Adversarial Masking for Pretraining ECG Data Improves Downstream Model Generalizability

Jessica Y. Bo
Dept. of Medicine
Harvard University
Boston, MA 02115
jbo@mit.edu

Hen-Wei Huang
Dept. of Medicine
Harvard University
Boston, MA 02115
henwei@mit.edu

Alvin Chan
Dept. of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge MA, 02139
alvincgw@mit.edu

Giovanni Traverso
Dept. of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge MA, 02139
cgt20@mit.edu

Abstract

Medical datasets often face the problem of data scarcity, as ground truth labels must be generated by medical professionals. One mitigation strategy is to pretrain deep learning models on large, unlabelled datasets with self-supervised learning (SSL), but this introduces the issue of domain shift if the pretraining and task dataset distributions differ. Data augmentations are essential for improving the generalizability of SSL-pretrained models, but they tend to be either handcrafted or randomly applied. We use an adversarial model to generate masks as augmentations for 12-lead electrocardiogram (ECG) data, where masks learn to occlude diagnostically-relevant regions. Compared to random augmentations, models pretrained with adversarial masking reaches better accuracy under a domain shift condition and in data-scarce regimes on two diverse downstream tasks, arrhythmia classification and patient age estimation. Adversarial masking is competitive with and even reaches further improvements when combined with state-of-art ECG augmentation methods, *3KG* and *random lead masking (RLM)*, demonstrating the generalizability of our method[1].

1 Introduction

Across medical applications, deep learning is increasingly used to automate disease diagnosis [Miotto et al., 2018]. In some cases, neural networks have even reached or exceeded the performance of expert physicians [Hannun et al., 2019]. One such application is with 12-lead electrocardiogram (ECG) data, which is commonly collected to screen for various cardiovascular disorders [Fesmire et al., 1998]. There has been a recent surge in ECG-based deep learning research, largely enabled by challenges like the annual PhysioNet/Computing in Cardiology Challenge [Perez Alday et al., Reyna et al., 2021]. While these research outcomes show substantial progress within the field, their performance only reflects training on large-scale, labelled dataset. In contrast, real world medical datasets are likely much smaller due to the extensive resources required to collect medical labels.

[1]Our code is available at: https://github.com/jessica-bo/advmask_ecg

Workshop on Learning from Time Series for Health, 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Data scarcity is a well-documented issues that effect deep learning training. Models trained on small datasets lack generalizability to unseen data and cannot be deployed reliably [Kelly et al., 2019]. To mitigate issues associated with small training datasets, large yet unlabelled dataset can be leveraged to pretrain deep learning models with robust representations, which is commonly done in the ECG domain [Sarkar and Etemad, 2020, Weimann and Conrad, 2021, Liu et al., 2021, Kiyasseh et al., 2021, Diamant et al., 2022, Mehari and Strodthoff, 2022, Oh et al., 2022]. However, if the two datasets are collected under different environments, with different sensors, or across different populations, the transferrability of the model to the downstream task can be greatly impacted [Koh et al., 2021].

Contrastive self-supervised learning (SSL) is a pretraining technique that does not require a labelled dataset and can induce robust representations in the model [Chen et al., 2020a]. The encoding model is trained to maximize the similarity between latent representations of augmented pairs of data while separating from the representations of other data samples. SLL-pretrained models learn the underlying structure of the data invariant to the augmentations, which means that selecting good augmentations is crucial for learning useful representations. State-of-art within the ECG domain is patient-level SSL (PL-SSL), where pairs of data are taken from the same patient at different points in time [Kiyasseh et al., 2021, Diamant et al., 2022]. The performance of PL-SSL techniques are powerful, but can be improved in combination with other augmentations.

For time-series data, augmentations do not reach uniform and predictable gains across different datasets and tasks, making the augmentation selection process difficult [Iwana and Uchida, 2021]. Augmentations are typically selected from a standard pool like noise injection, baseline shift, time-domain masking, and more recent advances using generative adversarial models (GANs) to create synthetic data [Wen et al., 2021]. In the ECG data domain, a recent work *3KG* develops physiologically-consistent spatial augmentations for ECGs that reach good performance on small datasets [Gopal et al., 2021]. Another introduces *random lead masking (RLM)*, where leads of the ECG are fully masked out at random [Oh et al., 2022]. However, such augmentations are either handcrafted with manually tuned parameters or randomly applied, while we investigate the use of an adversarial method to optimize augmentation parameters.

