Appendix for PulseImpute

Contents

36th Conference on Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks.

A1 mHealth versus Clinical Pulsative Waveforms

In our proposed challenge, we draw from clinical pulsative waveform datasets to mimic mHealth pulsative waveforms, and in this section, we provide additional justification for this approach.

A1.1 What is the rationale for constructing a dataset for mHealth signal imputation from equivalent signals connected in the clinical setting?

While there are differences between clinical pulsative signals collected in a hospital setting and mHealth pulsative signals collected in the field, this was a necessary approach due to the scarcity of large, publicly-available mHealth datasets (e.g. PPG-DaLiA, an mHealth dataset, has 15 subjects whereas our curated MIMIC-III PPG dataset, derived from a clinical dataset, has 18,210 subjects).

We can mimic real-world mHealth settings by applying realistic patterns of mHealth missingness. The original ablated samples are the ground truth, which makes it possible to quantify and visualize the imputation accuracy.

A1.2 What are the differences in how the ECG/PPG sensors collect pulsative signals across both settings?

An ECG signal is a recording of the electrical activity of the heart. The electrical activity is measured along the axis connecting two electrodes, and an ECG signal corresponding to a specific axis is referred to as an ECG lead. There are many specific ECG leads that are well-established within the medical field (e.g. Lead I, Lead II, Lead aVR, etc.), and each ECG lead measures the heart's activity along a specific direction.

In a clinical hospital setting, the patients are stationary, and therefore, it is simple to attach many electrodes onto the patient for diagnosis or monitoring purposes, allowing for multiple ECG leads to be recorded at once. However, in an mHealth field setting, ECG signals are recorded using wearables, such as a smart watch [1] or a band [5], on a user who may be constantly moving. Therefore, to prevent creating an unacceptable burden, single-lead recording is typically the only acceptable approach. This difference in total-leads-used is why during curation, we treat each lead as a separate waveform, and propose a univariate imputation problem rather than a multivariate one.

Previous work [6, 9] has demonstrated that mHealth ECG sensors are able to record clinically-accurate ECG signals, very similar to those collected in a hospital setting, in both healthy subjects and subjects with underlying cardiac disease.

A PPG signal measures blood volume changes to assess how the heart pumps blood to the periphery and typically does so with a pulse oximeter sensor, which works by measuring the changes in light absorption on the skin.

In clinical hospital settings, the pulse oximeter device is clipped to a stationary patient's finger, so the signal is stable with a high data quality [13]. In mHealth, PPG signals are typically collected on a watch, so there will be more noise and missingness resulting from movement [13].

While there are some differences between wrist mHealth PPG and finger clinical PPG (namely in signal shape structure) [15], both types of PPG signals are used to model the same health paradigms. Both of them can been used to model the same morphological-based phenomena such as Pulse Arrival Time [15] and the same rhythm-based phenomena such as Heart Rate Variability [13]. PPG signals collected in the mHealth setting may be adapted to be used for clinical marker calculations originally designed for clinical PPG signals [15, 13]. This suggests that domain gap issues between clinical and mHealth settings, while they exist, may not be not a major obstacle.

A1.3 How do the populations differ in these two settings?

Generally, patients in a clinical setting are in a worse health condition than users in a mHealth setting. In the hospital, patients may be in the ICU with ECG/PPG sensors to monitor their already-poor health condition. Conversely, mHealth technology has a young consumer base and is generally used by individuals for maintenance of healthy behaviors [20]. Therefore, clinical signals will be more variable than those originating from mHealth devices, due to the diverse set of cardiopulmonary diseases that may be afflicting the hospital patients. However, this is not a limitation for our challenge design, as this allows us to present a more challenging and interesting task for the ML community to tackle. ML methods must rely on learning to impute missing signals based on the signal that is present, rather than learning to create a general-purpose imputation template that mimics standard healthy behavior.

A1.4 How does clinical missingness differ from mHealth missingness?

There are similarities in missingness patterns across the clinical and mHealth domains. For example, with respect to participant compliance, both clinical patients and mHealth users can remove sensors, resulting in blocks of missing data. Likewise, participant movement in both contexts can result in artifacts (e.g. tugging at an attached sensor in the hospital vs adjusting an uncomfortable strap of a mHealth wearable). At the same time, the mHealth environment is more challenging for data capture and may experience more missingness overall.

However, we would like to clarify that comparisons between clinical and mHealth missingness do not affect our findings and experimental design because:

1. We do not make any claims about the suitability of our approach for addressing the issue of clinical missingness.
2. There is no clinical missingness present in our benchmark dataset.

Our contribution is on introducing a benchmarking suite for pulsative signals with realistic mHealth missingness, and our data curation process (described in Sections 3.1, 3.2, A2.1, A2.2) ensured that signals with clinical missingness were removed from the dataset.

A2 Curation of MIMIC-III Waveform and Heartbeat Detection Task Details

For each of these curations, we intentionally utilize an aggressive filtration method to ensure that we have clean signals. The sheer volume of the MIMIC-III Waveforms dataset allows us to filter out many unsuitable signals and still curate the largest ECG/PPG waveform dataset.

Curating a clean version of MIMIC-III Waveform is both critical for our imputation challenge design and is very advantageous for the broader biosensor ML field. For PulseImpute, we need clean signals for training imputation models to reconstruct the signal structure and not noise. In a broader context, we want to match the high quality level of other datasets such as PTB-XL, in which 77.01% of the signal data are of highest assessed quality [18]. This matching enables researchers to combine datasets in the future or to train transfer learning approaches with our curated datasets, potentially leveraging self-supervised representation learning.

A2.1 MIMIC-III ECG Curation

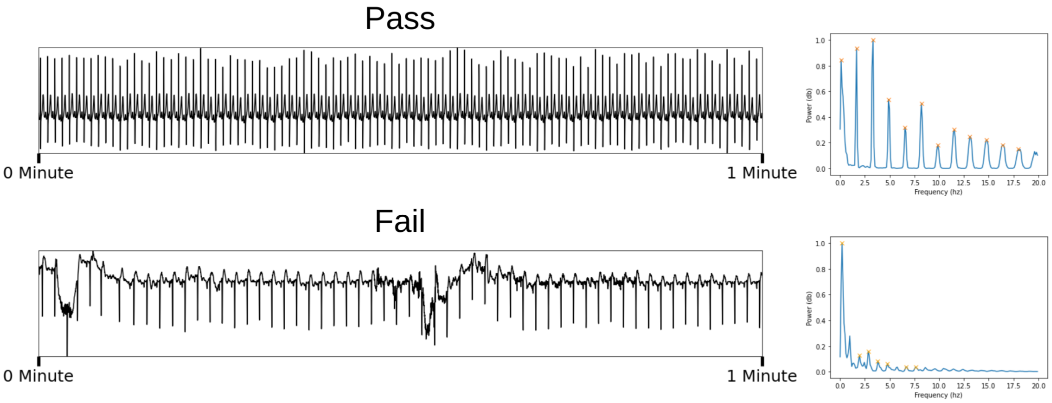
MIMIC-III Waveform [12] has 4,799,017 ECG signal files, which we curate down to 440,953 clean ECG signal files. Below is the MIMIC-III ECG curation procedure we utilized, and please see our code for specific implementation details.

1. For a given ECG Signal, resample the waveform from 125 Hz to 100 Hz and conduct linear interpolation to fill in NA values.

2. Utilize Welch's method [19] to obtain the periodogram and conduct peak detection on the periodogram with a strict minimum peak distance requirement. In a clean ECG signal, regularly spaced peaks in the periodogram correspond to the harmonics of a QRS complex, especially those in the upper frequency bands [2].

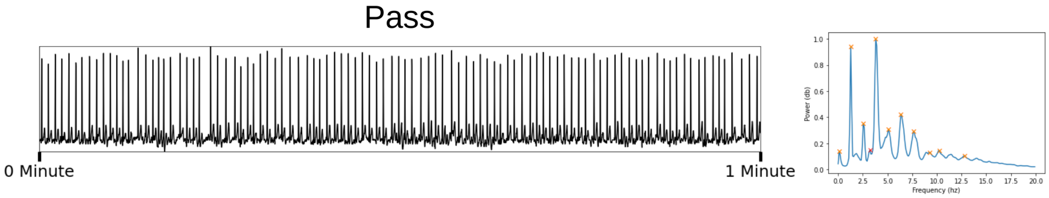
   Therefore, if the detected peaks are regularly spaced and there is a peak detected at > 10 Hz, then the ECG signal is marked as clean, and we skip to step 4. Else If the peaks are not too irregularly spaced, then we move to step 3 for another chance for the signal to pass the quality check. Else, the peaks are too irregularly spaced, and we reject the signal.

   See below for examples of ECG signals with their associated periodogram. The top demonstrates a clean ECG signal with regularly spaced peaks in its periodogram and detected peaks past 10 Hz. The bottom represents a noisy ECG signal with a periodogram with irregularly spaced peaks and no significant peaks detected past 10 Hz.



3. A new peak detection on the periodogram is conducted with a relaxed minimum peak distance requirement, and new peaks are compared to old peaks. These new peaks are designated by the red x's in the below periodogram, and the old peaks are designated by the orange x's.

   If number of new peaks is not too high or if the new peaks are far away from the old peaks, then the signal is marked clean, and we move to step 4. Else, we reject the signal. This allows for non-normal heart rhythms in which the heart rate fluctuates to pass. The below example demonstrates a signal with heart rate irregularities, but still passes our filter.



