learning tasks were also explored.

Results: Experimental results showed smaller distribution distances for all metrics between the synthetically generated set of SCG and a test set of human SCG, compared to distances from an animal dataset ($1.14 \times$ SWD), Gaussian noise ($2.5 \times$ SWD), or other comparison sets of data. The input and output features also showed minimal error (95% limits of agreement for pre-ejection period [PEP] and left ventricular ejection time [LVET] timings are 0.03 ± 3.81 ms and -0.28 ± 6.08 ms, respectively). Experimental results for data augmentation for a PEP estimation task showed 3.3% accuracy improvement on an average for every 10% augmentation (ratio of synthetic data to real data).

Conclusion: The model is thus able to generate physiologically diverse, realistic SCG signals with precise control over AO and AC features. This will uniquely enable dataset augmentation for SCG processing and machine learning to overcome data scarcity.

Key words: seismocardiogram, transformer neural networks, machine learning, cardiovascular

INTRODUCTION

The seismocardiogram (SCG) is a cardiovascular mechanical signal that records chest wall acceleration associated with the heart's contraction and ejection of blood.¹ The SCG waveform captures timing features correlated to cardiac events such as aortic valve opening

(AO), and aortic valve closing (AC). These features can be used to derive cardiac time intervals such as pre-ejection period (PEP), and left ventricular ejection time (LVET).² SCG recordings have been successfully used for various cardiovascular health monitoring tasks,

such as heart failure monitoring, detecting the effects of noninvasive neuromodulation and stressors for stress applications, and estimating a variety of hemodynamic variables such as blood pressure and stroke volume.^{3–8}

However, to effectively apply modern machine learning techniques to SCG data, dataset sizes should be relatively large. Dataset diversity is a key factor as well for model generalization.⁹ Yet, collecting such large, diverse datasets from human or animal subjects is often challenging. Aside from approaches such as transfer learning that use alternative sets of data,¹⁰ one common approach to addressing the challenge of limited data, and insufficiently diverse data, is through synthetic data augmentation via generative modeling.^{11,12}

Recent studies on generative models that can generate clinically relevant signal modalities for dataset augmentation have focused on ECG or PPG signals.^{13–16} Prior work on generative models for SCG signals have focused on correlation of dynamic processes using learned latent factors, sensor placement-induced changes, and ballistocardiogram (BCG) to SCG conversion.^{17,18} To the best of our knowledge, however, generative modeling for synthetic SCG signal generation has not been explored in the literature since SCG signals usually have morphologically complex datasets, which are usually smaller than ECG and PPG datasets. In this study, we introduce an SCG generator model based on transformers neural networks that generates synthetic SCG beats from clinically relevant SCG features.¹⁹

From a clinical utility standpoint, the 3 waveforms provide different information that can be fused and harvested towards improving decision support. Specifically, the ECG captures the electrophysiological health of the heart, and can be used to detect arrhythmias, hypertrophy, rate disturbances, and other electrical conduction abnormalities of the heart.²⁰ The PPG captures vascular health by quantifying the blood volume pulsatility at a peripheral site and is generally used to determine oxygen saturation, assessing/diagnosing peripheral artery disease and providing a distal timing reference for pulse transit time and/or pulse wave velocity.²¹ The SCG is a mechanical signal that captures the pumping action of the ventricles primarily, thus providing an indication of hemodynamic function of the heart together with the timings of valve activities.¹ One key difference in the SCG is a greater degree of interparticipant variability in signal morphology, which renders this signal a prototypical example for the training and evaluation of the algorithm presented in this work.¹

Nevertheless, the framework proposed in this work is easily expandable to other cardiovascular signals including ECG and PPG, as well as other cardio mechanical signals such as the ballistocardiogram (BCG) and phonocardiogram.

The contributions of this work include: (1) designing a transformer-based generative model that can be used to generate synthetic SCG beats similar to real human SCG beats; (2) demonstrating that our synthetic SCG generator model can generate SCG beats with AO and AC timings that are strongly correlated with the desired AO and AC parameters input to the model; and (3) showing that the generated synthetic SCG signals can be used for data augmentation during model training to improve performance.

MATERIALS AND METHODS

Datasets

In this work, we used 4 human participant datasets for training and validation, and 1 animal dataset for validation purposes only. The

human participant datasets contain recordings from a total number of 82 participants (32 females, and 50 males). All of these human participant datasets were fully deidentified and were collected previously under protocols approved by the Georgia Institute of Technology Institutional Review Board (IRB). The animal dataset contains recordings from 6 pigs which is used for evaluation purposes. This dataset was collected under a protocol approved by the Institutional Animal Care and Use Committees (IACUC) of the Georgia Institute of Technology, Translational Testing and Training Labs Inc. and the Department of the Navy Bureau of Medicine and Surgery. The datasets demographics are summarized in [Table 1](#) with a detailed description for each dataset in the [Supplementary Materials \(Supplementary Section S1\)](#).

Preprocessing

Noise reduction and segmentation

The z-axis SCG signals from the rest and baseline periods of the protocols were extracted as these segments contain minimal motion noise. Then the signals are filtered using a Kaiser window band-pass filter with cutoff frequencies of 1–40 Hz.^{17,22} The z-axis acceleration also known as the dorso-ventral component of the SCG signal has been focused on as the main SCG component in literature.¹ Using the R-peaks of the ECG signals collected concurrently with the SCG signals, the filtered signals were heartbeat-separated. A signal quality indexing (SQI) method for SCG beats was used to identify and exclude the beats contaminated with noise above a certain threshold.²⁶ Finally, the beats are min-max normalized and centered around 0.5 to ensure model generalization.