In this work, we elect to focus on time-domain masking, a type of augmentation with high potential but is underexploited for periodic time-series data like ECGs:

- We adapt the image-based adversarial masking method of Shi et al. [2022] to generate masks as augmentations for each given ECG sample during PL-SSL pretraining. To our knowledge, this is the first work to implement adversarial masking for time-series data.
- We show that incorporating adversarial masking improves performance in two downstream tasks compared to baseline augmentations. We also find orthogonal benefits when combined with state-of-art ECG augmentations, *3KG* and *RLM*.

2 Methods

The overview of the adversarial masking pretraining framework with the encoding model $\boldsymbol{E}$ and adversarial masking model $\boldsymbol{A}$ and the downstream transfer learning step is show in Figure 1.

2.1 Self-Supervised Pretraining

CMSC Objective: We use Contrastive Multi-segment Coding (CMSC) as the PL-SSL strategy from Kiyasseh et al. [2021]. Given a batch of 12-lead ECG data $\left\{\boldsymbol{x} \in \mathbb{R}^D\right\}_{i=1}^{B}$, the positive pair is created as $\boldsymbol{x}_i = \boldsymbol{x}^{1 \cdot \cdot \frac{D}{2}}, \boldsymbol{x'}_i = \boldsymbol{x}^{\frac{D}{2} \cdot \cdot D}$, where $\boldsymbol{x}_i, \boldsymbol{x'}_i$ represent temporally non-overlapping ECG segments from the same patient. We further transform a data sample $T(\boldsymbol{x'}_i)$, where $\boldsymbol{T}$ represents one or more augmentations. The encoding model $\boldsymbol{E}$ is trained to align the feature representations of the positive pair of data, $\boldsymbol{h}_i = \boldsymbol{E}(\boldsymbol{x}_i)$ and $\boldsymbol{h'}_i = \boldsymbol{E}(T(\boldsymbol{x'}_i))$, while separating them from all other negative samples within the batch where $\boldsymbol{x}_i \neq \boldsymbol{x}_j$. We perform training with the CMSC objective (same as the SimCLR objective [Chen et al., 2020b]) described in Equation 1, where $\tau = 0.1$ is a temperature scaling term and the *sim* function represents cosine similarity.

$$\mathcal{L}_{\text{CMSC}}(\boldsymbol{x}; \boldsymbol{E}) = -\log \frac{\exp(sim(\boldsymbol{h}_i, \boldsymbol{h'}_i)/\tau)}{\sum_{i \neq j} \exp(sim\left(\boldsymbol{h}_i, \boldsymbol{h'}_j\right)/\tau)} \tag{1}$$

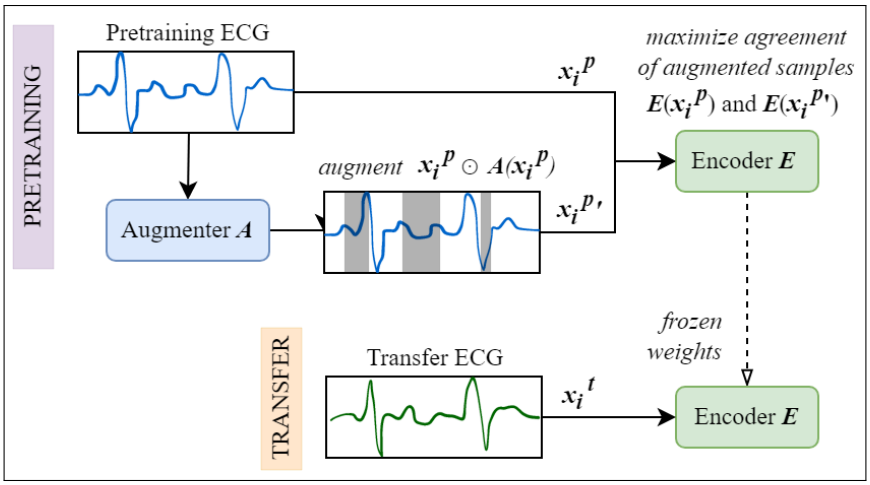

Figure 1: Adversarial masking pretraining and downstream transfer phases.

