SILICONET : A SIAMESE LEAD INVARIANT CONVOLUTIONAL NETWORK FOR VENTRICULAR HEARTBEAT DETECTION IN ELECTROCARDIOGRAMS (ECG)

Pierre Aublin[1], **Jacques Felblinger**[1,2], **Julien Oster**[1,2]
1. IADI, Inserm U1254
2. CIC-IT 1433, Université de Lorraine, Inserm, CHRU de Nancy
Bâtiment Recherche, CHRU Nancy Brabois,
Rue du Morvan, 54500 Vandœuvre-lès-Nancy, France
pierre.aublin, jacques.felblinger, julien.oster@inserm.fr

ABSTRACT

Pretraining deep learning models on a large corpus of unlabelled data using self supervised learning approaches can be an efficient mitigation strategy to deal with the lack of annotated data. We proposed to use a siamese framework for the pretraining of a convolutional neural network on the Computing in Cardiology 2021 dataset therefore making it invariant to ECG lead configuration changes. The obtained representation was then trained and tested on a heartbeat classification task on the MIT BIH Arrhythmia database, and on an external independent set, namely the INCART database. The proposed model reached a median F1 score of 0.89 on the MIT BIH Arrhythmia database comparable to the 0.90 F1 score obtained without pretraining. However, the pretrained model obtained a median F1 score of 0.74 on average over the different leads, compared to 0.53 the model without pretraining. The proposed pretraining approach, leveraging the availability of relatively large database of un-(or weakly)annotated ECG data, allows for the training of more generalisable, lead-agnostic, heartbeat classification models. Such an approach would ensure avoiding overfitting complex deep learning models on the small MIT-BIH arrhythmia database.

1 INTRODUCTION

Electrocardiogram (ECG) signals represent the electrical activity of the heart, and are used routinely by cardiologist and clinicians to diagnose cardiovascular diseases. A ventricular heartbeat (V) (cf. Figure 1), is an abnormal heart contraction, resulting from the instantaneous depolarisation of ventricle cells. Frequent V heartbeats (Al-Khatib et al., 2018) are related to other types of cardiovascular diseases and monitoring them can be useful for therapeutic aims (Lustgarten et al., 2020). These heartbeats can be visually identified on the ECG signal, and several automated methods have been developed (Chazal et al., 2004). More recently, deep learning approaches were tested for heartbeat classification (Zhang et al., 2021). However, there is a lack of large fully annotated data. The most popular widely available database, MIT BIH Arrhythmia database (Moody & Mark, 2001), contains only 48 recordings of 30-minute long 2 leads ECG signals. Therefore, even if this database includes tens of thousands of heartbeats, the recordings only provide a very limited amount of different heartbeat morphologies as they are sourced form few different subjects. Another limitation of the MIT-BIH arrhythmia database (Moody & Mark, 2001; Hong et al., 2020), apart from containing few patients, is that it contains only two-lead ECG signals, which means that the models trained on this dataset may not generalise well on other lead configurations. Self-supervised learning may be a way to circumvent this lack of labeled data. Among self-supervised approaches, siamese networks consist in sharing the weights of neural networks provided with several different views of a given input (Chen & He, 2021). By doing so, the networks are trained to become invariant to the transformations between these views. It has been shown that the 12 ECG leads are not equally important to detect different heart diseases (Green et al., 2007) and ECG classifiers can be affected by the lead positions and configurations (Reyna et al., 2021; 2022). Several approaches have been attempted to develop lead or patient invariant data representations, and have been successfully ap-

