2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Self-Supervised Learning with Electrocardiogram Delineation for Arrhythmia Detection

Byeong Tak Lee* Seo Taek Kong* Youngjae Song Yeha Lee†

***Abstract*— Electrocardiogram (ECG) signals convey immense information that, when properly processed, can be used to diagnose various health conditions including arrhythmia and heart failure. Deep learning algorithms have been successfully applied to medical diagnosis, but existing methods heavily rely on abundant high-quality annotations which are expensive. Self-supervised learning (SSL) circumvents this annotation cost by pre-training deep neural networks (DNNs) on auxiliary tasks that do not require manual annotation. Despite its imminent need, SSL applications to ECG classification remain under-explored. In this work, we propose an SSL algorithm based on ECG delineation and show its effectiveness for arrhythmia classification. Our experiments demonstrate not only how the proposed algorithm enhances the DNN's performance across various datasets and fractions of labeled data, but also how features learnt via pre-training on one dataset can be transferred when fine-tuned on a different dataset.**

***Index Terms*— Electrocardiography, arrhythmia classification, self-supervised learning**

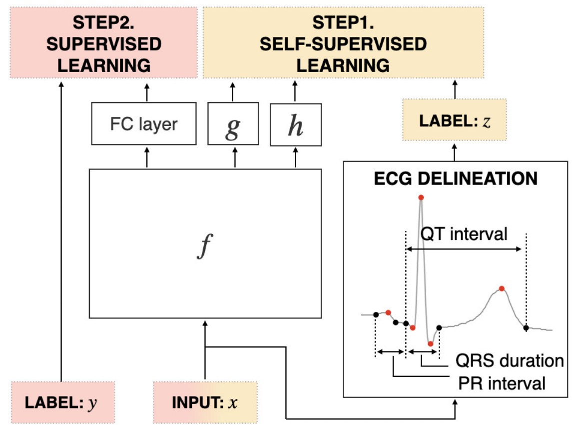

Fig. 1. Schematic diagram for pre-training and fine-tuning.

I. Introduction

Electrocardiogram (ECG) signals are recordings of the heart activity, and a primary inspection step in diagnosing cardiovascular diseases including heart failure prediction involves analyzing ECG signals. Modern deep learning (DL) algorithms have applied convolutional and/or recurrent network architectures to classify short segments of ECG signals collected offline [1], [2], [3], [4], [5], but much of their success can be attributed to massive amounts of labeled data used for training. Constructing such a rich annotated dataset is extremely expensive, and this training set acquisition cost is an immediate hurdle that must be overcome for DL to be applied to medical diagnosis.

To alleviate the annotation costs associated with DL, self-supervised learning (SSL) algorithms have been devised to leverage unlabeled data. SSL pre-trains a network on auxiliary tasks to learn features relevant to the downstream task, and has emerged as a crucial component when DNNs are trained on difficult tasks with insufficient labeled data [6], [7], [8], [9]. While SSL algorithms are often evaluated on tasks with few labeled but abundant unlabeled data, even acquiring large amounts of unlabeled data is infeasible for medical modalities, and general-purpose SSL techniques may not be able to achieve as high of an accuracy on such difficult tasks. Moreover, most SSL applications are limited to natural image classification and language processing, and algorithms specific to ECG classification are absent up to date.

* Equal contribution
VUNO Inc, Seoul, South Korea.
† Correspondence to yeha.lee at vuno.co

In this work, we investigate whether known clinical features based on ECG delineation can be used for SSL pre-training to reinforce DNNs in detecting arrhythmia. Our experiments demonstrate how learning key-features in ECG signals before performing classification can consistently enhance the performance of DNNs. The proposed algorithm is shown to outperform all general-purpose SSL baseline algorithms, confirming that domain-specific knowledge should be incorporated when available. Furthermore, we show how semantics learnt from one dataset can be transferred to other datasets which is clinically invaluable to deploy DL-based algorithms to practical clinical settings.