4. For all signals marked clean, we sample a 5-minute segment, run an ECG peak detection algorithm, and if HR is within an acceptable physiological range, then the *ECG signal is accepted.*

A2.2 MIMIC-III PPG Curation

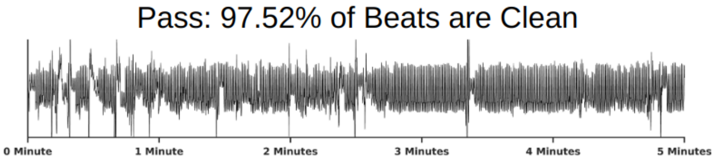
MIMIC-III Waveform [12] has 3,162,804 PPG signal files, which we curate down to 151,738 clean PPG signal files. Below is the MIMIC-III PPG curation procedure we utilized, and please see our code for specific implementation details.

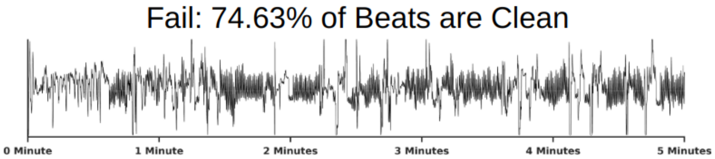
1. For a given PPG Signal, select a 5-minute segment and resample the waveform from 125 Hz to 100 Hz.

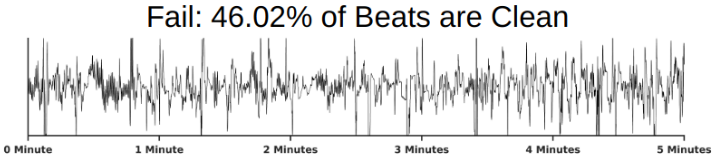
2. Segment the waveform into each beat with a peak detection algorithm and extract the PPG beat template with an ensemble averaging-based approach [17]. If a template is failed to be found, then reject the signal.

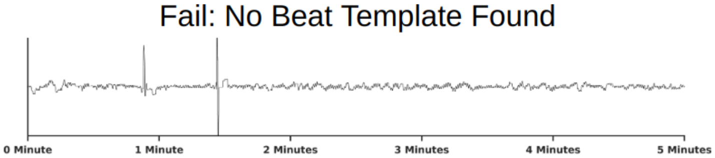
3. Calculate the DTW-based quality metric (bounded between 0 and 1) for each beat [17]. This is done by using DTW to align the template with the beat and calculating the correlation coefficient. If the correlation is negative, the similarity is clamped to zero.

4. If 95% of the beats have a quality greater than 0.5, the *PPG signal is accepted*. Else, reject the signal. See below for examples of accepted and rejected PPG signals.









A2.3 ECG/PPG mHealth Missingness Extraction

To generate ECG mHealth Missingness patterns, the Autosense [5] device in our mHealth field study [3] used an ECG data quality assessment algorithm [11] to detect noise and missingness.

However, this Autosense device does not record PPG signals, and thus we do not have access to PPG mHealth missingness patterns. We cannot use ECG mHealth missingness patterns to model PPG mHealth missingness because PPG signals may have different missingness patterns due to the differing types of sensor attachment. ECG signals can be collected on a chest band, as is done in Autosense, whereas PPG signals are typically collected with a wrist-mounted smartwatch.

Therefore, we seek to extract missingness patterns from the public mHealth PPG dataset, PPG-DaLiA [16], with the procedure outlined below.

1. For a given PPG Signal in PPG-DaLiA, resample the waveform from 64 Hz to 100 Hz.
2. Segment the waveform into individual beats with PPG-DaLiA's provided ground-truth peaks and extract PPG beat template with ensemble averaging [17].
3. Calculate DTW-based quality metric (bounded between 0 and 1), as described in A2.2.
4. Create a binary time-series by marking segments with DTW-based quality metric $< .5$ as missing (0) and segments with the metric $\geq .5$ as not-missing (1).
5. Split the binary time-series into 5 minute segments to serve as a PPG mHealth missingness pattern.

A2.4 Heartbeat Detection via Peak Detection in ECG/PPG

Peak detection is essential for segmenting and localizing individual heart beats, which is a core capability that supports a variety of widely-used mHealth markers such as heart rate and heart rate variability, and we use ECG/PPG Heartbeat detection on the curated MIMIC-III waveform datasets as a downstream task within our challenge to evaluate imputation.

Below are the formulations of each of the metrics that we use in this task.

$$\text{Sens} = \frac{TP}{TP + FN} \quad \text{Prec} = \frac{TP}{TP + FP} \quad \text{F1} = \frac{2 * \text{Prec} * \text{Sens}}{\text{Prec} + \text{Sens}}$$

Given a peak that was identified from the imputation, we center a 50 ms window around this peak, as done by [14]. If there was a peak originally in this window before being ablated for imputation, then this is a *True Positive*. If there was no peak originally in this window, this is a *False Positive*. *False Negatives* are peaks that were in the original signal that were ablated but were not detected in the reconstructed imputation with this procedure.

The peak detection procedures were Stationary Wavelet Transform peak detector from [8] for ECG signals and a neighbor comparison with threshold and peak prominence filters for PPG signals.

A3 mHealth Missingness Visualizations for ECG and PPG

A3.1 ECG Extracted mHealth Missingness

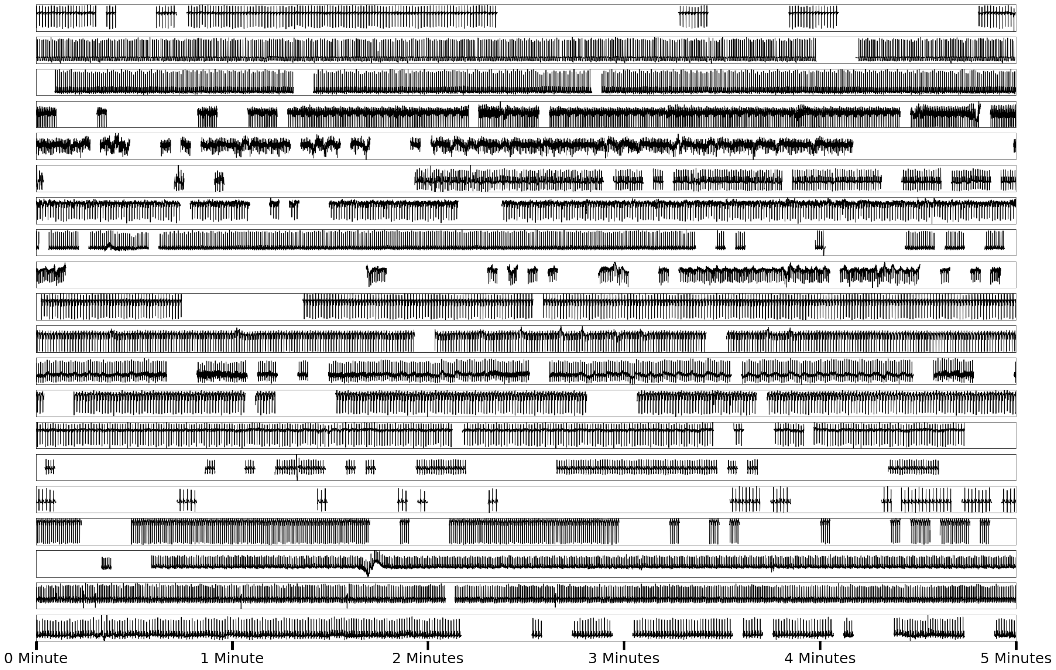


Figure A1: Various extracted ECG mHealth signal missingness patterns (shown by the gaps between the black signal) applied on different ECG waveforms. These are examples of the inputs used in the ECG Imputation and Heartbeat Detection Task. ECG missingness patterns are very complex in terms of their frequency and their duration. The ECG signals visualized here also are heterogeneous with many different morphologies (e.g. some signals have large peaks while others have large valleys) and different rhythms (e.g. signals have varying density). Additionally, each signal may be any particular lead within a wide range of possible leads.

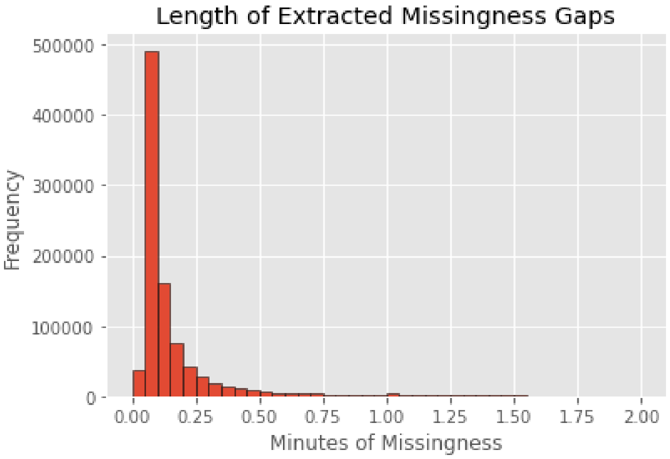


Figure A2: Histogram of Missingness Gap Length found in Extracted ECG mHealth Missingness Patterns. The missingness gaps' lengths have a wide range: the majority of missingness gaps are 3-9 seconds long but some gaps can last more than a minute.