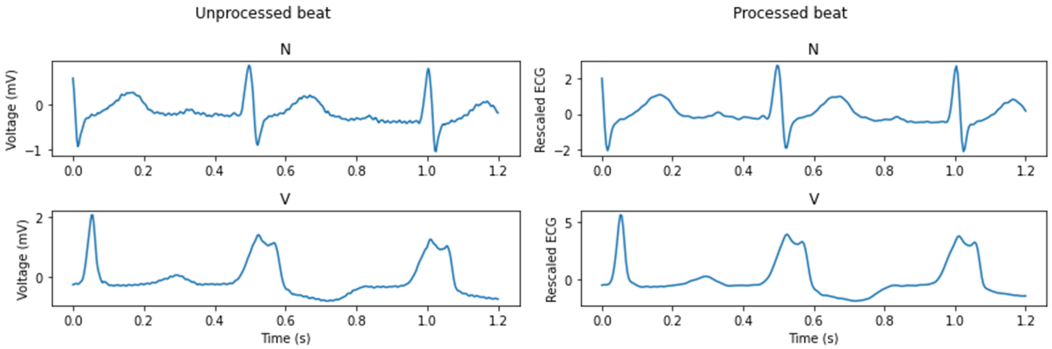

Figure 1: Illustration of N and V heartbeats before and after preprocessing.

plied for heart disease detection (Kiyasseh et al., 2021; Oh et al., 2022). However, these approaches were considering a whole ECG recording and not analysing individual heartbeats.

Our contributions are: (i) We pretrained a deep neural network following a siamese scheme on a large ECG database to be invariant to ECG lead configurations; (ii) We used the obtained data representation to develop a heartbeat classifier on a scarce database and compared it to a randomly initialized neural network; (iii) We improved the generalisability of this heartbeat classifier, and robustness to ECG lead configuration changes.

2 METHODS

2.1 DATASETS

The PhysioNet Computing in Cardiology 2021 challenge dataset (CinC) (Reyna et al., 2022) was used for pre-training. ECG signals were provided with a SNOMED-CT code diagnosis for each recording, but no individual heartbeat annotation was provided. For monitoring purpose during training, the CinC dataset was randomly split into a training set containing 80% of the recordings while 20% of the recordings were kept for validation. Recordings from the INCART database (Tihonenko et al., 2008) initially included in the CinC dataset were removed (to preserve independence of the testing database), resulting in a pretraining dataset including 88179 ECG recordings. A heartbeat classification (normal beats N vs V beats) task was performed as the downstream evaluation. The MIT-BIH arrhythmia database (Moody & Mark, 2001) (MITAR), containing 48 30-minute recordings of 2-lead ECGs with individual heartbeat annotations was used for validation and testing. The data was split in two sets (DS1/DS2) following (Chazal et al., 2004), ensuring patient stratification between sets. During training, DS1 was split into 50% train, and 50% validation sets, stratified by heartbeat class. The generalisability of the model and the impact of lead positions were assessed on the INCART database (Tihonenko et al., 2008), containing 12-lead ECG recordings from 74 subjects with individual heartbeat annotations.

2.2 ARCHITECTURE

The proposed architecture, inspired by (Vogt, 2019; Aublin et al., 2022), consists in an 11-layer CNN detailed in Table 3 of the appendix A, followed by a global max pooling layer leading to 256 features. A projection head and a predictor were added following (Chen et al., 2020; Chen & He, 2021) for the pretraining purpose. For the downstream task, the projection head and predictor were removed and replaced by a single linear layer followed by a softmax activation function.

2.3 PREPROCESSING

Preprocessing of the input consisted in band-pass filtering between 0.5 and 60 Hz, and resampling to 250Hz. As no heartbeat annotations are available on the CinC Dataset, a QRS detector (Pan & Tompkins, 1985) was run to get an estimate of QRS positions. On MITAR, QRS positions were obtained from the manual annotations. For each heartbeat, a time-window of 1s before the QRS position, and 0.2s after was extracted, resulting in a 301-sample input vector.
The transformation used for the pretraining task consisted in inputing 2 different (randomly selected) views of each heartbeat. For the downstream task on MITAR, analysis was only performed on the

Name	Pretrained weights	Train Embedding	Initial learning rate (Embedding)	Initial learning rate (Classifier)
randinit	no	yes	1.10^{-3}	1.10^{-3}
randfeatures	no	no	-	1.10^{-3}
pretrained	yes	yes	1.10^{-3}	1.10^{-3}
finetune	yes	no	-	1.10^{-3}
2lr	yes	yes	1.10^{-5}	1.10^{-2}