Adversarial Objective: We add an adversarial masking model $\boldsymbol{A}$ that generates a set of N masks for a given data sample, $\boldsymbol{m}_i \in \mathcal{R}^{N \times D} = \boldsymbol{A}(\boldsymbol{x}_i)$. Multiple masks ensure that the masked regions of the ECG are alternated, as only one mask is sampled, binarized, and applied in each training step. To overcome potential catastrophic forgetting due to overactive masking, we only apply the sampled mask with a probability of *p=0.8*. We use a differentiable soft binarization function to drive the mask values towards either 0 or 1, shown in Equation 2 with $\gamma = 25$. This enables back-propagation while simulating a shape-preserving binary mask without significant distortions to the ECG signals.

$$\boldsymbol{m}_{binarized} = \frac{1}{1 + \exp(-\gamma(\boldsymbol{m} - 0.5))} \tag{2}$$

The masking model $\boldsymbol{A}$ acts in opposition to the encoding model $\boldsymbol{E}$ by generating difficult augmentations, hence maximizing the CMSC objective. We also adapt the sparse penalty from Shi et al. [2022] to limit the amount of masking, described by Equation 3 with $\alpha = 0.1$ as a weighting term. This prevents $\boldsymbol{A}$ from fully occluding the data or not occluding anything at all. The full *min-max* loss objective optimizing $\boldsymbol{E}$ and $\boldsymbol{A}$ is described by Equation 4:

$$\mathcal{L}_{\text{sparse}}(\boldsymbol{x}; \boldsymbol{A}) = \sin\left(\frac{\pi}{D}\sum_{d=1}^{D} \boldsymbol{m}^d\right)^{-1} \tag{3}$$

$$\min_{\boldsymbol{E}} \max_{\boldsymbol{A}} \mathcal{L}_{\text{CMSC}}(\boldsymbol{E}, \boldsymbol{A}) - \mathcal{L}_{\text{sparse}}(\boldsymbol{A}) \tag{4}$$

Architectures: For the encoder backbone, we use a 1D ResNet-18 with a hidden dimension of 512 [He et al., 2016], which is the best performing architecture across multiple ECG tasks out of eight backbones [Nonaka and Seita, 2021]. We also use a two-layer projection head typical in contrastive SSL to convert the encoder's outputs to a 128-dimension space, but the projector is not transferred to the downstream task.

For the adversarial model, we adapt an open-source 1D U-Net with four downsampling and upsampling layers. The number of out channels is equal to the number of masks *N*. The outputs are passed through either a *softmax* or *sigmoid* function depending on *N*.

Pretraining Details: All models are implemented with PyTorch Lightning. Training and testing are performed with a single NVIDIA Volta V100 GPU on the MIT Supercloud [Reuther et al., 2018]. Pretraining is performed with 12-lead ECG datasets from the PhysioNet/Computing in Cardiology Challenge 2020 [Perez Alday et al.], which comprises over 66 thousand patient recordings from six institutions in four countries.

Table 1: Downstream performance for two tasks across multiple training dataset sizes (denoted as *X% DS*) reported for encoders pretrained on multiple augmentations and a *Scratch* baseline.