A3.2 PPG Extracted mHealth Missingness

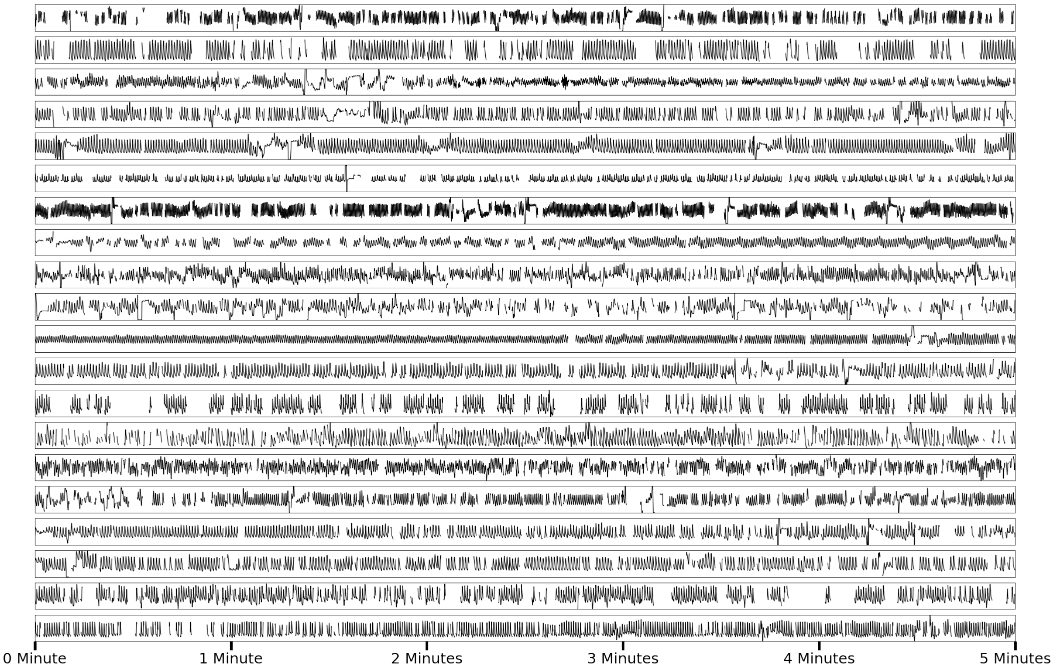


Figure A3: Various extracted PPG mHealth signal missingness patterns (shown by the gaps between the black signal) applied on different PPG waveforms. These are examples of the inputs used in the PPG Imputation and Heartbeat Detection Task. The PPG missingness patterns are different from those found in ECG, with much shorter gaps comparatively. The PPG signals are generally of simpler shapes, and there is more noise found in these PPG signals compared to the ECG signals.

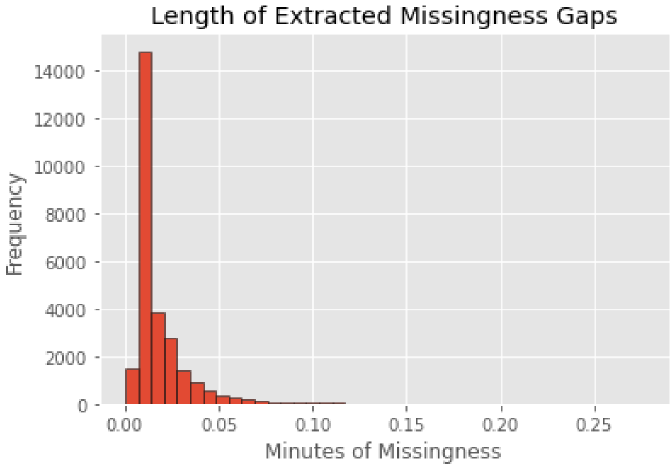


Figure A4: Histogram of Missingness Gap Length found in Extracted PPG mHealth Missingness Patterns. As visually seen while comparing Figure A1 and A3, the missingness gaps in PPG mHealth signals are shorter than those found in ECG.

A4 Experimental Setup Details

Our PulseImpute repo (`www.github.com/rehg-lab/pulseimpute`) contains the code needed to reproduce results, including a script to download the data and model checkpoints. Models were trained on Titan Xp GPUs for 24 hours or until convergence, whichever came first, on an internal Georgia Tech GPU Cluster. For each model trained on the 10-second-long ECG data used in the extended loss scenario for the ECG Imputation and Cardiac Classification Task, their model weights were used to initialize the model for the 5-minute-long ECG Imputation and Heartbeat Detection task before being further fine-tuned.

BRITS and NAOMI + BRITS w/ GAIL were implemented with their original papers' code bases found `www.github.com/caow13/BRITS` and `www.github.com/felixykliu/NAOMI`, respectively. The training procedures were set up to be identical to the original, with the only modification being how missingness was simulated during training. Rather than their default missingness procedure of dropping out individual time-points independently and at random, they were trained on task-specific missingness patterns, as described in Sections 3.1, 3.2, 3.3.

For the transformer models, the longformer's dilated sliding window attention was used for the 5-minute-long data in the ECG Imputation and Heartbeat Detection task. Conv9 uses the maximum kernel size for conv self-attention in its prior work [10], and our BDC module's query/key transformations have receptive fields of 883 ($\sim$9 sec). Each of the transformer-based architectures used follow the architecture scheme of one 1D Convolution Layer for embedding, two Transformer Encoder Layers, followed by one 1D Convolution Layer for projection for imputation. The transformer models were trained with a Masked Predictive Coding procedure, introduced in [7], inspired from the original Masked Language Prediction procedure, introduced in [4]. Given a block in which missingness would like to be ablated for training, there is a 80% probability that it is replaced with a 0 vector, 10% probability that sinusoidal vector is added as noise, and 10% probability that the block is kept the same. L2 Loss is then calculated between the imputed result and ground-truth.

Please see our code repo for further details on hyperparameters, experimental set-up, and reproducibility.

A5 Extra Results and Visualizations

In this section, we show extra visualizations of the performance of each of the imputation models, grouped by their downstream task: ECG Imputation and Heartbeat Detection, PPG Imputation and Heartbeat Detection, and ECG Imputation and Cardiac Pathophysiology Classification.

- ECG/PPG Imputation and Heartbeat Detection Tasks benchmark imputation by applying extracted mHealth missingness patterns on 5-minute-long ECG/PPG data.
- ECG Imputation and Cardiac Classification benchmarks imputation by systematically varying amount of missingness with the extended and transient loss missingness models on 10-second-long ECG data.

The purpose of this section is to visually evaluate the reconstruction quality of each of the models, as well as understanding the variation of the imputation model performance with MSE density plots.

A5.1 ECG Imputation and Heartbeat Detection

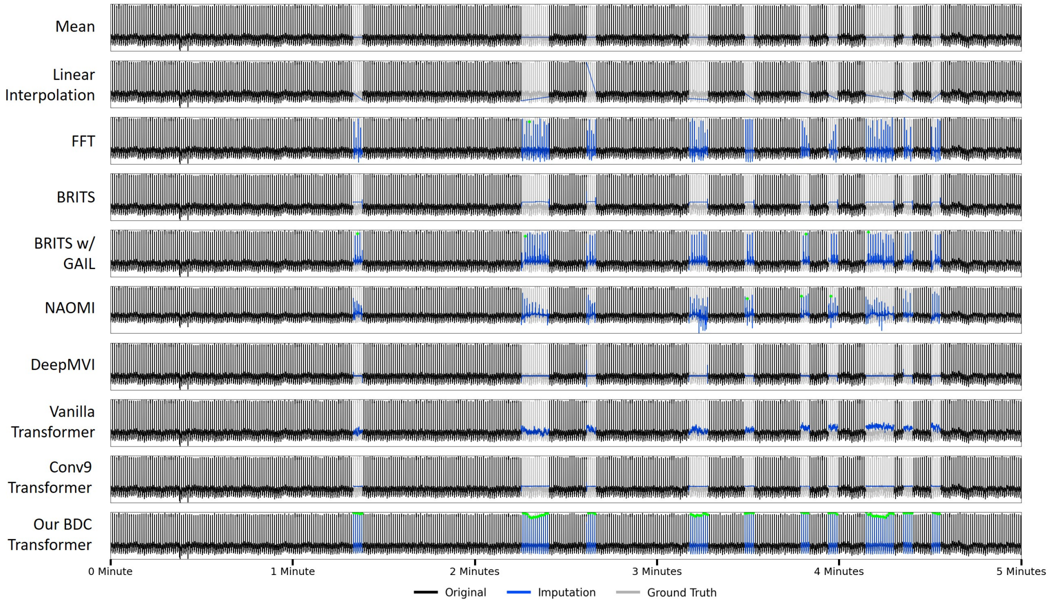


Figure A5: Extra visualization #1 of 5 minutes of imputation results from ECG Heartbeat Detection. The green dots designate True Positive reconstructed heartbeat peak detection. Given a signal with shorter missingness gaps, our BDC transformer is able to reconstruct the signal and rhythm very well, shown by the large amount of green dots.

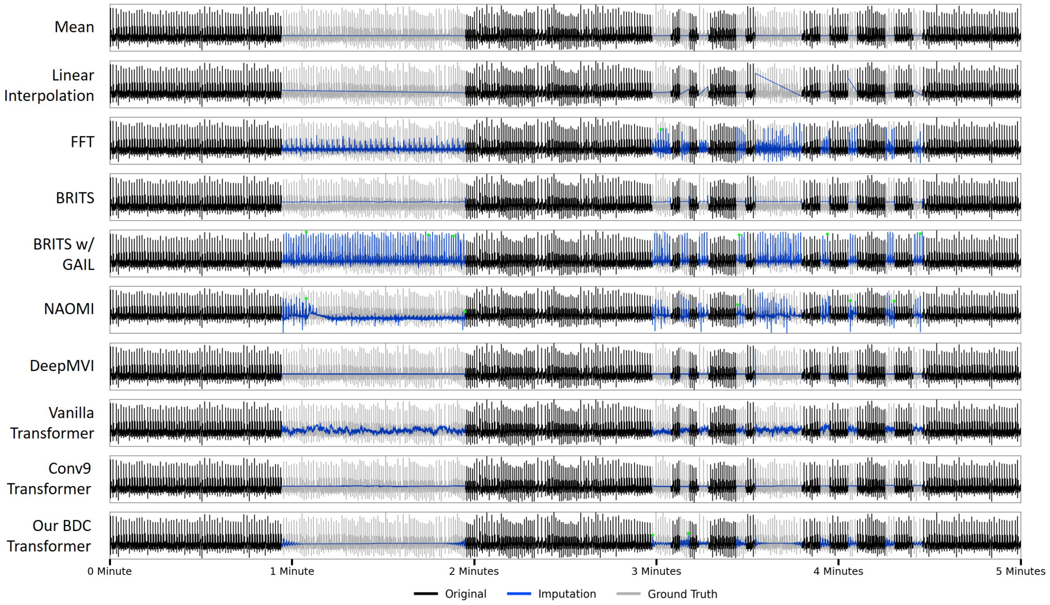


Figure A6: Extra visualization #2 of 5 minutes of imputation results from ECG Heartbeat Detection. The green dots designate True Positive reconstructed heartbeat peak detection. Here we see none of the models are able to perform well with long gaps of missingness. For example, BRITS w/ GAIL may seem to do well, but the small amount of green dots signifies that the rhythm of the original signal was unable to be recovered.