Table 1: Training configurations for the heartbeat classifier

		randfeatures	randinit	pretrained	finetune	2lr
DS2	MLII Median	0.49	**0.90**	0.87	0.65	0.89
	MLII Q1	0.45	**0.89**	0.85	0.65	0.87
	MLII Q3	0.51	0.90	0.87	0.66	**0.91**
INCART	Leads median	0.31	0.53	0.64	**0.74**	0.72
	Leads Q1	0.21	0.28	0.52	**0.72**	0.62
	Leads Q3	0.38	0.68	0.74	**0.77**	**0.77**

Table 2: Leads results for the different training configurations on DS2 and INCART

first lead (generally modified lead II). Normalisation was also performed globally with the mean and standard deviation extracted on the training set of DS1.
Generalisation of classifier performance was tested independently on the 12 available leads of the INCART database.

2.4 TRAINING

i) Pretraining step. A negative cosine similarity loss with $StopGradient$ (Chen & He, 2021) was chosen, with monitoring of the validation loss. Adam algorithm with weight decay decoupling (AdamW) (Loshchilov & Hutter, 2019) was used with a stepwise decay schedule learning rate. The base learning rate was set to 1.10^{-3} and multiplied by 0.75 every 10 epochs.
ii) Downstream task step. Cross-entropy loss was chosen and F1 score was monitored on the validation set. Five configurations were tested and are summarized in Table 1. AdamW optimizer was also used but with a base learning rate of 10^{-3} and a cosine decay schedule until 8.10^{-4} over 50 epochs, except for the 2lr configuration where the network parameter optimization were split into two parameter groups : the CNN embedding with a base learning rate of 1.10^{-6} and the classifier with a base learning rate of 10^{-2}. Both groups had a cosine learning rate scheduling over 50 epochs until a rate of 1.10^{-3}.
Early stopping (if no improvement in the monitored metric was observed after 20 epochs) was implemented in both steps.
Optimiser details are summarized in Appendix A.

3 RESULTS

F1 scores (median, first and third quartile) obtained on DS2 and INCART are assembled in table 2. Lead by lead analysis is given in table 4 and depicted in figure 5 in appendix B.
Similar results, with no statistical significant difference (Wilcoxon test), were obtained on DS2 for the randinit, pretrained and 2lr approaches. Pretrained models (finetune, 2lr) however outperform other approaches on the INCART database. The finetune technique outperforms all other approaches especially on precordial leads V1, V2, V3 where performance for all other techniques show a significant drop (cf table 4).
Pretrained models also show significantly lower F1 variability over multiple runs (standard deviation 0.02 for pretrained and 2lr, 0.01 for finetune) than other approaches (0.06 for randfeatures and 0.04 for randinit).

Figure 2 depicts the F1 score per lead and model. The cluster of lower F1 score on leads V1 to V3 for 2lr, pretrained, and randinit techniques is clearly visible.