AO and AC extraction

The proposed model in this paper receives SCG features as input and generates a realistic SCG beat that has the corresponding features. For each beat in the dataset, AO and AC amplitudes and locations were extracted from the beats using the simplified consistent peak tracking algorithm employed in prior work.⁶ Using the AO and AC amplitudes and locations extracted from the target real SCG beats, a simplistic SCG signal was created consisting of 2 Gaussian waveforms with means located at AO and AC locations, and amplitudes relative to AO and AC amplitudes of the target SCG beat. In this work, we refer to this simplistic signal as a “skeleton” signal. We propose that the model can translate the information embedded in this representation of features to a realistic SCG beat. A sample of the skeleton signal is shown in [Figure 2A](#). For more details about the preprocessing steps, refer to the [Supplementary Section S2](#).

Model architecture

The transformer neural network model architecture relies on the attention mechanism to learn input-output temporal dependencies, rather than recurrence. Transformers have demonstrated superior performance to recurrent neural networks (RNNs) in certain sequence-to-sequence translation tasks while overcoming the vanishing and exploding gradient problems.^{19,27,28} The translation task here, in particular, is translating “skeleton” signals to realistic human-like SCG beats with those features. To leverage this power of transformers for SCG beat generation, we adapted the previously validated text-to-speech (TTS) transformer model with some modifications.²⁸ Since speech, like SCG signals, is a continuous oscillatory waveform which is segmented into its component parts for analysis, we propose that adapting the TTS transformer architecture is a natural choice for creating a generative model for SCG signals.

Table 1. Datasets demographic information

Dataset	Participants count	Age (years)	Weight (kg)	Height (cm)	Duration (min)
(1) Shandhi et al ^{18,22}	26 (10f, 16m)	25.9 ± 3.5	70.4 ± 14.0	171.9 ± 10.9	14.5
(2) Hersek et al ¹⁸	10 (5f, 5m)	21.9 ± 0.6	65.4 ± 9.9	172.3 ± 9.8	16
(3) Ashouri et al ²³	10 (5f, 5m)	24.7 ± 2.3	70 ± 10.5	170 ± 11.6	20
(4) Gurel et al ⁴	16 (6f, 10m)	26.7 ± 3.2	N/A	N/A	24
(5) Chan et al ²⁴	20 (6f, 14m)	26.52 ± 2.5	71.9 ± 14	173.7 ± 9.4	30
(6) Zia et al ²⁵	6 (pigs)	N/A	87.2 ± 35.7	N/A	10

In order to relate SCG signals to the NLP paradigm, we first converted SCG beats into a sequence of embedding vectors. Prior work for the text-to-speech task used mel spectrogram as the fixed embedding method for speech signals followed by an additional trainable embedding layer.²⁸ To apply the same notion for SCG signals, we proposed 3 types of fixed embeddings for SCG signals: spectrogram, maximal overlap discrete wavelet transform (MODWT), and a pre-trained encoder (Figure 2B). The prenetworks before the encoder and decoder blocks operate as the trainable part of the embedding layer after the fixed embeddings to allow for the projection of the fixed embeddings to a more flexible subspace as suggested by Li et al²⁸ (Figure 2C and E). Supplementary Table S1 shows a summary of the model architecture and each block is explained in the Supplementary Materials (Supplementary Section S3).

Another important clinical feature of SCG signals is morphology. The morphology of an SCG beat is affected by several factors such as sensor placement, respiration, and interparticipant variation; however interparticipant variability specifically has a major impact on morphology variation.^{1,29} To consider interparticipant morphology variation in the model presented in this work, we appended a random ID token to the input tokens (similar to the [CLS] token in vision transformers), proposing that this token is responsible for the SCG morphology variation and decouples the participant-specific morphology information from other SCG features.³⁰ This synthetic ID is kept constant for all beats belonging to each participant during training to keep intraparticipant similarities while varying the synthetic ID for different participants to account for interparticipant variabilities.

Training setup

We used a single Nvidia GeForce GTX 1080 GPU to train our model on the SCG beats dataset (64 894 training samples) extracted from a total of 82 healthy participants. L1-norm was chosen as the loss function to minimize the error between the generated SCG waveform and the ground truth SCG waveform from the human participant signals Equation (1).

$$\mathcal{L}(\theta) = \frac{1}{N_{mb}} \sum_i |f_\theta(x_i) - y_i|_1 \quad (1)$$

where N_{mb} is the number of mini-batch samples, θ is the parameters of the network, f_θ is the network model with parameters θ , x_i is the input skeleton, and y_i is the output ground truth SCG beat. The $f_\theta(x_i)$ and y_i are 1D SCG signals with \mathbb{R}^m dimension, where m is the length of the signal.

Adam optimizer was used with the addition of a learning rate warm-up algorithm introduced by Ashish et al¹⁹ in which the learning rate is increased linearly for a number of steps and then decreases proportional to the inverse square root of the number of steps.

For the training of this model, 12 participants were held out for the test set, 6 for validation and the rest used for training. The model

was then trained on the training split and hyperparameters were tuned using the validation set (Supplementary Section S4). After tuning the hyperparameters, we merged the training and validation splits and retrained the model on this merged dataset for 65 epochs and tested using the held out test set.