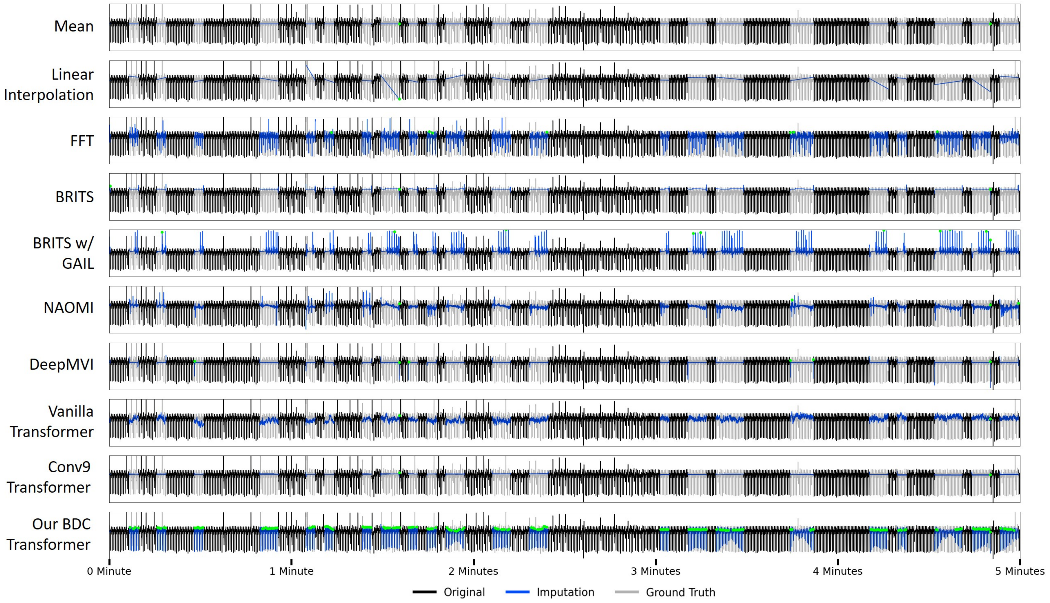


Figure A7: Extra visualization #3 of 5 minutes of imputation results from ECG Heartbeat Detection. The green dots designate True Positive reconstructed heartbeat peak detection. Even when the number of missingness gaps increase, because the gap length is short, our BDC transformer is able to perform well, signified by the high amount of green dots.

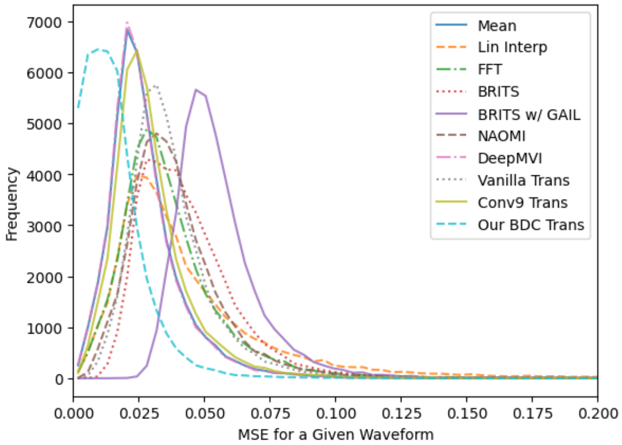


Figure A8: Density of MSE values across all ECG waveforms, for each imputation model. This demonstrates the variation of performance across all imputation methods, which shows that all existing imputation models have poor performance, with many models unable to achieve better MSE distributions than mean and linear interpolation. However, our BDC model consistently has the lowest MSE.

A5.2 PPG Imputation and Heartbeat Detection

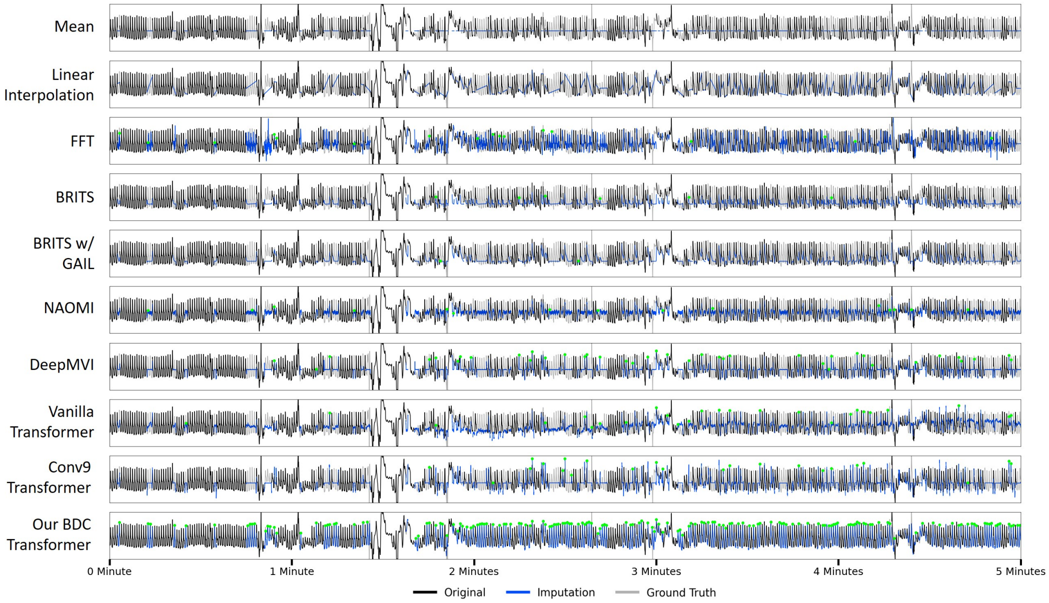


Figure A9: Extra visualization #1 of 5 minutes of imputation results from PPG Heartbeat Detection. The green dots designate True Positive reconstructed heartbeat peak detection. We can see that imputation models perform better in this PPG setting, with the simpler morphologies and shorter gaps of missingness. Across all ML models, there are more green dots, and thus more correct peak reconstructions. In general, our BDC transformer has very strong performance, able to reconstruct the signal nearly perfectly.

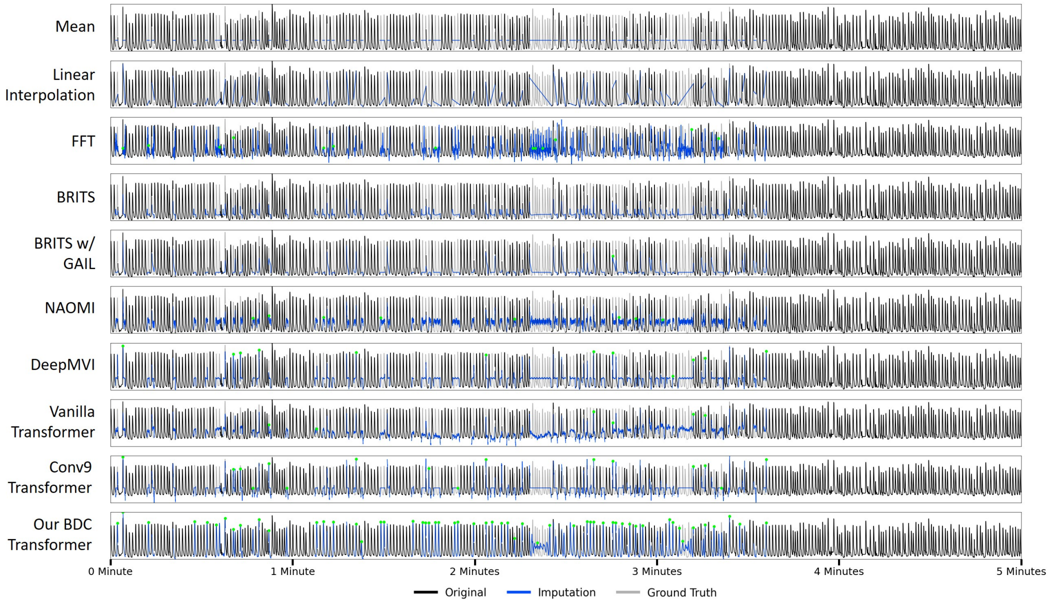


Figure A10: Extra visualization #2 of 5 minutes of imputation results from PPG Heartbeat Detection. The green dots designate True Positive reconstructed peak detection. This further shows how all models perform better within this PPG setting, and the stronger imputation performance of our BDC transformer compared to other models.