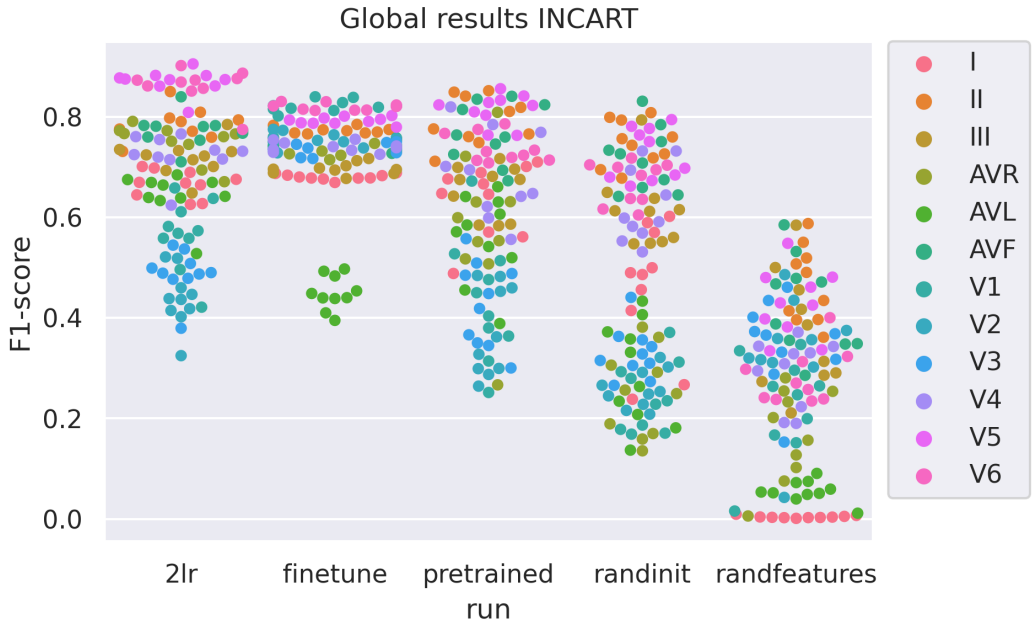

Figure 2: F1 scores swarmplot for the INCART database.

4 Discussion

Comparison between the randfeatures and the finetune results seems to be suggesting that the pretraining effictively learns a transferable representation of the ECG signal.
The result variability indicates that despite pretraining overfitting on the DS1 database is still occurring. This may be showing that the DS1 dataset is too small for training of a DL model and indicates the need for the release of a larger database of fully annotated ECG signals. On the other hand, training only the final linear layer (finetune) seems not to be flexible enough for high performance on DS2. This motivated the 2lr approach, allowing slight changes on the CNN feature extractor, to finely adapt the representation for the downstream task.
To overcome the limited size of MITAR database, even though many recently proposed DL solutions were trained on it (Hong et al., 2020), other larger ECG databases were also released but most often contain annotation from automated software, which implies that training DL on such database might only replicate performance from currently available medical software.
As demonstrated by the pefomance of pretrained models on the INCART database, pretraining improves the robustness of the classifier to different lead configurations, which is consistent with the capacity of learning invariance when using siamese networks (Chen & He, 2021).
In this study we have only tested the SimSiam pretraining approach so far. Future studies will include the evaluation of other self-supervised techniques such as BYOL, contrastive learning (CLOCS, Lead-agnostic) (Kiyasseh et al., 2021; Oh et al., 2022).
It may become possible to skip certain denoising steps, by applying representation learning to be invariant to a family of transformations mimicking the presence of physiological noise. This may be highly beneficial for situations where robust and efficient noise suppression is currently not possible. A good example of such a case consists of ECG signals acquired during magnetic resonance imaging examinations, where magnetohydrodynamic (MHD) effect distorts ECG signals and accurate MHD suppression remains challenging (Oster et al., 2014; Oster & Clifford, 2017).

5 Conclusion

In this study, we discussed the generalization difficulties of automated heartbeat classifiers developed with the MIT BIH arrhythmia database, and illustrated how classifier performance can be affected by different lead configurations. We addressed that issue by applying a Siamese neural network to develop lead invariant feature extractors, rendering subsequent automated heartbeat classifiers less sensible to lead configurations.

ACKNOWLEDGMENTS

PA and JO were supported by a grant from the ERA-CVD Joint Translational Call 2019, MEIDIC-VTACH (ANR-19-ECVD-0004). We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU partly used for this research.

REFERENCES

Sana M. Al-Khatib, William G. Stevenson, Michael J. Ackerman, William J. Bryant, David J. Callans, Anne B. Curtis, Barbara J. Deal, Timm Dickfeld, Michael E. Field, Gregg C. Fonarow, Anne M. Gillis, Christopher B. Granger, Stephen C. Hammill, Mark A. Hlatky, José A. Joglar, G. Neal Kay, Daniel D. Matlock, Robert J. Myerburg, and Richard L. Page. 2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death: Executive Summary. *Circulation*, 138(13):e210–e271, 2018. doi: 10.1161/CIR.0000000000000548. URL https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000000548. _eprint: https://www.ahajournals.org/doi/pdf/10.1161/CIR.0000000000000548.