II. Method

A. Electrocardiogram Delineation

Electrocardiogram delineation describes ECG signals as a collection of critical features including heart rate and intervals in the QRS complex which is the exemplar waveform of ECG signals. In designing an SSL algorithm for ECG classification, we adopted two ECG delineation algorithms to extract interval-based [10] and axis-based features [11]. The heart rate, QRS duration, PR-interval, QT-interval, and QT-corrected features are extracted by detecting the QRS complex, searching over scales of maximum modulus lines [12], and identifying the onset and offset of the QRS characterized by its first and last significant slopes. Because the slopes and critical points occur at different time instances within the cardiac cycle, both operations are performed in the wavelet domain spanned by quadratic spline basis functions [13]. The P and T wave detection and delineation algorithm follows

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

from the window-based thresholding in [10]. Based on the QRS complex and P&T waves obtained above, the electrical axes were computed using leads I and III as in the method developed by [11] with the net voltage obtained as follows. The QRS net voltage was computed by subtracting the Q and S amplitudes from the R peak, and the net voltages of P and T waves were determined by measuring the difference between the onset and peak amplitudes.

B. Self-Supervised Learning

The ECG delineation described above has been used as a dimension reduction technique using the derived features to classify cardiac abnormalities using traditional machine learning [14] and DL algorithms [15]. ECG delineation has been used in combination with other high-level features extracted using convolutional neural networks (CNNs) in the latter work. Here, we propose an approach to learn the above features before fine-tuning for arrhythmia detection.

Fig. I illustrates the overall procedure. All 5 interval and 3 axes features are associated with L (# leads) and N (# time steps) dimensional vectors, respectively. Each k^{th} feature $z_k \in \mathcal{Z}_k$ is classified as normal $\mathcal{N}(\mathcal{Z}_k)$ and abnormal subjects using a standard reference [16]. The features' mean μ and diagonal variance matrix Σ are computed, and a DNN is decomposed into a feature extraction module f preceding regression $g = (g_1, \ldots, g_K)$ and classification $h = (h_1, \ldots, h_K)$ fully-connected (FC) layers. Our SSL algorithm then trains the DNN using a regression and classification (cross-entropy) loss:

$$g^*, h^* = \arg\min_{g,h} \alpha \left\| g\left(f(x)\right) - \Sigma^{-1}\left(z - \mu\right) \right\|_2^2 + (1-\alpha) H\left(\mathbf{1}\left\{z \in \mathcal{N}(\mathcal{Z})\right\}, h(f(x))\right) \quad (1)$$

where the former mean-squared error (MSE) loss aligns the DNN's latent space $\mathcal{Z} = \otimes_{k=1}^{K} \mathcal{Z}_k$ induced by ECG signals to clinically-relevant features and the classification loss guides the network to identify normal and abnormal features (in contrast to the subject's condition of arrhythmia or normal). After training until convergence, the DNN is fine-tuned to classify arrhythmia using the ECG signals with true annotations indicating the subject's condition.

A preliminary experiment showed how inaccurately extracted features deteriorate the downstream classification task, and we processed the features to filter out noise. First, the sample-wise standard deviation $\hat{\sigma}_k$ was computed for each feature $z_k \in \mathbb{R}^{d_k}$ with $d_k = NL$ for interval and $d_k = N$ for axis features. Any loss values computed over features whose standard deviation $\hat{\sigma}_k$ exceeded some threshold, found via hyperparameter search, were masked to avoid abrupt gradient changes or noisy targets. Next, the features were smoothed by replacing the raw vectors z_k with its median, resulting in scalar-valued features $\mathcal{Z}_k \subset \mathbb{R}, \forall k$.

Initialization has played a key-role in modern DL, where proper initialization is now necessary to avoid gradient explosion/vanishing [17] or transfer learning is employed when the downstream task lacks training samples. Recent studies revealed how ImageNet pre-training may simply speed up the rate of convergence [18], limiting its applicability to non-natural image modalities which have drastically different features. In contrast, the auxiliary task described above initializes a network such that it can extract domain-specific features informative of cardiovascular diseases in ECG signals unlike general-purpose SSL algorithms.