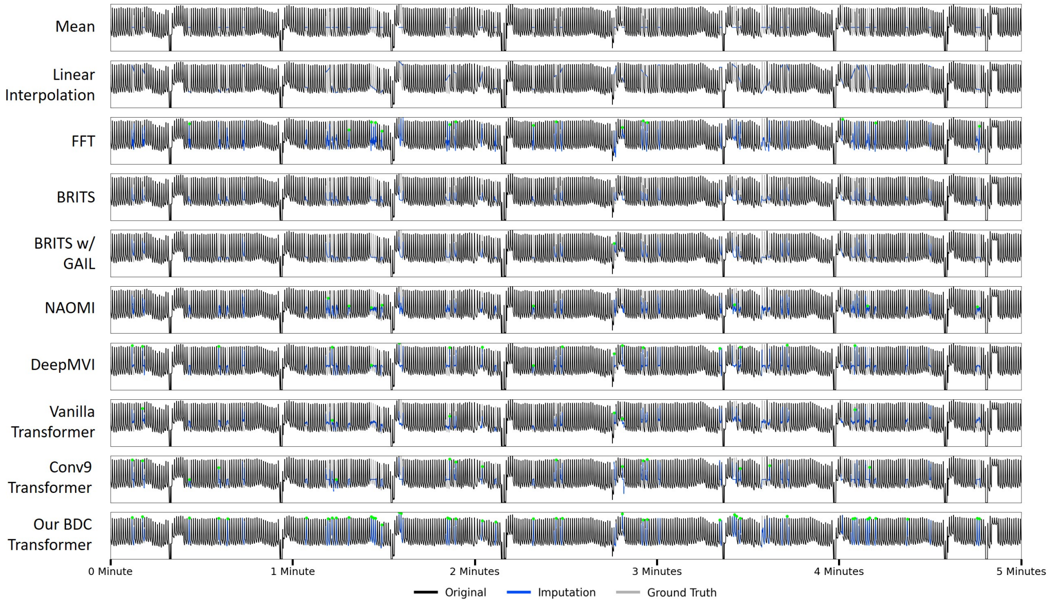


Figure A11: Extra visualization #3 of 5 minutes of imputation results from PPG Heartbeat Detection. The green dots designate True Positive reconstructed peak detection. This further shows how all models perform better within this PPG setting, and the stronger imputation performance of our BDC transformer compared to other models.

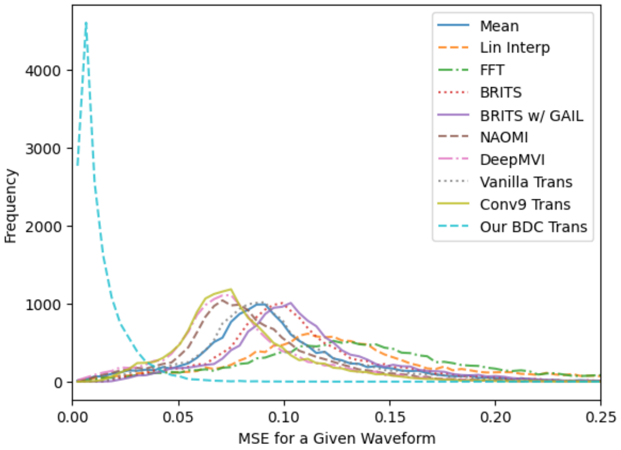


Figure A12: Density of MSE values across all PPG waveforms, for each imputation model. This demonstrates the variation of performance, which shows that compared to the ECG task, there are only a few models with better MSE distributions than the mean imputation model, namely Conv9 Transformer, DeepMVI, NAOMI, and our BDC transformer, with BDC consistently having the lowest MSE. In general, this PPG imputation task is easier for the ML models than the ECG imputation counterpart, likely due to the simpler morphologies and shorter missingness gaps present.

A5.3 ECG Imputation and Cardiac Pathophysiology Classification

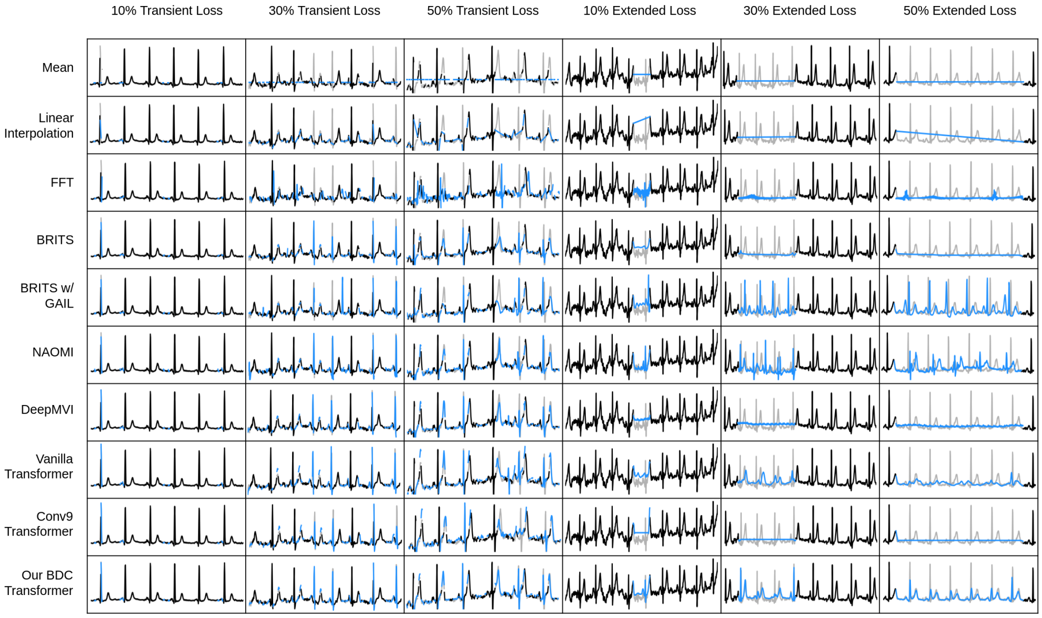


Figure A13: Extra visualization of imputation results from ECG Cardiac Classification, where each plot corresponds to a 6-second subsequence of the full 10-second signal. Grey designates the ground truth, blue the imputation results, and black the not-missing data. In the transient setting, all models perform well. However, in the extended setting, we see the GAN methods (e.g. NAOMI and BRITS w/ GAIL) will reconstruct signals that do not mimic the non-ablated data, and the other methods (e.g. BRITS and DeepMVI) have flat-line imputations. Our BDC transformer matches the rhythm of the ablated signal well, but struggles with reproducing the R peaks (the tallest steepest peaks in a quasiperiod).

