CONTRASTIVE HEARTBEATS: CONTRASTIVE LEARNING FOR SELF-SUPERVISED ECG REPRESENTATION AND PHENOTYPING

Crystal T. Wei[⋆] *Ming-En, Hsieh*[⋆] *Chien-Liang Liu*[†] *Vincent S. Tseng*[⋆]

[⋆]Institute of Data Science and Engineering, [†]Department of Industrial Engineering and Management
National Yang Ming Chiao Tung University, Taiwan, ROC

ABSTRACT

The non-invasive and easily accessible characteristics of electrocardiogram (ECG) attract many studies targeting AI-enabled cardiovascular-related disease screening tools based on ECG. However, the high cost of manual labels makes high-performance deep learning models challenging to obtain. Hence, we propose a new self-supervised representation learning framework, contrastive heartbeats (CT-HB), which learns general and robust electrocardiogram representations for efficient training on various downstream tasks. We employ a novel heartbeat sampling method to define positive and negative pairs of heartbeats for contrastive learning by utilizing the periodic and meaningful patterns of electrocardiogram signals. Using the CT-HB framework, the self-supervised learning model learns personalized heartbeat representations representing the specific cardiology context of a patient. Evaluations on public benchmark datasets and a private large-scale real-world dataset with multiple tasks demonstrate that the learned semantic representations result in better performance on downstream tasks and retain high performance while supervised learning suffers performance degradation with fewer supervised labels in downstream tasks.

***Index Terms*—** Electrocardiogram, Self-supervised learning, Representation learning, Contrastive learning

1. INTRODUCTION

Electrocardiogram (ECG) is a valuable biosignal that measures the electrical activity of the heart and is commonly used to diagnose cardiac arrhythmias with the merits of being non-invasive, fast to acquire at a relatively low-cost. Owing to numerous clinical usages, ECG has become a promising medium for various studies to apply deep learning in the medical domain. The high bar for model's performance in the medical domain poses significant challenges to obtain large amounts of labeled data used to train high-performance supervised models. Due to the high cost of manual labels, some studies proposed to utilize unlabeled data conveniently collected from electronic medical records through unsupervised learning [1][2] and self-supervised learning [3][4][5][6]. Recently, self-supervised learning methods have proven to learn good representations for various downstream tasks through non-manually labeled datasets. Among state-of-the-art self-supervised approaches, contrastive methods outperform others and reached comparable performance to supervised learning in computer vision [7][8][9].

We propose the contrastive heartbeats (CT-HB) framework, a novel approach to learn generalized ECG representations through contrastive self-supervised learning, allowing efficient training on various downstream tasks. We achieve higher performance than other state-of-the-art unsupervised learning and self-supervised learning methods in ECG applications. The proposed CT-HB framework can even outperform supervised learning under the low data paradigm, which alleviates costly labels in the medical domain.

Our contributions are summarized as the following:

1. We propose a novel contrastive learning approach, the CT-HB framework, to utilize the periodic and meaningful patterns from electrocardiogram signals.
2. The electrocardiogram representations learned using the CT-HB framework outperform representations of other state-of-the-art methods when applied to various disease prediction tasks.
3. The proposed CT-HB framework shows the potential to alleviate the challenge of costly labels in medical applications as it can outperform supervised learning in the low data paradigm with a simple non-linear classifier.

2. METHODOLOGY

2.1. Contrastive Heartbeats (CT-HB) Framework

We are motivated by Pretext-Invariant Representation Learning [10], which tries to learn semantic representations by forcing similar representation between the augmented image and its original image. Similar intuitions can be adopted for ECG. For the same person, each heartbeat is slightly different but

This research was supported in part by Ministry of Science and Technology Taiwan under grant no. MOST 110-2634-F-A49-002.

represents the same cardiology context. Each heartbeat can be seen as slight augmentations to the unique cardiology context of this person. We train our model to produce similar representations between the same person's heartbeats through contrastive loss. Namely, we design the contrastive heartbeat framework to learn personalized heartbeat representations representing the specific cardiology context of a patient. From the experiment results, we show that these semantic representations benefits a variety of downstream tasks.

An overview of the proposed CT-HB framework is shown in **Figure 1**. In order for the target model f_θ to learn personalized heartbeat representations, we split the recording X into heartbeats x_1 to x_T. For each heartbeat x, we start by extracting its corresponding representation y through model f_θ. We then project the embedding through a fully connected projection layer g_θ before passing through the contrastive loss for training. The pre-trained model f_θ apply as a feature extractor for heartbeat representations targeting various downstream tasks. We evaluate those representations following standard linear evaluation using logistic regression.