III. Experiments

A. Benchmarks

Datasets: The SSL algorithms were evaluated on 3 publicly available datasets with train/validation/test proportions of 70/15/15%. The public datasets we considered comprise $L = 12$ lead ECG signals which were split into 10 second segments sampled at frequency $f_s = 500$Hz. The task of interest varied across datasets: **CPSC** [19] consists of 6,877 ECG signals recorded over 6 to 60s. The model was trained to perform 9-class classification differentiating among normal, atrial fibrillation, first-degree atrioventricular block, left bundle branch block, right bundle branch block, premature atrial contraction, premature ventricular contraction, ST-segment depression, and ST-segment elevated. **PTBXL** [20] consists of 21,837 ECG records recorded for 10s. The model was trained to distinguish myocardial infarction from any condition in normal, cardiomyopathy/heart failure, bundke branch block, dysrhythmia, myocardial hypertrophy, valvular heart disease, myocarditis, and healthy controls. **Shaoxing** [21] consists of 10,615 ECG signals each with 10s segments. The model was trained to perform 7-class classification among sinus bradycardia, sinus rhythm, atrial fibrillation, sinus tachycardia, atrial flutter, sinus irregularity, and supraventricular tachycardia.

Baseline Algorithms: To demonstrate the efficacy of ECG-specific pre-training, we compared the median per-

F1 (%)	CPSC					PTB					Shaoxing				
Labeled (%)	1	3	10	30	100	1	3	10	30	100	1	3	10	30	100
None	30.13	46.07	66.37	72.15	79.00	53.84	60.71	65.95	69.09	**72.58**	48.91	62.90	78.90	84.91	87.37
BYOL	33.07	49.03	67.44	73.74	79.65	57.50	61.55	66.08	69.77	71.42	52.59	69.17	80.11	86.46	89.19
SimCLR	32.52	53.29	68.39	75.02	80.35	59.67	**63.33**	66.47	69.81	72.46	54.13	68.66	81.32	85.40	**89.86**
DC	33.85	55.93	68.04	74.01	79.47	57.84	62.17	66.41	70.00	71.78	54.40	69.64	80.70	85.82	88.42
ECG	**41.56**	**59.13**	**69.18**	75.01	**81.22**	**60.19**	62.55	**66.86**	**70.65**	72.49	**56.35**	**69.86**	**81.49**	**86.80**	89.41

TABLE I

Classification performances when pre-trained and fine-tuned on the same dataset with different fractions of labeled data. Bold indicates the highest performances exceeding the next highest by at least 0.05% (half the 2nd least significant digit).

Tuning layers	Source	CPSC			PTB			Shaoxing		
	Target	CPSC	PTB	Shaoxing	CPSC	PTB	Shaoxing	CPSC	PTB	Shaoxing
All	BYOL	79.65	72.26	88.77	79.84	71.42	88.84	79.65	72.19	89.19
	SimCLR	80.35	**72.90**	89.06	79.22	72.46	89.20	79.63	72.47	**89.86**
	DC	79.42	71.94	88.86	80.10	71.78	88.77	**80.34**	71.96	88.42
	ECG	**81.22**	72.54	**90.13**	**80.37**	**72.49**	**89.27**	80.25	72.43	89.41
Final	BYOL	50.41	58.07	70.27	52.75	55.62	70.15	50.98	57.79	71.28
	SimCLR	58.82	63.40	68.88	57.99	64.08	69.90	55.69	**62.32**	69.51
	DC	52.63	60.88	75.87	54.50	58.60	75.78	55.68	59.97	71.53
	ECG	**68.28**	**64.15**	**81.34**	**67.92**	**65.93**	**81.72**	**58.65**	61.40	**71.75**

TABLE II

TRANSFER LEARNING PERFORMANCES WHEN FINE-TUNED ON TARGET DATASETS AT (TOP) ALL OR (BOTTOM) FINAL LAYERS.

formances of the proposed algorithm with that of random initialization (None) and 3 state-of-the-art SSL algorithms. BYOL [22] and SimCLR [23] are built on the contrastive learning framework, maximizing the agreement between different views of the same input. Deep clustering (DC) [24] also maximizes the similarity between current and clustered previous representations. These algorithms require forming different views of the same signal, where we used stochastic augmentations identical to those used for fine-tuning, but with different probabilities p depending on the algorithm to maximize performance: Gaussian smoothing with window size uniformly sampled from 1 to 5 ($p = 0.5$), additive Guassian noise ($\sigma = 0.1, p = 0.5$), resampling up to $\pm 15\%$ of the original frequency ($p = 0.8/0.8/0.8/0.5$), and masking 0 to 50% of data points in the ECG signal ($p = 0.8/0.8/0/0$). None used the same augmentations as the proposed algorithm for fine-tuning.