	Arrhythmia Classification (accuracy %)			Age Regression (absolute error)		
	100% DS	*10% DS*	*1% DS*	*100% DS*	*10% DS*	*1% DS*
Scratch	*70.99 ± 1.59*	*63.92 ± 1.97*	*34.89 ± 0.54*	*14.56 ± 0.38*	*15.24 ± 0.48*	*25.98 ± 1.09*
Gaussian	79.82 ± 0.48	71.06 ± 0.98	59.82 ± 0.29	18.79 ± 0.31	20.36 ± 1.41	39.72 ± 3.2
Powerline	79.07 ± 2.07	69.84 ± 1.88	58.73 ± 1.04	19.51 ± 0.52	20.0 ± 0.36	38.96 ± 1.43
STFT	79.6 ± 1.37	70.62 ± 1.01	60.92 ± 1.32	19.11 ± 0.17	20.03 ± 0.74	39.02 ± 0.41
Wander	79.41 ± 0.65	71.12 ± 0.33	59.61 ± 1.37	18.76 ± 0.54	20.08 ± 0.88	38.81 ± 3.07
Shift	80.44 ± 1.58	70.74 ± 1.12	60.04 ± 2.26	19.32 ± 0.38	19.66 ± 0.68	38.03 ± 1.97
Mask	83.51 ± 0.83	75.44 ± 1.99	62.95 ± 1.53	18.77 ± 0.39	19.39 ± 0.28	41.81 ± 0.48
Blockmask	84.01 ± 2.31	74.22 ± 0.79	64.61 ± 1.14	18.79 ± 0.10	19.07 ± 0.31	40.85 ± 0.37
Peakmask	88.58 ± 0.22	82.92 ± 1.21	69.02 ± 0.92	17.63 ± 0.32	20.1 ± 0.76	42.48 ± 0.80
Adv Mask (AM)	89.86 ± 0.38	86.33 ± 1.14	76.57 ± 1.29	17.04 ± 0.28	18.64 ± 0.68	35.3 ± 3.03
3KG	91.55 ± 0.93	86.2 ± 1.05	75.72 ± 1.03	17.01 ± 0.52	**18.06 ± 0.46**	42.28 ± 0.36
AM + 3KG	**93.08 ± 0.82**	**90.93 ± 0.81**	82.38 ± 0.50	**16.71 ± 0.40**	18.81 ± 0.69	**33.13 ± 2.69**
RLM	90.58 ± 0.12	87.48 ± 0.56	76.03 ± 1.16	16.85 ± 0.48	20.09 ± 0.87	46.28 ± 0.25
AM + RLM	92.8 ± 0.82	90.74 ± 0.54	**83.32 ± 1.65**	17.21 ± 0.59	18.79 ± 0.99	33.37 ± 1.68

We use a learning rate of 0.0001 and an *Adam* optimizer for both the encoding model and adversarial model. The models are updated one after another with their respective losses within the same batch. The batch size is 32 with gradient accumulation over 4 batches, which is equivalent to an effective batch size of 128. To reduce computation efforts, we also use mixed precision training. An early stopping condition based on minimizing $\mathcal{L}_{\text{CMSC}}$ is implemented to reduce the training time.

2.2 Downstream Transfer Learning

We transfer the encoding models's learned representations to two downstream classification tasks by training a linear layer on top of the frozen model weights. The tasks, arrhythmia classification and age regression, are used as benchmarks by Diamant et al. [2022]. The former correlates to disease characteristics and the latter to patient identity, representing two diverse objectives. The downstream transfer dataset is the 12-lead Chapman-Shaoxing ECG dataset, which has over 10 thousand patients and four major classes of cardiac rhythm labels [Zheng et al., 2020]. The arrhythymia class balance is 36% atrial fbrillation (AFIB), 22% supraventricular tachycardia (GSVT), 21% sinus bradycardia (SB), and 21% sinus rhythm (SR). Patient ages range from 4 to 98 years.

Transfer learning is performed with a batch size of 256, learning rate of 0.01, and an *Adam* optimizer. A single *Linear* layer projects the output dimension of the encoding model to the number of output classes. There are no overlaps between the pretraining and downstream datasets, so our transfer learning setup constitutes a *domain shift*.

3 Results

3.1 Transfer Task Performance

We present transfer learning results using a train-validation-test split of 80%, 10% and 10%. We simulate real world data scarcity conditions by reducing the transfer training dataset to 100%, 10%, and 1% of the original size (8516 samples). The size of the test set is held constant and the test accuracy for all experiments is reported as the average across 3 random seeds.

Table 1 shows the performance metrics and standard deviation the tasks with models pretrained on all baseline and adversarial augmentations. The baseline augmentations are adapted from literature and details are available in Appendix A. Classification accuracy is reported for arrhythmia classification (higher is better) and the absolute error is reported for age regression (lower is better). For each task at each dataset size reduction, the best performing augmentation method for pretraining are in bold. Results from a finetuned *Scratch* baseline is also reported. Adversarial masking (**Adv Mask**) results are given with $N = 2$, which performs the best.