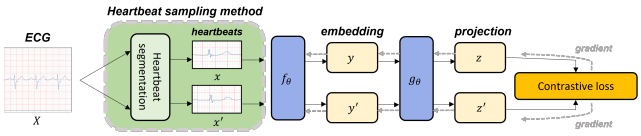

Fig. 1. Contrastive Heartbeats (CT-HB) framework.

2.2. Heartbeat Sampling Method

The heartbeat sampling method is the key component in CT-HB. To train the representation through contrastive loss, we define a positive pair as the anchor heartbeat with a positive heartbeat and a negative pair as the anchor heartbeat with a negative heartbeat. We sample the anchor and positive heartbeats from the same ECG and negative heartbeats from other ECG. This sampling method allows the model to learn personalized heartbeat representations representing the specific cardiology context of a patient utilizing the periodic and meaningful patterns from ECG signals. The anchor heartbeat is the baseline sample for other heartbeats to define similarity, which can be sampled from any ECG within the dataset. Let D be the whole ECG dataset and $X_i \in D$ as one ECG containing a set of heartbeats $\{x_1, ..., x_T\}$. We sample one anchor heartbeat $x^{anc} \in X_i$, M positive heartbeats $x^{pos} \in X_i$ where $x^{pos} \neq x^{anc}$, and K negative heartbeats $x^{neg} \in X_j$ where $X_j \in D \; and \; X_i \neq X_j$. An illustration of the heartbeat sampling method can be seen in **Figure 2**.

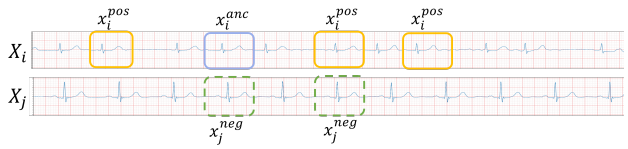

Fig. 2. Demonstration of the heartbeat sampling method.

2.3. Model Structure

We adopt the state-off-the-art time-series model, CausalCNN, proposed by Franceschi et al. [4] for our backbone model f_θ. The model has shown to be efficient compared to LSTMs while better at learning long-range dependencies compared to standard convolutional layers. We apply additional fully connected layer g_θ to project the features of the anchor, positive, and negative samples for contrastive learning.

2.4. Contrastive Loss

We employ multi-similarity (MS) loss [11], as shown in **Equation 1**, which is an alternative to triplet loss [4]. Compared to other triplet loss alternatives, MS loss takes into account all types of pair-based weighting allowing a better ability to map triplet pairs in the feature space [11]. We aggregate all negative samples within a mini-batch to reduce the sensitivity for outliers within a mini-batch for each anchor.

$$\frac{1}{D}\sum_{i=i}^{D}\{\frac{1}{\alpha}log(1+\sum_{k\in(x_m^{pos})}e^{-\alpha(S_{ik}-\lambda)})+\frac{1}{\beta}log(1+\sum_{k\in N_i}e^{-\beta(S_{ik}-\lambda)})\} \quad (1)$$

where S_{ik} is the (i, k) element in similarity matrix S,
N_i for all negatives in a batch and fixed hyperparameters λ, α, β

2.5. Heartbeats Ensemble Method

There are two types of ECG dataset labels, heartbeat labels and full-length ECG labels. We propose a heartbeats ensemble method to adopt the pre-trained model for full-length ECG labels, which cannot directly apply the learned heartbeat representation from self-supervised learning. We extract the heartbeats in the full-length ECG by using the Hamilton R-peak segmentation algorithm [12] implemented in BioSPPy [13] and ensemble them together. Since the normal range of heart rate is from 60 to 100 bpm when a person is awake, we exclude samples with less than eight heartbeats ($<$ 48 bpm) within the ten-second measurement. We randomly choose eight heartbeats from all the heartbeats extracted and average the eight logits for training and evaluation.