Implementation Details: We used the m-ResNet architecture [3] which outperformed the average of expert cardiologists in ECG classification. Both pre-training and fine-tuning used the Adam optimizer with identical configurations except for learning rates 10^{-3} and $5 \cdot 10^{-4}$, respectively, batch sizes 128 and 32, and weight decay 10^{-5} and 0. Regularization with dropout $p = 0.1$ for encoding and $p = 0.5$ for fully-connected layers was also employed. The α parameter used to control the convex combination between regression and classification loss in was found using grid search.

B. Classification Performance

The classification performance (F1-score) after SSL pre-training along with purely-supervised learning (None) is reported in Tab. I with increasing fractions of labeled training data. With 1% of the dataset annotated, the proposed algorithm attains an $\sim 9\%$ absolute gain in performances on average over a purely supervised algorithm. It is clear that the proposed algorithm is most effective when the number of labeled data is small, and performance gain is evident even when all labeled data is used. All SSL algorithms were able to attain relatively high classification accuracy compared to purely-supervised learning especially when the number of labeled data was small. This underlines how initialization schemes designed to efficiently train DNNs in general may not be sufficient to learn complicated tasks especially when insufficient annotations are provided.

C. Transfer Learning Performances

Because collecting medical data is difficult, the transferrability of features learnt from one dataset to another is paramount. To validate the transferability of features between different domains, we pre-trained a DNN using the SSL algorithms on source datasets and fine-tuned on different target datasets using all available annotations in the fine-tuning phase. As shown in Tab. II, all SSL algorithms excel at transferring features learnt from the source to target dataset, with ECG performing best on most of the tasks. From this table, it is difficult to determine whether the pre-training procedures help learn clinically-relevant features (representation) or if they are effectively yielding nice initialization schemes for better optimization as argued for natural modalities [18]. The former property may be necessary when a model need be interpretable, e.g. for medical applications, and the next section attempts to unveil their effects.

D. Importance of Clinically-Relevant Features

Here, the importance of ECG delineation as learning clinically-relevant features is highlighted by first training a DNN on all SSL algorithms, but then fine-tuning only the final linear layer when performing arrhythmia detection. Had other general-purpose SSL algorithms helped learn features relevant to ECG classification, a linear model should have also performed well on the downstream task. However, the proposed algorithm outperformed all other SSL schemes by a large margin when the linear model with fixed feature extractors was trained on both the same or different fine-tuning datasets using all available annotated data. In conclusion, pre-training on clinically relevant features enhances the arrhythmia detection performance in contrast to the observation for natural imaging modalities where fine-tuning only improved the rate of convergence, and the enhancement is persistent across training both deep and linear networks.

If the features obtained from ECG delineation contained enough information to describe arrhythmia, it would be sufficient to use those features as a dimension-reduction scheme and perform linear classification as in works described previously. The comparison in Tab. II demonstrates that this is not the case, and that fine-tuning all layers helps learn additional features absent in ECG delineation.

E. Ablation Study

We conducted an ablation study to understand how each component of the proposed algorithm affected the overall classification performance. Table III compares how the regression and classification loss functions, or their combination, affected performance. Each regression and classification losses are shown to improve overall performance by nearly equal margins on average, with their combination consistently enhancing the classification performance.

CPSC (%)	1	3	10	30	100
Regression	41.00	53.36	68.90	73.95	80.37
Classif.	35.79	57.27	68.57	74.35	79.79
Combined	41.56	59.13	69.18	75.01	81.22

TABLE III
ABLATION STUDY MEASURING THE EFFECT OF EACH COMPONENT.

IV. CONCLUSION

This work motivates the effectiveness of using domain-specific tasks to pre-train a DNN when the number of both labeled and unlabeled data is small. We designed an SSL algorithm addressing this lack of data based on ECG-delineation, training the network on auxiliary regression and classification tasks using interval- and axis-based features. Our experiments demonstrate that the proposed algorithm enhances the arrhythmia detection performance across various proportions of labeled training samples, and that the features learnt from one dataset properly transfers to another dataset. Moreover, the benefit of our SSL algorithm is shown for both a modern convolutional architecture and a linear layer attached to a deep feature extraction network, suggesting its potential strength across other various architectures. This result also verified the need to fine-tune the network after learning clinically-relevant features as opposed to using ECG delineation merely as a dimension reduction scheme.