Pierre Gabriel Aublin, Mouin Ben Ammar, Jérémy Fix, Michel Barret, Joachim Abraham Behar, and Julien Oster. Predict alone, decide together: cardiac abnormality detection based on single lead classifier voting. *Physiological Measurement*, 2022. URL http://iopscience.iop.org/article/10.1088/1361-6579/ac66b9.

Philip de Chazal, M. O'Dwyer, and R.B. Reilly. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51 (7):1196–1206, 2004. doi: 10.1109/TBME.2004.827359.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations, June 2020. URL http://arxiv.org/abs/2002.05709. arXiv:2002.05709 [cs, stat].

Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15750–15758, 2021.

Michael Green, Mattias Ohlsson, Jakob Lundager Forberg, Jonas Björk, Lars Edenbrandt, and Ulf Ekelund. Best leads in the standard electrocardiogram for the emergency detection of acute coronary syndrome. *Journal of Electrocardiology*, 40(3):251–256, 2007. ISSN 0022-0736. doi: https://doi.org/10.1016/j.jelectrocard.2006.12.011. URL https://www.sciencedirect.com/science/article/pii/S0022073606005346.

Shenda Hong, Yuxi Zhou, Junyuan Shang, Cao Xiao, and Jimeng Sun. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in Biology and Medicine*, 122:103801, 2020. ISSN 0010-4825. doi: https://doi.org/10.1016/j.compbiomed.2020.103801. URL http://www.sciencedirect.com/science/article/pii/S0010482520301694.

Dani Kiyasseh, Tingting Zhu, and David A. Clifton. CLOCS: Contrastive Learning of Cardiac Signals Across Space, Time, and Patients, May 2021. URL http://arxiv.org/abs/2005.13249. arXiv:2005.13249 [cs, eess, stat].

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Daniel L. Lustgarten, Gautham Rajagopal, Jerry Reiland, Jodi Koehler, and Shantanu Sarkar. Premature ventricular contraction detection for long-term monitoring in an implantable cardiac monitor. *Pacing and Clinical Electrophysiology*, 43(5):462–470, 2020. doi: https://doi.org/10.1111/pace.13903. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/pace.13903. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/pace.13903.

G.B. Moody and R.G. Mark. The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001. doi: 10.1109/51.932724.

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 Gerardo Flores, George H Chen, Tom Pollard, Joyce C Ho, and Tristan Naumann (eds.), *Proceedings of the Conference on Health, Inference, and Learning*, volume 174 of *Proceedings of Machine Learning Research*, pp. 338–353. PMLR, April 2022. URL https://proceedings.mlr.press/v174/oh22a.html.

Julien Oster and Gari Clifford. Acquisition of electrocardiogram signals during magnetic resonance imaging. *Physiological Measurement*, 38(119), 2017. doi: https://doi.org/10.1088/1361-6579/aa6e8c.

Julien Oster, Raúl Llinares, Stephen Payne, Zion Tsz Ho Tse, Ehud Jeruham Schmidt, and Gari Clifford. Comparison of three artificial models of the magnetohydrodynamic effect on the electrocardiogram. *Computer Methods in Biomechanics and Biomedical Engineering*, 18(13):1400 – 1417, April 2014. doi: 10.1080/10255842.2014.909090. URL https://hal.univ-lorraine.fr/hal-01729606.

Jiapu Pan and Willis J. Tompkins. A real-time qrs detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236, 1985. doi: 10.1109/TBME.1985.325532.

Matthew A Reyna, Nadi Sadr, Erick A Perez Alday, Annie Gu, Amit Shah, Chad Robichaux, Bahrami Ali 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.

Matthew A Reyna, Nadi Sadr, Erick Perez Alday, Annie Gu, Amit Shah, Chad Robichaux, Bahrami Ali Rad, Andoni Elola, Salman Seyedi, Sardar Ansari, Hamid Ghanbari, Qiao Li, Ashish Sharma, and Gari D Clifford. Issues in the automated classification of mutlilead ECGs using heterogeneous labels and populations, 2022. URL https://moody-challenge.physionet.org/2021/papers/2021ChallengePaperPMEA.pdf.