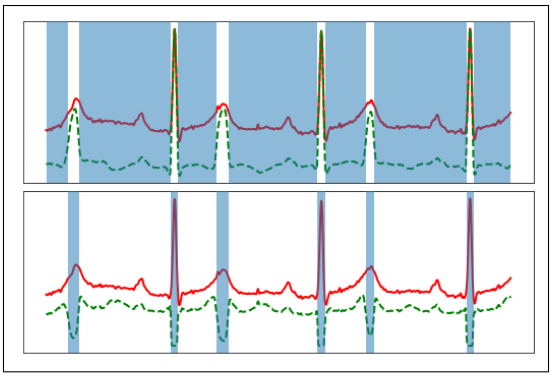

Figure 2: N=2 adversarially generated masks overlaying Lead II of the ECG sample.

For both downstream tasks, **Adv Mask** yields superior results to most baseline augmentations across multiple training dataset sizes, including other masking techniques. It is competitive to but does not consistently outperform *3KG* [Gopal et al., 2021] and *RLM* [Oh et al., 2022], two methods which are specific to ECGs and encourages learning invariances across leads. However, **Adv Mask** combined with *3KG* and *RLM* shows significant improvements over the respective augmentations on their own, with either **Adv Mask + 3KG** or **Adv Mask + RLM** reaching the best performance in nearly all transfer trials. This shows that benefits introduced by **Adv Mask** is orthogonal to other augmentations and consistent data-scarce and domain-shifted conditions.

3.2 Analysis of Augmentations

We visualize the generated masks in Figure 2, with more examples in Appendix B. Lead II of the ECG is displayed in red and the masked signal in green. Since we use a soft binarization method, the mask can take any value between 0 and 1. The blue overlay represent regions that are heavily driven to dropout. The shape of the original ECG is well-preserved outside of the masked regions, but may be augmented with a slight shift and scaling factor. In an analysis of the effect of data augmentation on model generalization, Huang et al. [2021] show richer augmentations lead to higher generalizability in downstream tasks, which supports why **Adv Mask** is effective.

With $N = 2$, one mask consistently covers the QRS complex and T-wave, while the other mask covers the remaining areas. We hypothesize this is similar to capturing diagnostically-relevant “semantic" content of the ECG, in the same way that Shi et al. [2022] demonstrates adversarial image masks cover semantically-coherent regions of the image. According to an analysis of the salient regions of ECG data, the QRS complex often carry high levels of disease-diagnostic information [Jones et al., 2020]. This suggests that the encoding model emphasizes learning the structual information of the highly salient regions of the ECGs during pretraining. While the results may indicate that learnt masks converge to peaks, we show that **Adv Mask** outperforms the *Peakmask* baseline, which uses statistical methods to mask out peaks in as a naive replication of **Adv Mask**’s behaviour.

4 Conclusions and Future Work

Adversarial masking for ECG data can be utilized positively in SSL pretraining of deep learning models. It induces the learning of generalizable representations for downstream transfer learning in highly data-scarce and domain-shifted tasks. Furthermore, masking as an augmentation scheme is agnostic to the choice of architecture and training scheme, so the method can be extended to other SSL frameworks beyond SimCLR, and even to other time-series data modalities.

Future work includes evaluating across more datasets to understand where adversarial masking can bring the most benefit. To strengthen our results, we would also like to extend our experiments to include generative models as an augmentation baseline and tailor the sparsity penalty to leverage specific properties of time-series data.

References

Tianlong Chen, Sijia Liu, Shiyu Chang, Yu Cheng, Lisa Amini, and Zhangyang Wang. Adversarial Robustness: From Self-Supervised Pre-Training to Fine-Tuning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020*, pages 699–708, 2020a. URL https://proceedings.mlr.press/v149/nonaka21a.html.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the 37th International Conference on Machine Learning*, pages 1597–1607. PMLR, 2020b. URL https://proceedings.mlr.press/v119/chen20j.html.

Nathaniel Diamant, Erik Reinertsen, Steven Song, Aaron D. Aguirre, Collin M. Stultz, and Puneet Batra. Patient contrastive learning: A performant, expressive, and practical approach to electrocardiogram modeling. *PLoS Computational Biology*, 18(2):1–16, 2022.

F. M. Fesmire, R. F. Percy, J. B. Bardoner, D. R. Wharton, and F. B. Calhoun. Usefulness of Automated Serial 12-Lead ECG Monitoring During the Initial Emergency Department Evaluation of Patients With Chest Pain. *Annals of Emergency Medicine*, 31(1):3–11, 1998. ISSN 0196-0644. doi: 10.1016/S0196-0644(98)70274-4.