This work could be extended by applying the proposed scheme to recurrent architectures or larger datasets. The fact that general purpose SSL algorithms outperformed a purely supervised learning algorithm insists that the algorithms enjoy benefits in various applications, but are not yet capable of achieving maximal performance attainable by learning domain-specific features. This work builds on top of recent interest of applying SSL to medical modalities where both labeled and unlabeled data are scarce [25], and further work comparing domain-specific algorithms with general-purpose SSL could stimulate research in both directions.

REFERENCES

[1] Zachi I Attia et al., "An artificial intelligence-enabled ecg algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction," *The Lancet*, vol. 394, no. 10201, pp. 861–867, 2019.

[2] Irene Fernández-Ruiz, "Artificial intelligence to improve the diagnosis of cardiovascular diseases," *Nature Reviews Cardiology*, vol. 16, no. 3, pp. 133–133, 2019.

[3] Awni Y Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, pp. 65, 2019.

[4] Philip Warrick and Masun Nabhan Homsi, "Cardiac arrhythmia detection from ecg combining convolutional and long short-term memory networks," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

[5] Jinwoo Cho et al., "Artificial intelligence algorithm for screening heart failure with reduced ejection fraction using electrocardiography.," *ASAIO Journal (American Society for Artificial Internal Organs: 1992)*, 2020.

[6] Spyros Gidaris, Praveer Singh, and Nikos Komodakis, "Unsupervised representation learning by predicting image rotations," in *International Conference on Learning Representations*, 2018.

[7] Mehdi Noroozi and Paolo Favaro, "Unsupervised learning of visual representations by solving jigsaw puzzles," in *European Conference on Computer Vision*. Springer, 2016, pp. 69–84.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[9] Alexander Kolesnikov, Xiaohua Zhai, and Lucas Beyer, "Revisiting self-supervised visual representation learning," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2019, pp. 1920–1929.

[10] Juan Pablo Martínez, Rute Almeida, Salvador Olmos, Ana Paula Rocha, and Pablo Laguna, "A wavelet-based ecg delineator: evaluation on standard databases," *IEEE Transactions on biomedical engineering*, vol. 51, no. 4, pp. 570–581, 2004.

[11] PN Singh and M Sajjad Athar, "Simplified calculation of mean qrs vector (mean electrical axis of heart) of electrocardiogram," *Indian journal of physiology and pharmacology*, vol. 47, no. 2, pp. 212–216, 2003.

[12] Cuiwei Li, Chongxun Zheng, and Changfeng Tai, "Detection of ecg characteristic points using wavelet transforms," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, pp. 21–28, 1995.

[13] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, 1992.

[14] Syed Anwar, Maheen Gul, Muhammad Majid, and Majdi Alnowami, "Arrhythmia classification of ecg signals using hybrid features," *Computational and Mathematical Methods in Medicine*, vol. 2018, pp. 1–8, 11 2018.

[15] Binhang Yuan and Wenhui Xing, *Diagnosing Cardiac Abnormalities from 12-Lead Electrocardiograms Using Enhanced Deep Convolutional Neural Networks*, pp. 36–44, 10 2019.

[16] Tomas B Garcia, *12-lead ECG: The art of interpretation*, Jones & Bartlett Publishers, 2013.

[17] Boris Hanin and David Rolnick, "How to start training: The effect of initialization and architecture," in *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., pp. 571–581. Curran Associates, Inc., 2018.

[18] K. He, R. Girshick, and P. Dollar, "Rethinking imagenet pre-training," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 4917–4926.

[19] Feifei Liu et al., "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368–1373, 2018.

[20] Patrick Wagner et al., "Ptb-xl, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, no. 1, pp. 1–15, 2020.

[21] 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, 12 2020.

[22] Jean-Bastien Grill et al., "Bootstrap your own latent: A new approach to self-supervised learning," 2020.

[23] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton, "A simple framework for contrastive learning of visual representations," 2020.

[24] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze, "Deep clustering for unsupervised learning of visual features," 2019.

[25] Hari Sowrirajan, Jingbo Yang, Andrew Y. Ng, and Pranav Rajpurkar, "Moco pretraining improves representation and transferability of chest x-ray models," 2020.