Evaluation

Generative model evaluation

To evaluate the model introduced in this work, distribution distance metrics were used to quantify the closeness of the generated synthetic SCG beats to real human SCG beats. For this, we created 6 datasets: a training dataset, which is the same dataset that the model was trained with; a test dataset, which is our held out test set; a pig dataset which contains SCG beats recorded from 6 pigs; a skeleton dataset containing skeleton input signals (Figure 2A) that were fed as input to the model; a noise dataset containing Gaussian noise; and a synthetic SCG beat dataset that contains synthetic beat samples generated by the proposed model.

The 3 distance metrics chosen were Maximum Mean Discrepancy (MMD), Sliced-Wasserstein Distance (SWD), and Kullback-Leibler divergence (KLD) which were used in prior work as metrics for evaluating generative models for signal synthesis.^{14,31,32} Using each metric, distances were calculated between each pair of these 6 datasets. These metrics are explained in Supplementary Materials (Supplementary Section S5).

Dataset augmentation evaluation

In order to study the benefit of using the proposed work to augment training datasets with synthetic data, we added different amounts of synthetic data generated by the model introduced in this work, to a dataset of human SCG signals. And compared the performance of models trained for an SCG related task on a held out set of human SCG signals. We reproduced a prior work that used SCG signals for PEP estimation and augmented the dataset using synthetic data generated by the model.³³ The synthetic beats for augmentation are generated with random morphologies and clinical features controlled using the model inputs.

RESULTS

A high level block diagram of the synthetic SCG generator designed in this work is illustrated in Figure 1. The model receives 2 groups of inputs, a synthetic participant identification (ID) illustrated in Figure 1A that is responsible for keeping the SCG morphology unique to the participant, and clinically relevant features that operate as figurative knobs that control the AO and AC features of the generated SCG beat (Figure 1B). This separation of inputs decouples participant-specific morphology information from other features of the SCG beat. Figure 1C represents the generative model which receives Figure 1A and B as inputs and outputs a realistic synthetic

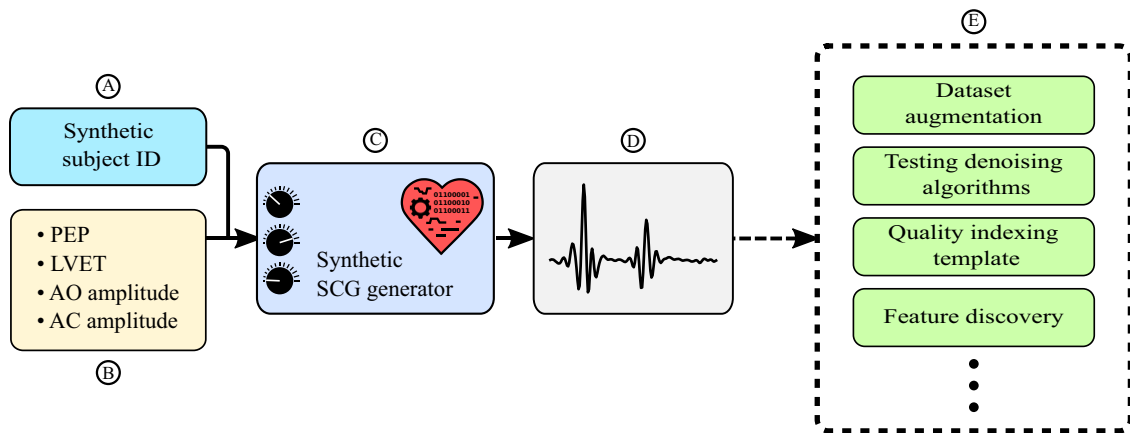


Figure 1. Overview of the synthetic seismocardiogram (SCG) generator model. (A) A participant identification (ID) is fed as input which controls the morphology of the generated beat. (B) The model receives SCG features as inputs, and outputs an SCG beat with those features. (C, D) This model can be used to generate sequences with controlled features. (E) This enables tasks such as dataset augmentation and template generation. It also sets the stage for simulating noisy SCG and testing denoising algorithms. PEP: pre-ejection period; LVET: left ventricular ejection time; AO: aortic opening; AC: aortic closing.

SCG beat shown in Figure 1D. Figure 1E lists the potential applications of the model. Although the primary application of this model is data augmentation,^{11,12} other applications include the development of denoising algorithms by adding noise to clean synthetic SCG beats,³⁴ treating the synthetic beats as ground truth (similar to speech denoising^{35,36}).

The designed transformer-based architecture is shown in Figure 2 resulted from modifying the TTS model in prior work to adapt to the task of generating SCG beats. The main modifications include the pre-net and post-net blocks as well as the embedding block to optimize the performance for SCG generation from clinical features. The architecture is discussed in details in the [Supplementary Material](#).

To validate the designed architecture based on the main hypotheses that the synthetic SCG beats generated are realistic and human-like, have controllable features, and can be used for dataset augmentation to enhance model performance, we performed a series of evaluations.

Realistic outputs

Distribution distance metrics explained in “Evaluation” were used to quantify the similarity of the synthetically generated beats to real SCG beats. For this purpose, the 6 datasets explained in “Evaluation” were used and distances were calculated between each pair of these datasets using the 3 distance metrics. Figure 3B and C illustrates the results of these calculations. Figure 3B visualizes the SWD distances between each pair of datasets. And Figure 3C shows the distances between the synthetic SCG dataset (source) to the remaining 5 datasets (targets) using all 3 distance metrics. These results show smaller distribution distances between generated SCG beats and human SCG signals compared to animal SCG ($1.14\times$ SWD), skeleton signals, and Gaussian noise ($2.5\times$ SWD) for all distance metrics.