Bryan Gopal, Ryan Han, Gautham Raghupathi, Andrew Ng, Geoff Tison, and Pranav Rajpurkar. 3KG: Contrastive Learning of 12-Lead Electrocardiograms using Physiologically-Inspired Augmentations. *Proceedings of Machine Learning Research*, 158:156–167, 2021. URL https://proceedings.mlr.press/v158/gopal21a.html.

Awni Y. Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H. Tison, Codie Bourn, Mintu P. Turakhia, and Andrew Y. Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine 2019 25:1*, 25(1): 65–69, 2019.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.

Weiran Huang, Mingyang Yi, and Xuyang Zhao. Towards the Generalization of Contrastive Self-Supervised Learning. 11 2021. doi: 10.48550/arxiv.2111.00743. URL http://arxiv.org/abs/2111.00743.

Brian Kenji Iwana and Seiichi Uchida. An empirical survey of data augmentation for time series classification with neural networks. *PLoS ONE*, 16(7 July):1–43, 2021. ISSN 19326203. doi: 10.1371/journal.pone.0254841.

Yola Jones, Fani Deligianni, and Jeff Dalton. Improving ECG Classification Interpretability using Saliency Maps. In *IEEE 20th International Conference on Bioinformatics and Bioengineering*, pages 675–682. Institute of Electrical and Electronics Engineers Inc., 2020. doi: 10.1109/BIBE50027.2020.00114.

Christopher J. Kelly, Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(195): 1–9, 2019. doi: 10.1186/S12916-019-1426-2.

Dani Kiyasseh, Tingting Zhu, and David A Clifton. CLOCS: Contrastive Learning of Cardiac Signals Across Space, Time, and Patients. In *Proceedings of the 38th International Conference on Machine Learning, PMLR*, pages 5606–5615. PMLR, 2021. URL https://proceedings.mlr.press/v139/kiyasseh21a.html.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning*, pages 5637–5664. PMLR, 2021.

Han Liu, Zhenbo Zhao, and Qiang She. Self-supervised ECG pre-training. *Biomedical Signal Processing and Control*, 70:103010, 2021. doi: 10.1016/J.BSPC.2021.103010.

Temesgen Mehari and Nils Strodthoff. Self-supervised representation learning from 12-lead ECG data. *Computers in Biology and Medicine*, 141:105114, 2022. doi: 10.1016/j.compbiomed.2021.105114.

Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T. Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6):1236–1246, 2018. doi: 10.1093/BIB/BBX044.

Naoki Nonaka and Jun Seita. In-depth Benchmarking of Deep Neural Network Architectures for ECG Diagnosis. In *6th Machine Learning for Healthcare Conference*, volume 149 of *Proceedings of Machine Learning Research*, pages 414–439. PMLR, 2021. URL `https://proceedings.mlr.press/v149/nonaka21a.html`.

Jungwoo Oh, Hyunseung Chung, Joon-myoung Kwon, Dong-gyun Hong, and Edward Choi. Lead-agnostic Self-supervised Learning for Local and Global Representations of Electrocardiogram. In *Proceedings of the Conference on Health, Inference, and Learning*, volume 174 of *Proceedings of Machine Learning Research*, pages 338–353. PMLR, 2022. URL `https://proceedings.mlr.press/v174/oh22a.html`.

Erick A Perez Alday, Annie Gu, Amit Shah, Chad Robichaux, An Kwok, Ian Wong, Chengyu Liu, Feifei Liu, Ali Bahrami Rad, Andoni Elola, Salman Seyedi, Qiao Li, Ashish Sharma, Gari D Clifford, and Matthew A Reyna. Classification of 12-lead ECGs: the PhysioNet/ Computing in Cardiology Challenge 2020.

Albert Reuther, Jeremy Kepner, Chansup Byun, Siddharth Samsi, William Arcand, David Bestor, Bill Bergeron, Vijay Gadepally, Michael Houle, Matthew Hubbell, Michael Jones, Anna Klein, Lauren Milechin, Julia Mullen, Andrew Prout, Antonio Rosa, Charles Yee, and Peter Michaleas. Interactive supercomputing on 40,000 cores for machine learning and data analysis. In *2018 IEEE High Performance extreme Computing Conference (HPEC)*, pages 1–6. IEEE, 2018.