Vikto Tihonenko, Alexander Khaustov, Sergey Ivanov, Alexei Rivin, and Evgeniy Yakushenko. St Petersburg INCART 12-lead Arrhythmia Database. *PhysioBank, PhysioToolkit, and PhysioNet*, 2008. doi: 10.13026/C2V88N.

Nora Vogt. CNNs, LSTMs, and Attention Networks for Pathology Detection in Medical Data. *CoRR*, abs/1912.00852, 2019. URL http://arxiv.org/abs/1912.00852. _eprint: 1912.00852.

Yatao Zhang, Junyan Li, Shoushui Wei, Fengyu Zhou, and Dong Li. Heartbeats Classification Using Hybrid Time-Frequency Analysis and Transfer Learning Based on ResNet. *IEEE Journal of Biomedical and Health Informatics*, 25(11):4175–4184, 2021. doi: 10.1109/JBHI.2021.3085318.

A Implementation Details

A.1 Siamese architecture

Figure 3 illustrates the used neural network for pretraining and the heartbeat classification task. Details on the network layers for the different stages of the architecture are provided in Table 3. On pretraining, the neural network is given two views from an ECG. One view is passed forward until the Predictor, while the other is passed only until the Projector. Then the gradient backpropagation is blocked on the branch which was passed only until the projector, in order to avoid mode collapse.

On the downstream task, the weights from the Embedding that was pretrained on the CinC dataset are reused in the Embedding of the new neural network.

A.2 Loss functions and optimizers

The SimSiamLoss is based on the cosine similarity distance. Given 2 vectors x_1, x_2

$$d(x_1, x_2) = 1 - \frac{x_1.x_2}{max(x_1, x_1, \epsilon)} \tag{1}$$

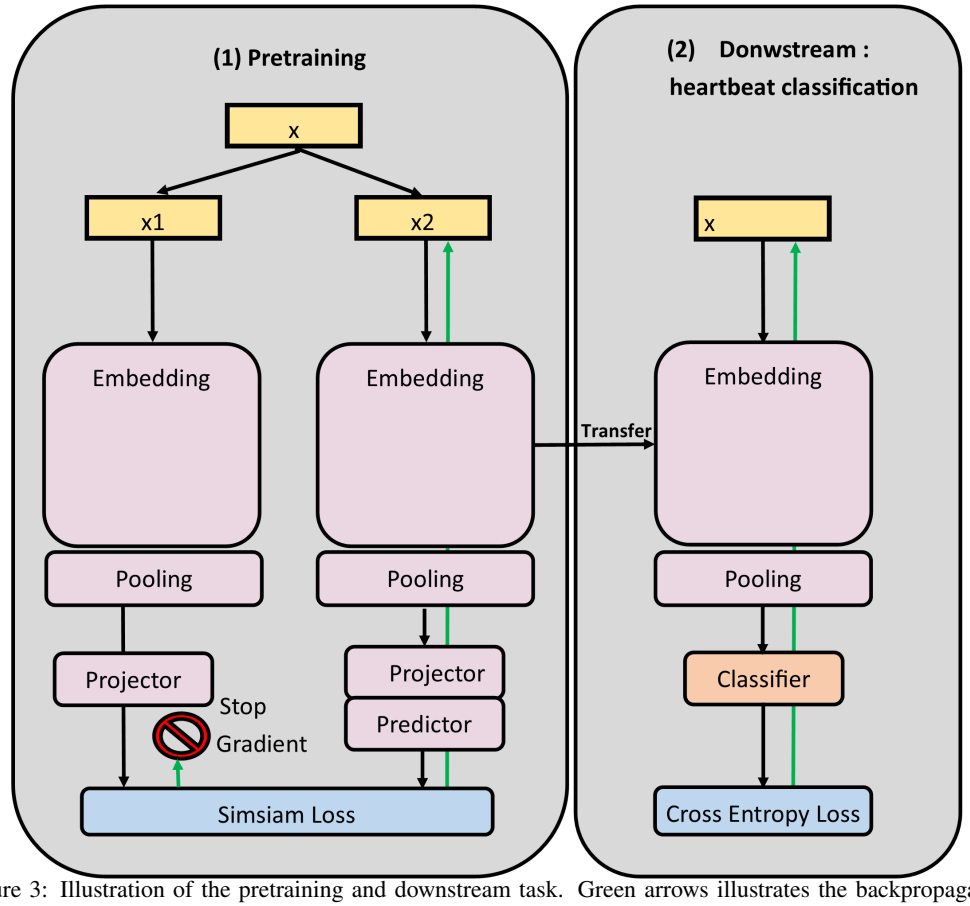

Figure 3: Illustration of the pretraining and downstream task. Green arrows illustrates the backpropagation paths.