A t-Distributed Stochastic Neighbor Embedding (t-SNE) plot of the 6 datasets is also presented in Figure 3A to visualize how synthetic beats cluster with beats from other datasets. We can observe that the synthetically generated beats cluster closer with the real human beats compared to other datasets.

Controllable output features

To validate that the generative model presented here can generate synthetic SCG beats with controlled clinical features, we fed in a series of 16 input sequences (different random ID tokens representing 16 fake participants; Figure 1A) each with linearly varying PEP and LVET parameters and random amplitudes. These values were chosen from clinically meaningful ranges extracted from the training dataset (64 human participants from 5 datasets). AO locations from the R-peak (PEP) are in the range of 65.74 ± 25.58 ms, and AC locations from the R-peak (PEP+LVET) are in the range of 333.93 ± 46.80 ms.

After feeding the generated outputs to our preprocessing pipeline (see “Preprocessing”), we extracted SCG features from these generated beats. Figure 4B and C shows Bland-Altman plots for PEP and LVET errors between the generator-output and the input skeleton signals. Notably, this figure shows 95% limits of agreement of 0.03 ± 3.81 ms for PEP error and -0.28 ± 6.08 ms for LVET error between the input output features.

Qualitatively, Figure 4A shows a sequence of generated SCG beats from a fake participant plotted on top of each other with AO and AC annotations. Note that the random ID token is kept constant for all beats belonging to 1 fake participant but it is varied between participants.

Dataset augmentation

We propose that by augmenting human SCG datasets with synthetic data that is very similar to human SCG signals, we can improve the performance of data-driven algorithms. For this, we reproduced prior work on PEP estimation in which several regression models were trained on a dataset of SCG signals to estimate the PEP value on a beat-by-beat basis.³³ Figure 5 shows the RMSE for PEP estimates with different ratios of synthetic SCG data to previously acquired human SCG data. Notably, the figure shows that when augmenting the dataset using the synthetic generator model, every 10% augmentation (ratio of synthetic data to real data) results in a 3.3% accuracy improvement on an average.

Embeddings comparison

The embeddings play an important role in the training of the transformer model and operate as a bridge between the input 1D signal

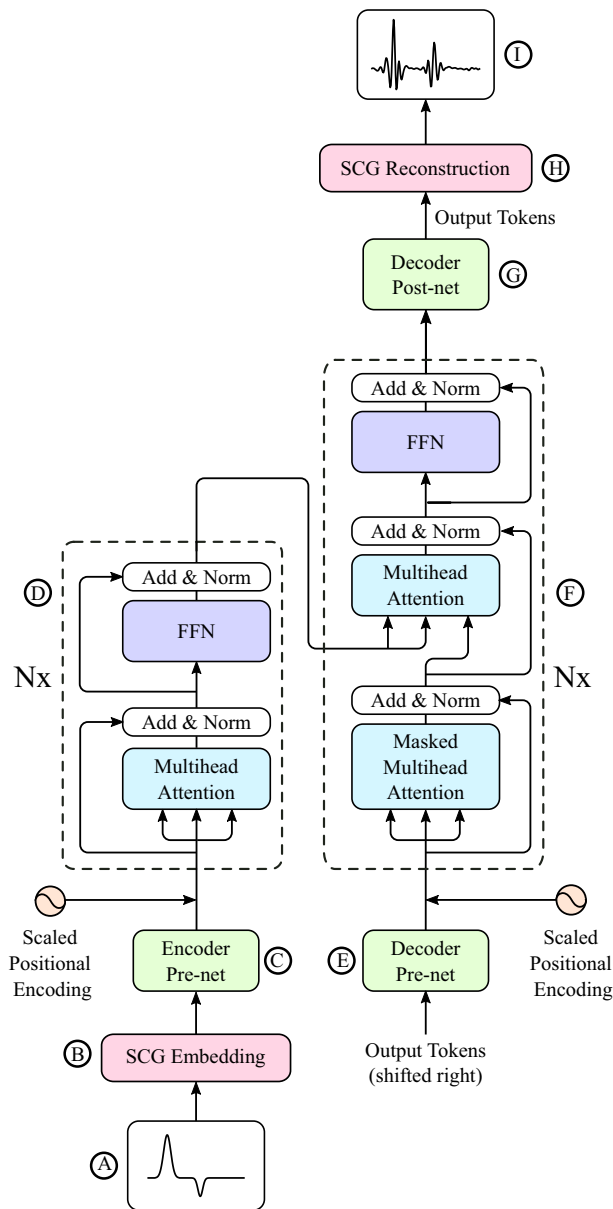


Figure 2. Architecture of the model, which is adapted from the text-to-speech transformer architecture for SCG feature-to-beat generation. (A) A simplistic signal consisting of 2 Gaussian waveforms at AO and AC locations with AO and AC amplitudes (referred to as "skeleton") is fed into the encoder. (B–D) The fixed embedding block and the encoder pre-net first convert input tokens to embedding vectors which are then fed into the transformer block. (E–G) The decoder generates an SCG beat with the same features. (H, I) At the output of the decoder a reconstruction block will convert the generated tokens with embeddings to an SCG beat. Nx: repeat block N times; FFN: feed forward network.

and the input that the transformer model receives. We proposed 3 different fixed embedding methodologies (MODWT, spectrogram, and a pretrained encoder) and compared their performances for the task of generating SCG beats. For evaluation, we calculated the distances between the synthetically generated SCG beat dataset and the human SCG signal datasets (train, and test) using the 3 distance metrics (SWD, MMD, and KLD) and repeated this for each embedding. Figure 6 shows that the model with MODWT embedding outperforms models with spectrogram and pretrained encoder embeddings.