3. EXPERIMENTAL DESIGN

3.1. Datasets

We evaluate our CT-HB performance on three datasets. Due to the heartbeat labeling method, MIT-BIH arrhythmia database [14] is chosen as our main comparison dataset. We follow all data processing methods from [15] and use lead II of the ECG signal for training to predict three arrhythmia

classes defined by the AAMI standard. The Chapman dataset [16] is composed of 12-lead ECG recordings from 10,646 patients alongside 11 different classes of cardiac arrhythmia. We follow the suggestion [16] to integrate labels into four classes. Our private large-scale ECG dataset contains 705,924 ten seconds 12-lead ECG recordings from the ECG database of a large National Medical Center in Taiwan. Among various cardiovascular-related diseases, we choose three heart diseases as classification targets, including Left ventricular hypertrophy (LVH), Type 1 atrioventricular block (1AVB), and Atrial flutter (AFL). For LVH, we include 61,422 valid patients having echocardiogram recordings within 30 days from the time point of ECG recordings and obtained labels from doctor annotation acquired from the echocardiogram. The 1AVB dataset is collected from the same patient set as LVH, and the labels are obtained from doctor annotations on the electrocardiogram. For AFL, we obtain the labels from doctor annotation on the ECG, and select all samples after 2013 to obtain 312,888 recordings from different patients. All data splitting process follows the patient non-overlapping method with the training, validation, and testing ratio of 7:1:2.

3.2. Experimental Setups

We adopt the logistic regression protocol for linear evaluation [7], which intend to evaluate whether the learned representations are linearly separable. We also evaluate on semi-supervised learning similar to settings in [7] where we use only a portion of labeled data during fine-tuning. It is common in medical scenarios to have much fewer labels in uncommon diseases. An unsupervised pre-trained model that can reach or even outperform supervised learning performance when trained on low label counts would be beneficial for various important downstream tasks. To avoid negative transfer on small datasets such as MIT-BIH, we only unfreeze the last causal convolutional block and linear layer in the pre-trained model. On the other hand, for our large-scale private dataset, we allow fine-tuning of the whole model. The different strategies for different dataset scales are common in transfer learning, where the representation corrupts when fine-tuning the pre-trained model with relatively small data samples. All CT-HB results are conducted with a batch size of 64 and an embedding size of 128 with M and K equal to 5.

3.3. Baselines Methods

We compared our CT-HB framework with state-of-the-art ECG unsupervised learning methods based on generative approaches [3][2] targeting the benchmarking MIT-BIH dataset, and state-of-the-art ECG self-supervised learning methods [5][6]. Ochiai et al. [3] utilizes convolutional denoising autoencoders to pre-train a model and fine-tuning the model with an additional fully connected layer for heart diseases classification. Nurmaini et al. [2] conducts similar experiments as Ochiai et al. [3] while applying deep autoencoders for pre-training. Both of these approaches do not evaluate the learned ECG representations using standard linear representation evaluations. Apart from generative approaches, Sarkar and Etemad [5] adopt the transformation-based method of self-supervised learning in ECG representation learning for emotion recognition. Kiyasseh et al. [6] proposed the self-supervised learning method for ECG representation learning utilizing contrastive learning to learn patient-specific representations. For rigorous comparison, we also compare with the state-of-the-art self-supervised learning method for general time-series data [4]. Franceschi et al. [4] proposed a general time-series self-supervised learning framework using the contrastive method with the CausalCNN model to learn representations and achieve good performance on benchmark time-series datasets with comprehensive experiment results.

3.4. Evaluation Metrics

For multi-class classification, we reported accuracy (ACC), macro area under receiver operating characteristic curve (AUROC), and Matthews correlation coefficient (MCC) as three primary metric candidates. According to [17], MCC is a better surrogate for accuracy in an imbalanced dataset. We also reported individual single class performance, including sensitivity (SEN), specificity (SPEC), and positive prediction value (PPV). For binary classification, we utilize standard AUROC and sensitivity. To form a fair comparison, we calculate the sensitivity value at a fixed operation point of 0.9 specificity.

4. EXPERIMENTAL RESULTS

4.1. Linear Evaluation of Representation

The MIT-BIH dataset results shown in **Table 1** indicates that the CT-HB framework surpass the best baseline by the margin of 2.47% for accuracy, 2.11% for macro AUROC and 9.20% for MCC. Besides, the CT-HB framework significantly surpass baselines in the minor class, SVEB. The Chapman dataset results shown in **Table 2** indicate that the CT-HB framework outperforms the state-of-the-art self-supervised learning method by nearly 2% in AUROC. To demonstrate that the proposed method is applicable to large-scale datasets in the real-world, we evaluated our CT-HB framework with the baseline methods on a large-scale private dataset on two cardiovascular diseases. As shown in **Table 3**, the CT-HB framework gets 4.19% absolute improvement on sensitivity and 1.23% on AUROC for LVH, and the performance of 1AVB is significantly better than the strongest baseline with a margin of more than 22% of improvement for sensitivity.