		Transient				Extended			
% Miss	Models	MSE	Rhy AUC	Form AUC	Diag AUC	MSE	Rhy AUC	Form AUC	Diag AUC
0	-	0	.949	.796	.845	0	.949	.796	.845
10	Mean	.0302 ± .00044	.922 ± .0112	.776 ± .0186	.827 ± .0120	.0300 ± .00042	.886 ± .0111	.786 ± .0166	.827 ± .0199
	Lin Interp	.0221 ± .00033	.927 ± .0108	.788 ± .0187	.833 ± .0117	.0438 ± .00107	.890 ± .0114	.787 ± .0156	.824 ± .0190
	FFT	.0482 ± .00058	.923 ± .0125	.786 ± .0169	.832 ± .0102	.0368 ± .00047	.905 ± .0005	.778 ± .0154	.826 ± .0115
	BRITS	.0059 ± .00013	.946 ± .0104	.793 ± .0184	.846 ± .0082	.0285 ± .00041	.887 ± .0106	.787 ± .0194	.834 ± .0192
	BRITS w/ GAIL	.0110 ± .00024	.939 ± .0101	.793 ± .0181	.846 ± .0094	.0486 ± .00049	.904 ± .0108	.787 ± .0166	.831 ± .0136
	NAOMI	.0306 ± .00019	.946 ± .0103	.793 ± .0180	.847 ± .0080	.0341 ± .00043	.929 ± .0096	.782 ± .0205	.843 ± .0100
	DeepMVI	.0039 ± .00010	.947 ± .0104	.794 ± .0185	.847 ± .0080	.0285 ± .00038	.875 ± .0110	.783 ± .0197	.824 ± .0213
	Van Trans	.0065 ± .00014	.946 ± .0105	.791 ± .0183	.846 ± .0081	.0260 ± .00039	.895 ± .0105	.787 ± .0180	.831 ± .0164
	Conv9 Trans	.0049 ± .00012	.944 ± .0104	.791 ± .0185	.847 ± .0083	.0288 ± .00041	.887 ± .0109	.787 ± .0182	.830 ± .0203
	Our BDC Trans	.0030 ± .00007	.948 ± .0104	.795 ± .0185	.847 ± .0078	.0116 ± .00027	.944 ± .0188	.790 ± .0086	.844 ± .0086
20	Mean	.0302 ± .00038	.876 ± .0118	.762 ± .0194	.805 ± .0217	.0301 ± .00039	.882 ± .0206	.775 ± .0230	.811 ± .0230
	Lin Interp	.0239 ± .00028	.910 ± .0114	.782 ± .0196	.813 ± .0134	.0454 ± .00104	.886 ± .0203	.774 ± .0215	.812 ± .0215
	FFT	.0477 ± .00049	.890 ± .0147	.765 ± .0177	.808 ± .0108	.0357 ± .00045	.871 ± .0226	.758 ± .0173	.799 ± .0126
	BRITS	.0075 ± .00013	.944 ± .0104	.793 ± .0160	.844 ± .0076	.0296 ± .00040	.883 ± .0113	.776 ± .0206	.825 ± .0242
	BRITS w/ GAIL	.0137 ± .00023	.929 ± .0100	.787 ± .0179	.838 ± .0095	.0523 ± .00045	.898 ± .0109	.773 ± .0164	.816 ± .0145
	NAOMI	.0102 ± .00018	.941 ± .0099	.792 ± .0158	.846 ± .0088	.0388 ± .00039	.903 ± .0113	.771 ± .0189	.827 ± .0136
	DeepMVI	.0045 ± .00010	.945 ± .0107	.794 ± .0172	.845 ± .0076	.0288 ± .00035	.845 ± .0136	.771 ± .0212	.803 ± .0279
	Van Trans	.0071 ± .00014	.944 ± .0104	.790 ± .0164	.845 ± .0081	.0251 ± .00036	.895 ± .0122	.774 ± .0191	.808 ± .0189
	Conv9 Trans	.0060 ± .00012	.938 ± .0105	.789 ± .0181	.842 ± .0088	.0293 ± .00039	.882 ± .0120	.777 ± .0204	.819 ± .0232
	Our BDC Trans	.0033 ± .00007	.947 ± .0105	.795 ± .0160	.847 ± .0076	.0136 ± .00027	.935 ± .0119	.780 ± .0196	.829 ± .0093
30	Mean	.0302 ± .00036	.828 ± .0119	.735 ± .0212	.782 ± .0240	.0304 ± .00040	.883 ± .0129	.760 ± .0197	.807 ± .0204
	Lin Interp	.0260 ± .00025	.870 ± .0119	.766 ± .0181	.792 ± .0207	.0454 ± .00109	.879 ± .0123	.761 ± .0190	.801 ± .0204
	FFT	.0470 ± .00046	.842 ± .0236	.738 ± .0198	.780 ± .0114	.0344 ± .00043	.870 ± .0182	.746 ± .0161	.768 ± .0137
	BRITS	.0095 ± .00014	.933 ± .0099	.783 ± .0179	.838 ± .0179	.0299 ± .00039	.880 ± .0115	.766 ± .0202	.823 ± .0188
	BRITS w/ GAIL	.0173 ± .00023	.909 ± .0101	.777 ± .0192	.826 ± .0192	.0535 ± .00043	.890 ± .0107	.768 ± .0180	.815 ± .0133
	NAOMI	.0124 ± .00018	.932 ± .0099	.781 ± .0167	.840 ± .0167	.0405 ± .00038	.899 ± .0119	.751 ± .0229	.808 ± .0113
	DeepMVI	.0056 ± .00011	.939 ± .0108	.790 ± .0155	.841 ± .0155	.0290 ± .00036	.856 ± .0136	.751 ± .0210	.797 ± .0223
	Van Trans	.0084 ± .00014	.936 ± .0107	.786 ± .0152	.841 ± .0152	.0226 ± .00035	.903 ± .0117	.758 ± .0194	.796 ± .0143
	Conv9 Trans	.0078 ± .00013	.930 ± .0106	.783 ± .0168	.837 ± .0168	.0294 ± .00038	.885 ± .0119	.761 ± .0207	.814 ± .0190
	Our BDC Trans	.0038 ± .00007	.945 ± .0105	.793 ± .0177	.844 ± .0177	.0159 ± .00028	.930 ± .0121	.773 ± .0197	.817 ± .0094
40	Mean	.0302 ± .00035	.784 ± .0128	.707 ± .0240	.752 ± .0279	.0307 ± .00039	.871 ± .0114	.761 ± .0172	.812 ± .0181
	Lin Interp	.0282 ± .00026	.827 ± .0130	.745 ± .0180	.766 ± .0275	.0460 ± .00106	.870 ± .0110	.750 ± .0188	.804 ± .0154
	FFT	.0461 ± .00047	.776 ± .0298	.698 ± .0235	.740 ± .0143	.0332 ± .00043	.843 ± .0176	.736 ± .0183	.762 ± .0134
	BRITS	.0121 ± .00016	.923 ± .0109	.778 ± .0161	.832 ± .0098	.0301 ± .00038	.870 ± .0107	.762 ± .0206	.825 ± .0190
	BRITS w/ GAIL	.0216 ± .00024	.870 ± .0101	.764 ± .0177	.807 ± .0182	.0543 ± .00041	.874 ± .0113	.754 ± .0184	.809 ± .0147
	NAOMI	.0150 ± .00018	.922 ± .0099	.775 ± .0171	.833 ± .0110	.0410 ± .00036	.876 ± .0122	.749 ± .0160	.801 ± .0166
	DeepMVI	.0072 ± .00012	.925 ± .0110	.781 ± .0169	.833 ± .0102	.0291 ± .00034	.830 ± .0127	.741 ± .0248	.799 ± .0297
	Van Trans	.0105 ± .00016	.929 ± .0107	.774 ± .0164	.835 ± .0103	.0270 ± .00036	.860 ± .0120	.755 ± .0182	.790 ± .0219
	Conv9 Trans	.0106 ± .00016	.915 ± .0109	.768 ± .0190	.827 ± .0122	.0295 ± .00037	.870 ± .0115	.760 ± .0182	.815 ± .0175
	Our BDC Trans	.0048 ± .00009	.944 ± .0109	.790 ± .0183	.841 ± .0084	.0181 ± .00029	.912 ± .0132	.762 ± .0194	.801 ± .0113
50	Mean	.0302 ± .00034	.758 ± .0137	.677 ± .0210	.717 ± .0294	.0312 ± .00040	.858 ± .0119	.742 ± .0217	.806 ± .0225
	Lin Interp	.0306 ± .00028	.772 ± .0130	.726 ± .0174	.743 ± .0312	.0467 ± .00104	.846 ± .0113	.740 ± .0237	.800 ± .0251
	FFT	.0448 ± .00046	.721 ± .0296	.665 ± .0216	.706 ± .0138	.0325 ± .00038	.830 ± .0314	.727 ± .0219	.766 ± .0129
	BRITS	.0155 ± .00018	.895 ± .0115	.761 ± .0151	.817 ± .0142	.0302 ± .00036	.850 ± .0103	.744 ± .0212	.825 ± .0242
	BRITS w/ GAIL	.0275 ± .00027	.841 ± .0117	.737 ± .0196	.786 ± .0188	.0549 ± .00040	.860 ± .0120	.749 ± .0181	.802 ± .0212
	NAOMI	.0179 ± .00019	.913 ± .0095	.765 ± .0155	.829 ± .0101	.0411 ± .00037	.863 ± .0123	.716 ± .0201	.786 ± .0167
	DeepMVI	.0098 ± .00014	.904 ± .0116	.767 ± .0170	.820 ± .0136	.0294 ± .00034	.818 ± .0120	.736 ± .0222	.797 ± .0225
	Van Trans	.0138 ± .00020	.903 ± .0108	.756 ± .0178	.826 ± .0149	.0290 ± .00035	.827 ± .0133	.748 ± .0163	.779 ± .0261
	Conv9 Trans	.0146 ± .00020	.877 ± .0117	.756 ± .0187	.812 ± .0210	.0296 ± .00035	.856 ± .0111	.744 ± .0217	.812 ± .0233
	Our BDC Trans	.0066 ± .00011	.936 ± .0109	.784 ± .0157	.836 ± .0083	.0209 ± .00029	.892 ± .0126	.754 ± .0176	.794 ± .0142

Table A1: Full tabulated results from the ECG cardiac classification task with Macro AUC values with 95% CI