Demystifying the transformer model

Interpretable machine learning is defined as the extraction of information about relationships learned by the black-box model.³⁷ To demystify the network architecture presented in this work, the attention weights of the transformer block are extracted and visualized by heatmaps in Supplementary Figure S3 and are discussed in the Supplementary Materials (Supplementary Section S7).

DISCUSSION

Prior work has shown that generative models have been successful in generating clinically relevant signal modalities for dataset augmentation. For instance, Zhu et al¹³ designed a bidirectional long short term memory convolutional neural network (LSTM-CNN) model to generate synthetic ECG data for dataset augmentation. Delaney et al¹⁴ instead used generative adversarial networks (GAN) to generate ECG signals. Hazra et al¹⁵ and Kiyasseh et al¹⁶ designed GAN models for photoplethysmogram (PPG) data augmentation.

To the best of our knowledge, however, generative modeling for synthetic SCG signal generation has not been explored in the literature. Two recent studies explore related topics. Zia et al¹⁷ demonstrated the correlation of dynamic processes and sensor placement with learned latent factors. Hersek et al¹⁸ introduced a U-Net model that generates ballistocardiogram (BCG) signals from SCG signals. Although both approaches involve some form of deep generative modeling for SCG, neither explored synthetic SCG generation from SCG features to enable data augmentation. This work is the first to show that the designed transformer-based architecture can be trained for synthetic SCG generation.

Evaluation results

By evaluating the model using distribution distance metrics, we demonstrate its ability to generate SCG beats that obey distributional characteristics more similar to real human SCG. The SWD results (Figure 3B) agree that the generated beats are closer to real SCG beats rather than random noise and also agree that the synthetic dataset is slightly closer to real human dataset compared to animal (pig) SCG dataset. In Figure 3C, the results from all 3 metrics agree that the synthetic dataset is closer to real SCG signals (human and pig) and is farther from non-SCG signals (skeleton, and noise) which implies that the model is generating realistic SCG beats. Further, focusing on the datasets that contain real SCG signals (train, test, and pig), we observe that the synthetic beats are closer to human SCG signals and farther from pig SCG signals. This entails that the proposed generative model is capable of generating SCG beats from a distribution closer to human SCG distribution rather than non-SCG or animal SCG distributions. From the t-SNE representation (Figure 3A), it can be seen that the synthetically generated SCG beats and real human SCG beats cluster together thus making them more difficult to separate. However, the animal, skeleton, and noise samples clusters can be more easily separated from the synthetically generated SCG beats.

Figure 4 shows strong correlations between features of the generated SCG beats and the desired features fed into the model. We can observe that the 95% limits of agreement for PEP and LVET are 0.03 ± 3.81 ms and -0.28 ± 6.08 ms, respectively. The increase in LVET error compared to PEP may be due to increased susceptibility of AC peaks to motion artifacts. In addition, comparing the amplitudes of the generator-output and the input skeleton signals, we achieved an R^2 of 0.71 for AO amplitudes and 0.43 for AC

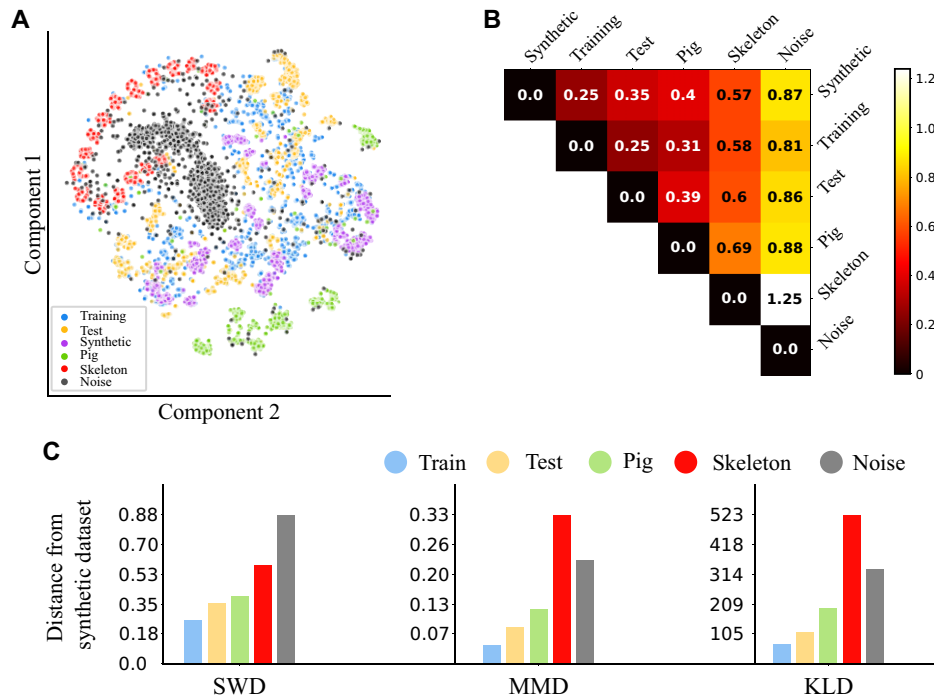


Figure 3. (A) t-SNE plot of SCG data from 6 different datasets. (B) Sliced Wasserstein distance (SWD) between the 6 datasets. (C) Bar plots of synthetic dataset distance from the other 5 datasets using 3 distance metrics. MMD: maximum mean discrepancy; SWD: sliced Wasserstein distance; KLD: Kullback-Leibler divergence.