4.2. Robustness to Unseen Patient Population

To evaluate the robustness of the learned representations when used on an unseen patient population, we evaluated the

Table 1. Linear evaluation performance on the MIT-BIH dataset.

Method	ACC	Normal			SVEB			VEB			Macro	MCC
	%	SEN	PPV	SPEC	SEN	PPV	SPEC	SEN	PPV	SPEC	AUROC	
Supervised learning:												
Mousavi and Afghah (II) [15]	**99.53**	**99.68**	**99.55**	**96.05**	**88.94**	**92.57**	**99.72**	**99.94**	**99.50**	**99.97**	-	-
Garcia et al. (II + V) [18]	92.40	94.00	98.00	82.60	62.00	53.00	97.90	87.30	59.40	95.90	-	-
CausalCNN (II)	94.89	97.20	97.37	77.00	44.24	55.77	98.59	92.03	82.03	98.52	0.9522	0.7323
Linear evaluation:												
Ochiai et al. [3]	72.09	73.67	96.83	78.88	26.80	6.57	85.24	76.19	30.15	87.66	0.8020	0.3033
Nurmaini et al. [2]	84.17	86.14	97.47	80.45	30.72	12.24	91.47	87.57	50.48	94.00	0.8716	0.4519
Sarkar and Etemad [5]	85.70	86.77	97.60	81.34	50.27	19.58	92.01	91.27	55.99	94.99	0.9216	0.5127
Franceschi et al. [4]	86.78	89.23	96.46	71.36	25.93	12.63	**93.06**	87.91	61.49	96.15	0.8826	0.4813
CT-HB	**89.25**	**90.01**	**98.23**	**85.81**	**66.45**	**24.33**	92.00	**91.89**	**79.08**	**98.30**	**0.9427**	**0.6047**

Table 2. Linear evaluation performance on Chapman dataset.

Models	Supervised (CausalCNN)	SimCLR [7]	CLOCS [6]	Franceschi et al. [4]	CT-HB
AUROC	0.977	0.775	0.902	0.893	**0.920**

Table 3. Linear evaluation performance on private dataset.

	LVH		1AVB	
Models	Sensitivity	AUROC	Sensitivity	AUROC
Supervised learning:				
CausalCNN	69.77	0.8300	-	-
Linear evaluation:				
Ochiai et al. [3]	43.66	0.8050	-	-
Nurmaini et al. [2]	44.49	0.8046	-	-
Sarkar and Etemad [5]	42.89	0.7747	-	-
Franceschi et al. [4]	45.73	0.8055	51.46	0.8255
CT-HB	**49.94**	**0.8178**	**73.65**	**0.9089**

learned representation on a larger set of patients where about 73 percent of patients are unseen in the original training data for representation learning. We want to show that the representations learned by our proposed CT-HB framework are general enough that we can utilize a relatively small dataset for the unsupervised learning phase to train the general representation of heartbeat for all other patients. **Table 4** shows that we can significantly improve the sensitivity for the highly imbalanced disease class, indicating that the representation not only generalized to different prediction tasks but also generalized to new unseen patients. This ability is crucial for the CT-HB framework when applied to real-world applications.

Table 4. Performance of the unseen patient population.

Models	AFL Sensitivity	AFL AUROC
Franceschi et al.[4]	68.58	0.8913
CT-HB	**75.14**	**0.9122**

4.3. Semi-supervised Learning

To understand whether fine-tuning improves the downstream performance under both the full dataset and in the low data paradigm, we employ semi-supervised learning, which fine-tunes the pre-trained model using a partial amount of labels. **Table 5** shows that fine-tuning can significantly increase the downstream performance on the MIT-BIH dataset under all drop ratios for macro AUROC. The results of the LVH dataset shows that the proposed method can consistently surpass supervised learning under both full and low data sizes. This demonstrates that our proposed method can alleviate the problem of requiring large amounts of labeled data to produce high-performance models compared to supervised learning.

Table 5. Performance of semi-supervised learning on MIT-BIH and LVH dataset.