		Transient				Extended			
% Miss	**Models**	**MSE**	**Rhy AUC**	**Form AUC**	**Diag AUC**	**MSE**	**Rhy AUC**	**Form AUC**	**Diag AUC**
0	-	0	.930	.752	.794	0	.930	.752	.794
10	Mean	.0334 ± .00014	.904 ± .0051	.743 ± .0055	.783 ± .0038	.0333 ± .00015	.885 ± .0053	.733 ± .0054	.773 ± .0036
	Lin Interp	.0229 ± .00011	.919 ± .0045	.747 ± .0053	.786 ± .0038	.0460 ± .00031	.880 ± .0056	.730 ± .0053	.773 ± .0039
	FFT	.0487 ± .00017	.917 ± .0047	.744 ± .0054	.774 ± .0036	.0406 ± .00017	.888 ± .0055	.734 ± .0051	.783 ± .0037
	BRITS	.0036 ± .00003	.931 ± .0042	.749 ± .0052	.795 ± .0037	.0336 ± .00016	.887 ± .0053	.741 ± .0052	.777 ± .0038
	BRITS w/ GAIL	.0378 ± .00017	.887 ± .0056	.729 ± .0057	.764 ± .0037	.4249 ± .00278	.883 ± .0051	.714 ± .0058	.766 ± .0038
	NAOMI	.0264 ± .00015	.914 ± .0045	.743 ± .0054	.783 ± .0038	.0409 ± .00021	.889 ± .0052	.731 ± .0057	.776 ± .0038
	DeepMVI	1.1129 ± .00108	.609 ± .0085	.598 ± .0061	.647 ± .0043	.0309 ± .00012	.878 ± .0055	.740 ± .0054	.775 ± .0038
	Van Trans	.0015 ± .00002	.930 ± .0042	.750 ± .0051	.794 ± .0037	.0304 ± .00012	.875 ± .0055	.735 ± .0053	.764 ± .0038
	Conv9 Trans	.0018 ± .00002	.928 ± .0043	.750 ± .0051	.793 ± .0037	.0323 ± .00015	.886 ± .0051	.734 ± .0054	.776 ± .0037
	Our BDC Trans	.0015 ± .00001	.932 ± .0040	.749 ± .0054	.794 ± .0037	.0132 ± .00010	.867 ± .0061	.713 ± .0053	.751 ± .0042
20	Mean	.0334 ± .00012	.878 ± .0055	.721 ± .0055	.756 ± .0037	.0336 ± .00014	.867 ± .0061	.713 ± .0053	.751 ± .0042
	Lin Interp	.0246 ± .00009	.896 ± .0057	.736 ± .0059	.774 ± .0037	.0477 ± .00031	.862 ± .0061	.713 ± .0051	.752 ± .0042
	FFT	.0490 ± .00015	.880 ± .0062	.724 ± .0056	.746 ± .0037	.0397 ± .00015	.870 ± .0060	.702 ± .0055	.752 ± .0037
	BRITS	.0042 ± .00003	.931 ± .0041	.748 ± .0050	.795 ± .0036	.0343 ± .00015	.871 ± .0055	.716 ± .0051	.759 ± .0041
	BRITS w/ GAIL	.0378 ± .00015	.831 ± .0072	.689 ± .0058	.718 ± .0039	.5566 ± .00293	.863 ± .0055	.691 ± .0056	.738 ± .0039
	NAOMI	.0252 ± .00013	.897 ± .0051	.730 ± .0053	.762 ± .0038	.0493 ± .00025	.855 ± .0057	.690 ± .0056	.745 ± .0041
	DeepMVI	.0038 ± .00002	.929 ± .0039	.748 ± .0054	.794 ± .0037	.0316 ± .00012	.836 ± .0070	.712 ± .0053	.734 ± .0039
	Van Trans	.0019 ± .00002	.929 ± .0043	.747 ± .0053	.792 ± .0037	.0286 ± .00010	.848 ± .0068	.728 ± .0054	.743 ± .0043
	Conv9 Trans	.0026 ± .00002	.924 ± .0046	.747 ± .0052	.791 ± .0037	.0326 ± .00013	.870 ± .0057	.715 ± .0052	.752 ± .0041
	Our BDC Trans	.0017 ± .00001	.931 ± .0042	.747 ± .0054	.793 ± .0036	.0153 ± .00009	.919 ± .0037	.735 ± .0050	.773 ± .0039
30	Mean	.0335 ± .00011	.847 ± .0060	.686 ± .0059	.723 ± .0040	.0340 ± .00014	.864 ± .0054	.691 ± .0056	.737 ± .0044
	Lin Interp	.0266 ± .00008	.869 ± .0051	.691 ± .0054	.740 ± .0041	.0487 ± .00032	.868 ± .0067	.721 ± .0060	.756 ± .0037
	FFT	.0489 ± .00014	.840 ± .0065	.694 ± .0057	.712 ± .0039	.0384 ± .00015	.868 ± .0058	.682 ± .0055	.727 ± .0038
	BRITS	.0051 ± .00003	.933 ± .0035	.745 ± .0051	.792 ± .0036	.0346 ± .00014	.865 ± .0053	.697 ± .0054	.748 ± .0041
	BRITS w/ GAIL	.0379 ± .00014	.767 ± .0077	.646 ± .0057	.669 ± .0039	.6916 ± .00279	.849 ± .0054	.674 ± .0053	.724 ± .0040
	NAOMI	.0249 ± .00011	.875 ± .0053	.705 ± .0055	.741 ± .0039	.0513 ± .00024	.841 ± .0061	.666 ± .0059	.718 ± .0042
	DeepMVI	.0048 ± .00003	.925 ± .0038	.743 ± .0053	.790 ± .0036	.0320 ± .00012	.834 ± .0072	.698 ± .0054	.736 ± .0037
	Van Trans	.0027 ± .00002	.926 ± .0043	.745 ± .0054	.791 ± .0037	.0289 ± .00010	.852 ± .0064	.704 ± .0057	.735 ± .0041
	Conv9 Trans	.0038 ± .00003	.917 ± .0045	.743 ± .0055	.787 ± .0037	.0329 ± .00013	.863 ± .0056	.691 ± .0055	.741 ± .0043
	Our BDC Trans	.0021 ± .00002	.930 ± .0041	.745 ± .0053	.792 ± .0037	.0174 ± .00009	.917 ± .0038	.719 ± .0053	.760 ± .0039
40	Mean	.0335 ± .00011	.806 ± .0074	.646 ± .0060	.683 ± .0042	.0345 ± .00014	.851 ± .0054	.684 ± .0058	.736 ± .0041
	Lin Interp	.0287 ± .00008	.850 ± .0057	.684 ± .0058	.734 ± .0041	.0496 ± .00031	.835 ± .0075	.696 ± .0060	.728 ± .0038
	FFT	.0485 ± .00014	.796 ± .0068	.657 ± .0058	.668 ± .0042	.0372 ± .00014	.840 ± .0062	.668 ± .0057	.712 ± .0038
	BRITS	.0063 ± .00003	.928 ± .0036	.738 ± .0053	.787 ± .0036	.0349 ± .00014	.852 ± .0054	.693 ± .0056	.742 ± .0041
	BRITS w/ GAIL	.0380 ± .00014	.699 ± .0081	.600 ± .0057	.619 ± .0041	.8261 ± .00253	.836 ± .0061	.661 ± .0062	.711 ± .0040
	NAOMI	.0251 ± .00010	.842 ± .0068	.676 ± .0056	.716 ± .0040	.0514 ± .00022	.797 ± .0085	.653 ± .0060	.693 ± .0042
	DeepMVI	.0064 ± .00003	.914 ± .0043	.738 ± .0052	.784 ± .0036	.0325 ± .00012	.821 ± .0067	.691 ± .0056	.731 ± .0039
	Van Trans	.0041 ± .00002	.921 ± .0043	.743 ± .0050	.788 ± .0037	.0302 ± .00010	.839 ± .0062	.689 ± .0057	.719 ± .0045
	Conv9 Trans	.0058 ± .00003	.902 ± .0050	.738 ± .0052	.782 ± .0037	.0332 ± .00013	.851 ± .0055	.684 ± .0058	.738 ± .0041
	Our BDC Trans	.0030 ± .00002	.927 ± .0042	.742 ± .0052	.789 ± .0037	.0203 ± .00009	.907 ± .0039	.705 ± .0053	.746 ± .0040
50	Mean	.0335 ± .00011	.753 ± .0075	.611 ± .0059	.638 ± .0043	.0351 ± .00014	.847 ± .0061	.681 ± .0053	.728 ± .0042
	Lin Interp	.0311 ± .00008	.849 ± .0058	.680 ± .0054	.731 ± .0041	.0506 ± .00031	.790 ± .0076	.666 ± .0060	.698 ± .0039
	FFT	.0476 ± .00014	.734 ± .0076	.619 ± .0062	.623 ± .0043	.0364 ± .00013	.834 ± .0062	.662 ± .0054	.698 ± .0042
	BRITS	.0080 ± .00003	.915 ± .0040	.731 ± .0053	.779 ± .0035	.0350 ± .00013	.851 ± .0057	.692 ± .0053	.736 ± .0044
	BRITS w/ GAIL	.0380 ± .00014	.620 ± .0083	.563 ± .0060	.578 ± .0044	.9647 ± .00183	.825 ± .0059	.661 ± .0057	.711 ± .0040
	NAOMI	.0261 ± .00010	.809 ± .0068	.642 ± .0053	.691 ± .0038	.0519 ± .00020	.775 ± .0073	.638 ± .0057	.679 ± .0043
	DeepMVI	.0090 ± .00004	.891 ± .0052	.726 ± .0052	.772 ± .0036	.0329 ± .00012	.820 ± .0070	.686 ± .0051	.725 ± .0041
	Van Trans	.0065 ± .00003	.907 ± .0044	.733 ± .0052	.780 ± .0037	.0318 ± .00011	.830 ± .0067	.680 ± .0054	.705 ± .0044
	Conv9 Trans	.0088 ± .00004	.881 ± .0058	.727 ± .0056	.770 ± .0037	.0334 ± .00012	.847 ± .0061	.682 ± .0055	.729 ± .0042
	Our BDC Trans	.0054 ± .00003	.916 ± .0042	.736 ± .0051	.782 ± .0038	.0235 ± .00010	.888 ± .0049	.691 ± .0055	.728 ± .0040

Table A2: For the sake of completeness, these are the full tabulated results from a Union of Leads PTB-XL ECG dataset for the Cardiac Classification task with Macro AUC values with 95% CI. However, assume that all other results, besides this table, presented with ECG cardiac classification task are with the Lead I only PTB-XL dataset.