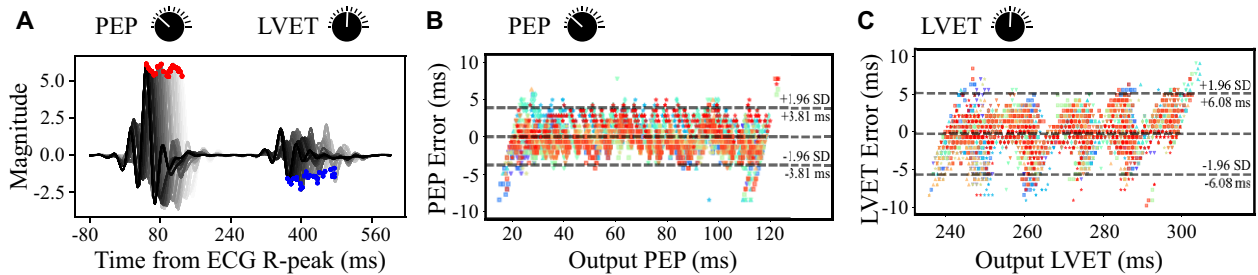


Figure 4. A generated sequence sample with PEP and LVET increasing linearly in a valid range. (A) segmented beats with AO, and AC annotations in the time range 0–250 ms and 250–500 ms respectively (constant amplitude for better visualization). (B, C) Bland-Altman plots of error between derived PEP, and LVET from synthetically generated SCG beats from 16 fake participants compared to the input parameters of the model picked from clinically valid ranges.

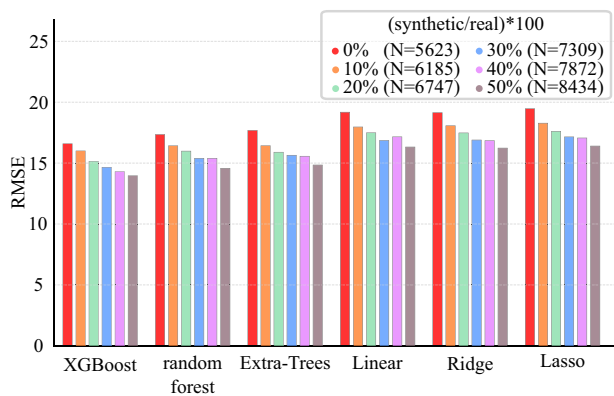


Figure 5. Comparing the RMSE for a PEP estimates task using the same models as in prior work³¹ with the dataset being augmented with different amounts of synthetic data generated by the proposed model.

amplitudes. The R^2 value for AC is lower due to the fact that AC has a lower SNR than AO and is notoriously difficult to annotate.³⁸

Importantly, beat morphology could be randomized through a random ID token (Supplementary Figure S1), making the model desirable for dataset augmentation by introducing diversity in the morphology of the SCG signals, while remaining in control over physiological variation. Further, we showed that the intrain-participant variability is less than inter-participant variability (Supplementary Section S6).

The results of the dataset augmentation experiment (Figure 5) implies that augmenting the training dataset with the synthetic data generated by the model introduced in this work can enhance the performance of the models on a held out test set. In addition, we observed that increasing the amount of synthetic data to the real data also improved the performance. Further, we showed that simply adding data without varying PEP and LVET does not affect the model performances, demonstrating that the dataset diversity

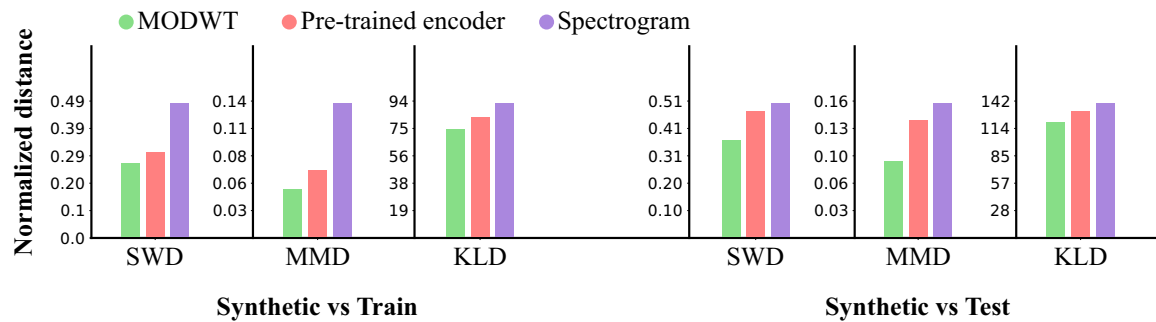


Figure 6. The distance between the synthetically generated SCG beats using 3 different embeddings using the 3 distance metrics explained in “Evaluation.”

enabled by our generator model is what drives the improved performance (Supplementary Figure S2).

A key contribution to the performance was the choice of embedding for SCG signals to make them transformer-readable. Using MODWT as an embedding for SCG signals enabled the model to learn the distribution of human SCG beats better (Figure 6).