	Dataset	MIT-BIH			LVH
Learning Method	Drop ratio	ACC%	Macro AUROC	MCC	AUROC
CT-HB	0%	**95.23**	**0.9614**	**0.7672**	**0.8378**
	50%	94.61	**0.9604**	**0.7385**	**0.8312**
	90%	**95.74**	**0.9501**	**0.7603**	**0.8104**
Supervised (CausalCNN)	0%	94.89	0.9522	0.7323	0.8300
	50%	**95.10**	0.9428	0.7324	0.8245
	90%	94.70	0.9203	0.7022	0.7898

5. CONCLUSION

We conducted a series of experiments on two public datasets and a private large-scale real-world ECG dataset to demonstrate that the proposed contrastive heartbeats (CT-HB) framework reduces the performance gap between supervised learning and unsupervised learning in downstream tasks such as ECG phenotyping. By utilizing self-supervised learning for pre-training, we alleviate the need to collect large amounts of labeled data compared to pure supervised learning approaches. We demonstrated that the proposed CT-HB framework is robust to unseen patient populations and could retain high performance in the low data paradigm compared to supervised learning approaches. Our proposed CT-HB framework enables self-supervised learning to be successfully adapted to the periodic and meaningful patterns of ECG signals and shows the potential of employing it in real-world ECG applications.

6. REFERENCES

[1] Shawn Tan, Guillaume Androz, Ahmad Chamseddine, Pierre Fecteau, Aaron Courville, Yoshua Bengio, and Joseph Paul Cohen, "Icentia11k: An unsupervised representation learning dataset for arrhythmia subtype discovery," *arXiv preprint arXiv:1910.09570*, 2019.

[2] Siti Nurmaini, Radiyati Umi Partan, Wahyu Caesarendra, Tresna Dewi, Muhammad Naufal Rahmatullah, Annisa Darmawahyuni, Vicko Bhayyu, and Firdaus Firdaus, "An automated ecg beat classification system using deep neural networks with an unsupervised feature extraction technique," *Applied Sciences*, vol. 9, no. 14, pp. 2921, 2019.

[3] Keiichi Ochiai, Shu Takahashi, and Yusuke Fukazawa, "Arrhythmia detection from 2-lead ecg using convolutional denoising autoencoders," in *Proceedings of the KDD*, 2018, vol. 18.

[4] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi, "Unsupervised scalable representation learning for multivariate time series," in *Advances in Neural Information Processing Systems*, 2019, pp. 4652–4663.

[5] Pritam Sarkar and Ali Etemad, "Self-supervised learning for ecg-based emotion recognition," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 3217–3221.

[6] Dani Kiyasseh, Tingting Zhu, and David A Clifton, "Clocs: Contrastive learning of cardiac signals across space, time, and patients," in *International Conference on Machine Learning*. PMLR, 2021, pp. 5606–5615.

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton, "A simple framework for contrastive learning of visual representations," in *Proceedings of the 37th International Conference on Machine Learning, ICML*. 2020, vol. 119 of *Proceedings of Machine Learning Research*, pp. 1597–1607, PMLR.

[8] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9729–9738.

[9] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Ávila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko, "Bootstrap your own latent - A new approach to self-supervised learning," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*, 2020.

[10] Ishan Misra and Laurens van der Maaten, "Self-supervised learning of pretext-invariant representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6707–6717.

[11] Xun Wang, Xintong Han, Weilin Huang, Dengke Dong, and Matthew R Scott, "Multi-similarity loss with general pair weighting for deep metric learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5022–5030.

[12] Pat Hamilton, "Open source ecg analysis," in *Computers in cardiology*. IEEE, 2002, pp. 101–104.

[13] Carlos Carreiras, Ana Priscila Alves, André Lourenço, Filipe Canento, Hugo Silva, Ana Fred, et al., "Biosppy: Biosignal processing in python," *Accessed on*, vol. 3, no. 28, pp. 2018, 2015.

[14] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[15] Sajad Mousavi and Fatemeh Afghah, "Inter-and intra-patient ecg heartbeat classification for arrhythmia detection: a sequence to sequence deep learning approach," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1308–1312.

[16] Jianwei Zheng, Jianming Zhang, Sidy Danioko, Hai Yao, Hangyuan Guo, and Cyril Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Scientific data*, vol. 7, no. 1, pp. 1–8, 2020.

[17] Giuseppe Jurman, Samantha Riccadonna, and Cesare Furlanello, "A comparison of mcc and cen error measures in multi-class prediction," *PloS one*, vol. 7, no. 8, 2012.

[18] Gabriel Garcia, Gladston Moreira, David Menotti, and Eduardo Luz, "Inter-patient ecg heartbeat classification with temporal vcg optimized by pso," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.