Matthew A. Reyna, Nadi Sadr, Erick A.Perez Alday, Annie Gu, Amit J. Shah, Chad Robichaux, Ali Bahrami Rad, Andoni Elola, Salman Seyedi, Sardar Ansari, Hamid Ghanbari, Qiao Li, Ashish Sharma, and Gari D. Clifford. Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021. *Computing in Cardiology*, 48:1–4, 2021. doi: 10.23919/CINC53138.2021.9662687.

Pritam Sarkar and Ali Etemad. Self-supervised ECG Representation Learning for Emotion Recognition. *IEEE Transactions on Affective Computing*, pages 1–13, 2020. doi: 10.1109/TAFFC.2020.3014842.

Yuge Shi, N Siddharth, Philip H S Torr, and Adam R Kosiorek. Adversarial Masking for Self-Supervised Learning. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 20026–20040. PMLR, 2022. URL `https://proceedings.mlr.press/v162/shi22d.html`.

Kuba Weimann and Tim O.F. Conrad. Transfer learning for ECG classification. *Scientific Reports*, 11 (1):1–12, 2021. doi: 10.1038/s41598-021-84374-8.

Qingsong Wen, Liang Sun, Fan Yang, Xiaomin Song, Jingkun Gao, Xue Wang, and Huan Xu. Time Series Data Augmentation for Deep Learning: A Survey. *IJCAI International Joint Conference on Artificial Intelligence*, pages 4653–4660, 2021. doi: 10.24963/ijcai.2021/631.

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*, 7(1), 2020. doi: 10.1038/s41597-020-0386-x.

A Augmentations

The baseline augmentations are sourced from ECG deep learning research. While the list is not exhaustive, we aim to be representative of typical random ECG augmentation techniques used in existing works. Note that a limitation of our results is that we only test with isolated augmentations, whereas combining two or more augmentations may intuitively yield better results. We also do not have generative models (GANs) as baseline, but we aim to include them in the future.

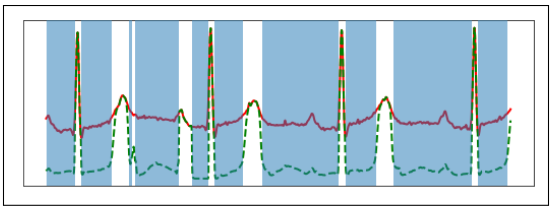

Figure 3: *N*=1 adversarially generated masks overlaying Lead II of the ECG sample.

Gaussian Noise: A vector of noise $v(t) \sim \mathcal{N}(0, 0.05)$ is sampled and added to each lead of the ECG. Gaussian noise injection is a very common augmentation used in SSL training for both time-series and image data.

Powerline Noise: Powerline noise $n(t) = \alpha cos(2\pi t k f_p + \phi)$, with $\alpha \sim \mathcal{U}(0, 0.5)$, $\phi \sim \mathcal{N}(0, 2\pi)$, and $f_p = 50Hz$ is added to each lead of the ECG [Mehari and Strodthoff, 2022, Oh et al., 2022].

Short-time Fourier Transform: STFT involves computing the Fourier transform of short segments of the time-series signal to generate a spectrogram. A random mask with values sampled from a beta distribution $B(\alpha = 5, \beta = 2)$ is applied to the spectrogram. Finally, the STFT operation is inversed to recover the time domain signal.

Baseline Wander: The ECG signal is perturbed with a very low frequency signal to simulate drifting: $n(t) = C \sum_{k=1}^{K} a \cos(2\pi t k \Delta f + \phi)$, with $C \sim \mathcal{N}(1, 0.5^2)$, $\alpha \sim \mathcal{U}(0, 0.5)$, $\phi \sim \mathcal{N}(0, 2\pi)$, and $\Delta f \sim \mathcal{U}(0.01, 0.2)$ [Mehari and Strodthoff, 2022, Oh et al., 2022].

Baseline Shift: A fraction $p = 0.2$ of the baseline of each lead of the ECG signal is shifted by a factor of $\alpha \sim \mathcal{N}(-0.5, 0.5)$ [Mehari and Strodthoff, 2022, Oh et al., 2022].

Mask: Two types of random masking are implemented, where Mask refers to any timestep being masked out to 0 with the probability $p = 0.2$.

Blockmask: Blockmask masks out a continuous portion $p = 0.2$ of each lead to 0. The main difference is that Blockmask would occlude larger structural regions of the ECG, whereas Mask only occludes local details.