Limitations and future work

This study has some limitations: (1) The performance of the model is limited for SCG AO and AC amplitudes (especially AC) as amplitudes are dependent on factors such as chest medium and sensor placement. Future work can explore model modifications and learn these dependencies to achieve better amplitude accuracy; (2) In this work, we focused on SCG beats from healthy population for the training of the model, thus the model is able to generate realistic beats analogous to healthy human SCG beats. Future work can explore synthetic generation of SCG signals for “fake” participants with different health conditions such as heart failure (HF), or post-traumatic stress disorder (PTSD) as these conditions have shown to produce heterogeneity in the SCG signal.^{6,7} However, this requires a large datasets of SCG signals collected from patients with such conditions.

In terms of applications, other than data augmentation, the model presented in this work can be used for testing denoising applications, quality indexing, and feature discovery. For testing denoising algorithms, a sequence of clean SCG beats with controlled features affected by real-time conditions can be generated to be fed into SCG simulators,³⁴ enabling noisy SCG signal collection to explore denoising methods. Some SCG quality indexing algorithms require a set of SCG beat templates with clean features.²⁶ Using the model presented in this work, we can generate clean SCG beat templates with diverse morphology and controllable features resulting in a richer template set, therefore enabling a more accurate quality indexing.

The framework proposed here in general can be expanded to other applications as well where data can be hard to collect. Specifically, this framework can be directly applied to any continuous, quasi-periodic signal. And further, the framework and the transformer architecture can be adapted to other medical domains such as remote patient monitoring.

CONCLUSION

We designed and validated a novel deep generative model based on transformer neural networks that was trained on a large combined dataset of real human SCG signals. This model generates synthetic

SCG signals analogous to human SCG with controlled clinical features. The use of synthetic SCG generator models such as the one elucidated herein can help increase the size of datasets by appending SCG signals with physiological and morphological diversity relevant to the task. This could not only be used for data augmentation and machine learning tasks, but could also be used for a variety of clinical tasks that require physiological diversity manifested in SCG signals (eg, medical training).

FUNDING

This work is based on material supported in part by the National Institutes of Health grant number R01GM139967; and the Office of Naval Research, grant numbers N00014-21-1-2036 and N00014-22-1-2325. The work of AHG was supported by a National Science Foundation Graduate Research Fellowship grant number DGE-2039655.

AUTHOR CONTRIBUTIONS

MN: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization; AHG: Conceptualization, Methodology, Investigation, Resources, Writing – review & editing; JZ: Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing – review & editing; SA: Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing – review & editing; DJL: Conceptualization, Methodology, Investigation, Resources, Writing – review & editing; OTI: Conceptualization, Methodology, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition; RK: Conceptualization, Methodology, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