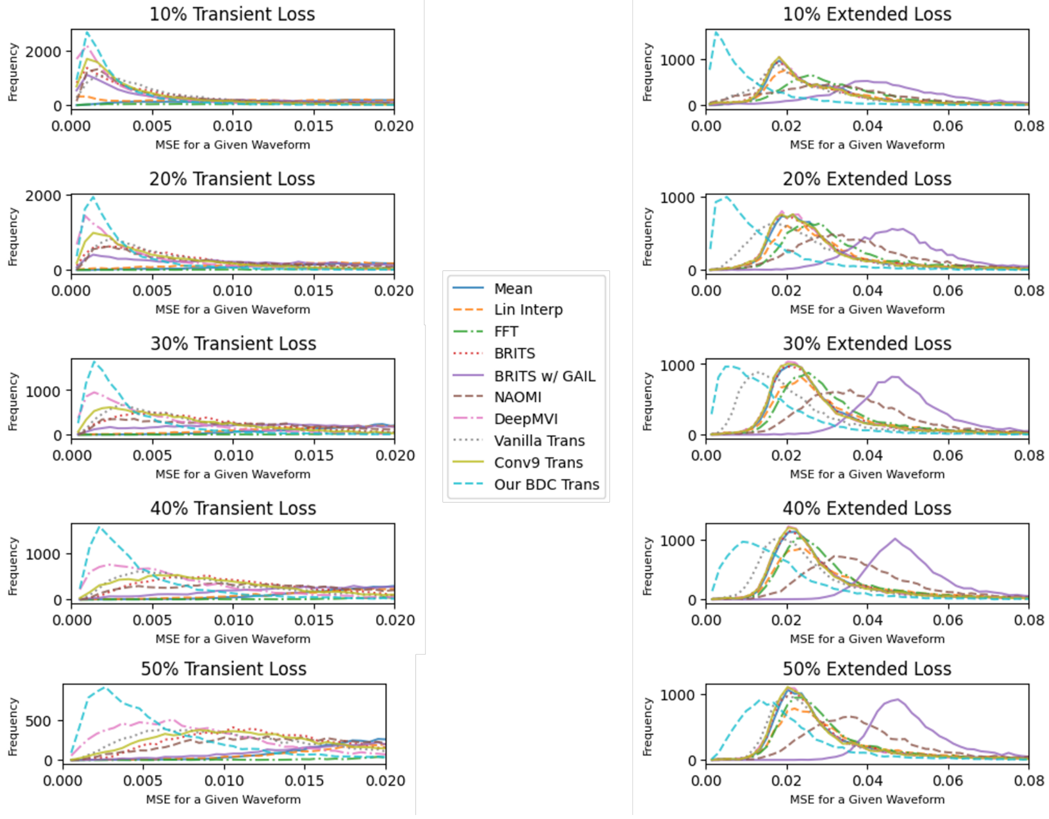


Figure A14: Density of MSE values across all ECG waveforms, for each imputation model for both the transient and extended missingness scenarios. These plots demonstrate imputation models have a much easier time modeling in the transient loss scenario, with most models performing well. However, as missingness duration increases in transient loss, many models exhibit decreased performance, with an increased MSE. In extended loss, most models' performance stays constant (except for BDC and Vanilla transformer), MSE seemingly independent from amount of missingness. This makes sense if we also look at the visualizations in Figure A13 above. The GAN methods (e.g. NAOMI and BRITS w/ GAIL) do not seem to depend heavily on the non-ablated data because their reconstructions do not mimic the non-ablated data, and the other methods (e.g. BRITS and DeepMVI) have simple flat-line imputations, regardless of the amount missing.

A6 Dataset and License details

The code repository for our benchmarking challenge can be found here, www.github.com/rehg-lab/pulseimpute, and it is licensed under the MIT License. The intended use of our curated datasets and missingness patterns is to be used in conjunction with our PulseImpute challenge framework. The training and evaluation procedures for our challenge can be found in our code repo and is described in the main text and Appendix A4.

Our curated datasets and missingness patterns can be found linked here, www.doi.org/10.5281/zenodo.7129965, and we license them under the Creative Commons Attribution 4.0 International. The ECG and PPG waveforms are a form of personal data, but the identifiers have been removed and its public redistribution is in public interest. The data we provide also does not contain any offensive content. We, the authors, bear all responsibility to withdraw our paper and data in case of violation of licensing or patient privacy rights, and confirmation of the data license. The curated data and missingness patterns are organized as shown below:

```
/pulseimpute_data/
 |__README.md
 |__missingness_patterns/
 |   |__mHealth_missing_ecg/
 |   |   |__missing_ecg_train.csv
 |   |   |__missing_ecg_val.csv
 |   |   |__missing_ecg_test.csv
 |   |__mHealth_missing_ppg/
 |       |__missing_ppg_train.csv
 |       |__missing_ppg_val.csv
 |       |__missing_ppg_test.csv
 |__waveforms/
     |__mimic_ecg/
     |   |__mimic_ecg_train.npy
     |   |__mimic_ecg_val.npy
     |   |__mimic_ecg_test.npy
     |   |__MIMIC_III_ECG_filenames.txt
     |__mimic_ppg/
     |   |__mimic_ppg_train.npy
     |   |__mimic_ppg_val.npy
     |   |__mimic_ppg_test.npy
     |   |__MIMIC_III_PPG_filenames.txt
     |__ptbxl_ecg/
         |__scp_statements.csv
         |__ptbxl_database.csv
         |__ptbxl_ecg.npy
```

The data is all stored as .npy files, with each row corresponding to a 100 Hz waveform. The missingness patterns are stored in csv files, with each row as a list of tuples of size 2, which represent the binary missingness pattern time-series. The first item in the tuple corresponds to missing (0) or not missing (1) with the second entry corresponding to the length of samples (in 100 Hz) that the missing or not missingness segment lasts. Each of MIMIC-III curated data and missingness patterns have been split into 80/10/10 training/validation/testing splits accordingly. If we concatenate train, validation, and test .npy files in that order, each data index corresponds to the file name at the corresponding line in the MIMIC_III_ECG_filenames.txt or the MIMIC_III_PPG_filenames.txt file.

For the cardiac classification tasks on the PTB-XL data, the labels can be found in the ptbxl_database.csv file and the waveform data in the ptbxl_ecg.npy. We use the original paper's proposed splits to divide the data into 40/10/50 training/validation/testing splits. Imputation models and downstream cardiac classification models are trained and tuned with the 40/10 split, with the classification model training on the clean non-imputed data. Then classification runs inference on the imputed test data in the 50 split to evaluate imputation quality. This large test split was done to allow for future work where the classification model trains directly on imputed data.

Our datasets originate from the curation of two different datasets, MIMIC-III Waveform [12] and PTB-XL [18]. MIMIC-III Waveform uses the Open Data Commons Open Database License v1.0 (linked here), which explicitly allows for the creation and distribution of derivative databases, which we have done via our curation described in Section A2. We have attributed the data to its original source throughout our paper, by citing [12]. PTB-XL uses the Creative Commons Attribution 4.0 International Public License (linked here), which explicitly allows for adaptation and redistribution of the data, and we have attributed the data to its original source throughout our paper, by citing [18]. The missingness patterns for PPG were extracted from analyzing PPG-DaLiA, which is hosted on the UCI Machine Learning Repository (linked here), and does not have an explicit license, but states others may use the dataset for scientific, non-commercial purposes, provided that credit is given, which we have done throughout our paper, by citing [16]. The binary missingness patterns for ECG were extracted from [3], our mHealth study. For additional documentation for the datasets we have mentioned, please see the original publications that the datasets originate from [12, 18, 16, 3].

References

[1] Apple. Take an ecg with the ecg app on apple watch, Dec 2021.

[2] Syed Khairul Bashar, Eric Ding, Allan J Walkey, David D McManus, and Ki H Chon. Noise detection in electrocardiogram signals for intensive care unit patients. *IEEE Access*, 7:88357–88368, 2019.

[3] Soujanya Chatterjee, Alexander Moreno, Steven Lloyd Lizotte, Sayma Akther, Emre Ertin, Christopher P Fagundes, Cho Lam, James M Rehg, Neng Wan, David W Wetter, et al. Smokingopp: Detecting the smoking'opportunity'context using mobile sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–26, 2020.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.

[5] Emre Ertin, Nathan Stohs, Santosh Kumar, Andrew Raij, Mustafa Al'Absi, and Siddharth Shah. Autosense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In *Proceedings of the 9th ACM conference on embedded networked sensor systems*, pages 274–287, 2011.

[6] Melanie RF Gropler, Aarti S Dalal, George F Van Hare, and Jennifer N Avari Silva. Can smartphone wireless ecgs be used to accurately assess ecg intervals in pediatrics? a comparison of mobile health monitoring to standard 12-lead ecg. *PLoS One*, 13(9):e0204403, 2018.

[7] Dongwei Jiang, Xiaoning Lei, Wubo Li, Ne Luo, Yuxuan Hu, Wei Zou, and Xiangang Li. Improving transformer-based speech recognition using unsupervised pre-training. *arXiv preprint arXiv:1910.09932*, 2019.

[8] Vignesh Kalidas and Lakshman Tamil. Real-time qrs detector using stationary wavelet transform for automated ecg analysis. In *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 457–461, 2017.

[9] T Hickey Kathleen, B Biviano Angelo, Hasan Garan, Robert R Sciacca, Teresa Riga, Kate Warren, Ashton P Frulla, Nicole R Hauser, Daniel Y Wang, and William Whang. Evaluating the utility of mhealth ecg heart monitoring for the detection and management of atrial fibrillation in clinical practice. *Journal of atrial fibrillation*, 9(5), 2017.

[10] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[11] MD2K. Cstress data dictionary documentation. `https://github.com/MD2Korg/CerebralCortex-Docs/blob/master/cc_book/content/data_dictionary/features/cstress.md`, Sep 2019.

[12] B Moody, G Moody, M Villarroel, G Clifford, and I Silva III. Mimic-iii waveform database (version 1.0), 2020.

[13] Mimma Nardelli, Nicola Vanello, Guenda Galperti, Alberto Greco, and Enzo Pasquale Scilingo. Assessing the quality of heart rate variability estimated from wrist and finger ppg: a novel approach based on cross-mapping method. *Sensors*, 20(11):3156, 2020.

[14] Qin Qin, Jianqing Li, Yinggao Yue, and Chengyu Liu. An adaptive and time-efficient ecg r-peak detection algorithm. *Journal of Healthcare Engineering*, 2017, 2017.

[15] Satu Rajala, Harri Lindholm, and Tapio Taipalus. Comparison of photoplethysmogram measured from wrist and finger and the effect of measurement location on pulse arrival time. *Physiological measurement*, 39(7):075010, 2018.

[16] Attila Reiss, Ina Indlekofer, Philip Schmidt, and Kristof Van Laerhoven. Deep ppg: large-scale heart rate estimation with convolutional neural networks. *Sensors*, 19(14):3079, 2019.

[17] Adriana N. Vest, Giulia Da Poian, Qiao Li, Chengyu Liu, Shamim Nemati, Amit Shah, and Gari D Clifford. cliffordlab/PhysioNet-Cardiovascular-Signal- Toolbox: PhysioNet-Cardiovascular-Signal-Toolbox 1.0, May 2018.

[18] Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I Lunze, Wojciech Samek, and Tobias Schaeffter. Ptb-xl, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1):1–15, 2020.

[19] P. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[20] Zhenzhen Xie, Ahmet Nacioglu, Calvin Or, et al. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst chinese adults: cross-sectional survey study. *JMIR mHealth and uHealth*, 6(4):e9002, 2018.