Peakmask: We introduce Peakmask as a statistical baseline that aims to replicate the peak-finding behaviour of the adversarially generated masks. Timesteps where the average value across leads exceeds the average value over the entire ECG are masked. We either sample the computed mask directly, or the opposite regions (non-peak areas). This emulates the **Adv Mask** technique with N=2.

3KG: Developed by Gopal et al. [2021], 3KG augments the 3D spatial representation of the ECG in vectorcardiogram (VCG) space and reprojects it back into ECG space, mimicking natural variations in cardiac structure and orientation. We take the best parameters reported in the paper, a random rotation $-45° \leq \theta \leq 45°$ and a scaling factor $1 \leq s \leq 1.5$.

RLM: Introduced by Oh et al. [2022], RLM fully occludes individual ECG leads with probability *p=0.5*, which was the parameter used in their paper. This reduces the dependency on requiring all 12 leads to extract useful information.

B More Adversarial Masks

Figures 3-5 depict more examples of adversarially generated masks with $N = 1, 3, 12$. When $N < 12$, one mask is randomly sampled and applied to all leads of the ECG. When $N = 12$, each mask is applied separately to each lead. As mentioned in the paper, $N \neq 2$ trials saw reduced performance so their results are not reported. Increasing *N* does not seem to be helpful as masks tend to cover either peaks or flat regions, which can be achieved with *N*=2.

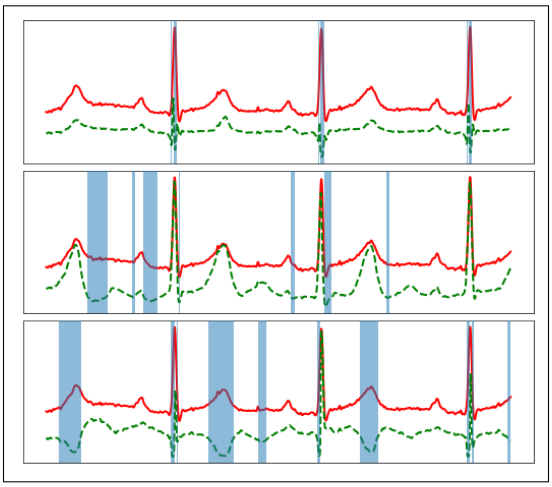

Figure 4: N=3 adversarially generated masks overlaying Lead II of the ECG sample.

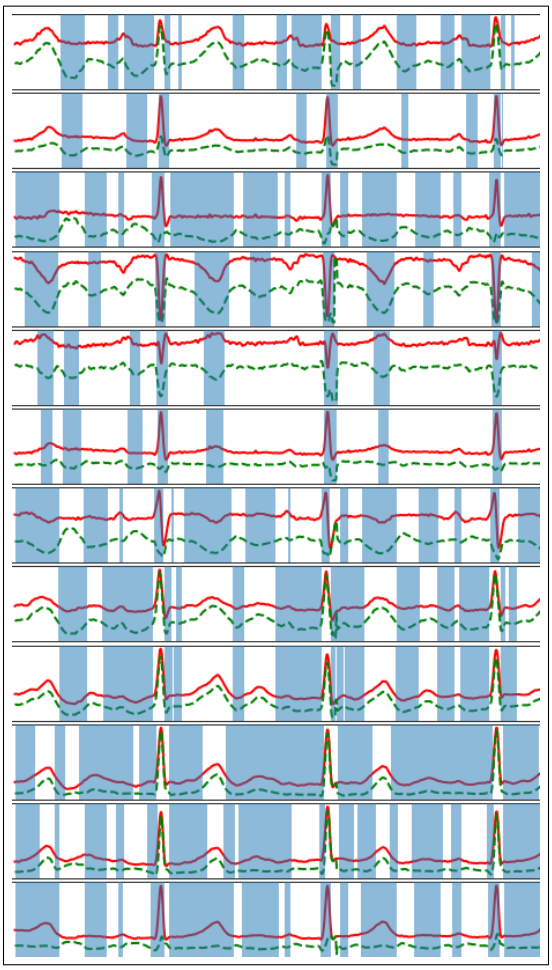

Figure 5: N=12 adversarially generated masks overlaying each lead of the ECG sample.