CONFLICT OF INTEREST STATEMENT

OTI is a cofounder and board member for Cardiosense, Inc.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Inan OT, Migeotte P-F, Park K-S, *et al.* Ballistocardiography and seismocardiography: a review of recent advances. *IEEE J Biomed Health Inform* 2015; 19 (4): 1414–27.
- Sørensen K, Schmidt SE, Jensen AS, Søgaard P, Struijk JJ. Definition of fiducial points in the normal seismocardiogram. *Sci Rep* 2018; 8 (1): 1–11.
- Inan OT, Baran Pouyan M, Javaid AQ, *et al.* Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients. *Circ Heart Fail* 2018; 11 (1): e004313.
- Gurel NZ, Jung H, Hersek S, Inan OT. Fusing near-infrared spectroscopy with wearable hemodynamic measurements improves classification of mental stress. *IEEE Sens J* 2019; 19 (19): 8522–31.
- Gurel NZ, Gazi AH, Scott KL, *et al.* Timing considerations for noninvasive vagal nerve stimulation in clinical studies. *AMIA Annu Symp Proc* 2020; 2019: 1061–70.
- Gazi AH, Sundararaj S, Harrison AB, *et al.* Transcutaneous cervical vagus nerve stimulation inhibits the reciprocal of the pulse transit time's responses to traumatic stress in posttraumatic stress disorder. In: 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Mexico; 2021: 1444–7. doi: [10.1109/EMBC46164.2021.9630415](https://doi.org/10.1109/EMBC46164.2021.9630415).
- Ganti VG, Gazi AH, An S, *et al.* Wearable seismocardiography-based assessment of stroke volume in congenital heart disease. *J Am Heart Assoc* 2022; 11 (18): e026067.
- Ganti V, Carek AM, Jung H, *et al.* Enabling wearable pulse transit time-based blood pressure estimation for medically underserved areas and health equity: comprehensive evaluation study. *JMIR Mhealth Uhealth* 2021; 9 (8): e27466.
- Jaipuria N, Zhang X, Bhasin R, *et al.* Deflating dataset bias using synthetic data augmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; Seattle, WA; 2020: 3344–53. doi: [10.1109/CVPRW50498.2020.00394](https://doi.org/10.1109/CVPRW50498.2020.00394).
- An S, Medda A, Sawka MN, *et al.* AdaptNet: human activity recognition via bilateral domain adaptation using semi-supervised deep translation networks. *IEEE Sensors J* 2021; 21 (18): 20398–411.
- Zheng Z, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In: Proceedings of the IEEE International Conference on Computer Vision; Venice, Italy; 2017: 3774–82. doi: [10.1109/ICCV.2017.405](https://doi.org/10.1109/ICCV.2017.405).
- Luo Y, Zhu L-Z, Wan Z-Y, Lu B-L. Data augmentation for enhancing EEG-based emotion recognition with deep generative models. *J Neural Eng* 2020; 17 (5): 056021.
- Zhu F, Ye F, Fu Y, Liu Q, Shen B. Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. *Sci Rep* 2019; 9 (1): 1–11.
- Delaney AM, Brophy E, Ward TE. Synthesis of realistic ECG using generative adversarial networks. arXiv, arXiv:1909.09150, 2019, preprint: not peer reviewed.
- Hazra D, Byun Y-C. Synsiggan: generative adversarial networks for synthetic biomedical signal generation. *Biology* 2020; 9 (12): 441.
- Kiyasseh D, Tadesse GA, Thwaites L, *et al.* Plethaugment: GAN-based PPG augmentation for medical diagnosis in low-resource settings. *IEEE J Biomed Health Inform* 2020; 24 (11): 3226–35.
- Zia J, Kimball J, Hersek S, Inan OT. Modeling consistent dynamics of cardiogenic vibrations in low-dimensional subspace. *IEEE J Biomed Health Inform* 2020; 24 (7): 1887–98.
- Hersek S, Semiz B, Shandhi MMH, Orlandic L, Inan OT. A globalized model for mapping wearable seismocardiogram signals to whole-body ballistocardiogram signals based on deep learning. *IEEE J Biomed Health Inform* 2020; 24 (5): 1296–309.
- Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Adv Neural Inform Process Syst* 2017; 30: 6000–10.
- Zimetbaum PJ, Josephson ME. Use of the electrocardiogram in acute myocardial infarction. *N Engl J Med* 2003; 348 (10): 933–40.
- Allen J. Photoplethysmography and its application in clinical physiological measurement. *Physiol Meas* 2007; 28 (3): R1–39.
- Shandhi MMH, Semiz B, Hersek S, Goller N, Ayazi F, Inan OT. Performance analysis of gyroscope and accelerometer sensors for seismocardiography-based wearable pre-ejection period estimation. *IEEE J Biomed Health Inform* 2019; 23 (6): 2365–74.
- Ashouri H, Inan OT. Automatic detection of seismocardiogram sensor misplacement for robust pre-ejection period estimation in unsupervised settings. *IEEE Sensors J* 2017; 17 (12): 3805–13.
- Chan M, Ganti VG, Heller JA, Abdallah CA, Etemadi M, Inan OT. Enabling continuous wearable reflectance pulse oximetry at the sternum. *Bio-sensors* 2021; 11 (12): 521.
- Zia J, Kimball J, Rolfs C, Hahn J-O, Inan OT. Enabling the assessment of trauma-induced hemorrhage via smart wearable systems. *Sci Adv* 2020; 6 (30): eabb1708.
- Zia J, Kimball J, Hersek S, Shandhi MMH, Semiz B, Inan OT. A unified framework for quality indexing and classification of seismocardiogram signals. *IEEE J Biomed Health Inform* 2020; 24 (4): 1080–92.
- Karita S, Chen N, Hayashi T, *et al.* A comparative study on transformer vs RNN in speech applications. In: 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU); Singapore; 2019: 449–56. doi: [10.1109/ASRU46091.2019.9003750](https://doi.org/10.1109/ASRU46091.2019.9003750).
- Li N, Liu S, Liu Y, Zhao S, Liu M. Neural speech synthesis with transformer network. *AAAI* 2019; 33 (1): 6706–13.
- Taeabi A, Solar BE, Bomar AJ, Sandler RH, Mansy HA. Recent advances in seismocardiography. *Vibration* 2019; 2 (1): 64–86.
- Liang Y, Ge C, Tong Z, Song Y, Wang J, Xie P. Not all patches are what you need: expediting vision transformers via token reorganizations. arXiv, arXiv:220207800, 2022, preprint: not peer reviewed.
- Hartmann KG, Schirrmeyer RT, Ball T. EEG-GAN: generative adversarial networks for electroencephalographic (EEG) brain signals. arXiv, arXiv:1806.01875, 2018, preprint: not peer reviewed.
- Weng L. From GAN to WGAN. arXiv, arXiv:1904.08994, 2019, preprint: not peer reviewed.
- Ashouri H, Hersek S, Inan OT. Universal pre-ejection period estimation using seismocardiography: quantifying the effects of sensor placement and regression algorithms. *IEEE Sens J* 2018; 18 (4): 1665–74.
- Nikbakht M, Lin DJ, Gazi AH, Inan OT. A synthetic seismocardiogram and electrocardiogram generator phantom. In: 2022 IEEE Sensors; Dallas, TX; 2022: 1–4. doi: [10.1109/SENSOR52175.2022.9967101](https://doi.org/10.1109/SENSOR52175.2022.9967101).
- Liu D, Smaragdis P, Kim M. Experiments on deep learning for speech denoising. In: Fifteenth Annual Conference of the International Speech Communication Association; 2014. <http://www.scopus.com/inward/record.url?scp=84910049527&partnerID=8YFLogxK>.
- Rethage D, Pons J, Serra X. A wavenet for speech denoising. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Calgary, AB, Canada; 2018: 5069–73. doi: [10.1109/ICASSP.2018.8462417](https://doi.org/10.1109/ICASSP.2018.8462417).
- Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci USA* 2019; 116 (44): 22071–80.
- Lin DJ, Kimball JP, Zia J, Ganti VG, Inan OT. Reducing the impact of external vibrations on fiducial point detection in seismocardiogram signals. *IEEE Trans Biomed Eng* 2022; 69 (1): 176–85.