Let z, p be respectively the projector and predictor output. Then given 2 views x_1, x_2 of the network :

$$L(x_1, x_2) = 1 + 0.5d(p_1, StopGradient(z_2)) + 0.5d(p_2, StopGradient(z_1)) \quad (2)$$

where $StopGradient$ means the backpropagation is blocked on the gradient computation.

Adam optimizer with weight decay decoupling Loshchilov & Hutter (2019) was used.

For both pretraining and downstream task, scheduler's initial and final learning rates were manually searched over the range $[10^{-6}10^{-2}]$.

B DETAILED LEAD RESULTS

Stage		Layer
Input		1 lead ECG beat (batch,1,301)
Embedding		BN, Conv(16),ReLU
	(x2)	Conv(32), BN, ReLU
		AvgPool
	(x2)	Conv(32), BN, ReLU
		AvgPool
	(x2)	Conv(64), BN, ReLU
		AvgPool
	(x2)	Conv(64), BN, ReLU
	(x2)	Conv(128), BN, ReLU
		Conv(256), BN, ReLU
Pooling		Global Max Pooling
Projector	(x2)	Linear(256, 256), BN
Predictor		Linear(256, 64), BN, GeLU
		Linear(64, 256)
Classifier		Linear(256, 2)

Table 3: Neural Network stages. BN is a batch normalization layer, Conv(n) is a 1d convolutional layer, with n filters, and kernel size of 7 and stride of 1. AvgPool is a 1d average pooling layer with kernel size of 2, stride of 2 and no padding.

		randfeatures	randinit	pretrained	finetune	2lr
DS2		0.50(0.06)	**0.89(0.04)**	0.87(0.02)	0.66(0.01)	**0.89(0.02)**
INCART	I	0.004(0.002)	0.46(0.12)	0.63(0.07)	**0.68(0.01)**	0.67(0.02)
	II	0.46(0.07)	0.74(0.04)	**0.79(0.05)**	0.77(0.01)	**0.79(0.03)**
	III	0.37(0.11)	0.61(0.07)	0.65(0.05)	0.70(0.01)	**0.71(0.03)**
	AVR	0.17(0.09)	0.25(0.09)	0.57(0.14)	0.74(0.01)	**0.75(0.04)**
	AVL	0.06(0.02)	0.29(0.10)	0.56(0.09)	0.45(0.01)	**0.64(0.04)**
	AVF	0.43(0.09)	0.72(0.06)	0.76(0.06)	0.74(0.01)	**0.77(0.03)**
	V1	0.23(0.09)	0.24(0.07)	0.43(0.12)	**0.82(0.01)**	0.56(0.07)
	V2	0.33(0.11)	0.26(0.04)	0.37(0.09)	**0.75(0.01)**	0.43(0.06)
	V3	0.35(0.08)	0.33(0.05)	0.43(0.08)	**0.73(0.01)**	0.49(0.04)
	V4	0.29(0.06)	0.62(0.07)	0.68(0.09)	**0.74(0.01)**	0.72(0.04)
	V5	0.42(0.06)	0.72(0.05)	0.81(0.03)	0.79(0.01)	**0.87(0.02)**
	V6	0.28(0.05)	0.67(0.04)	0.74(0.03)	0.82(0.01)	**0.86(0.03)**
	Leads Average	0.28(0.14)	0.49(0.21)	0.62(0.14)	**0.72(0.10)**	0.69(0.13)

Table 4: F1 score (Mean (and standard deviation)) for DS2 and on each lead of INCART. The line "Leads Average" corresponds to the mean F1 score over the INCART 12 leads.

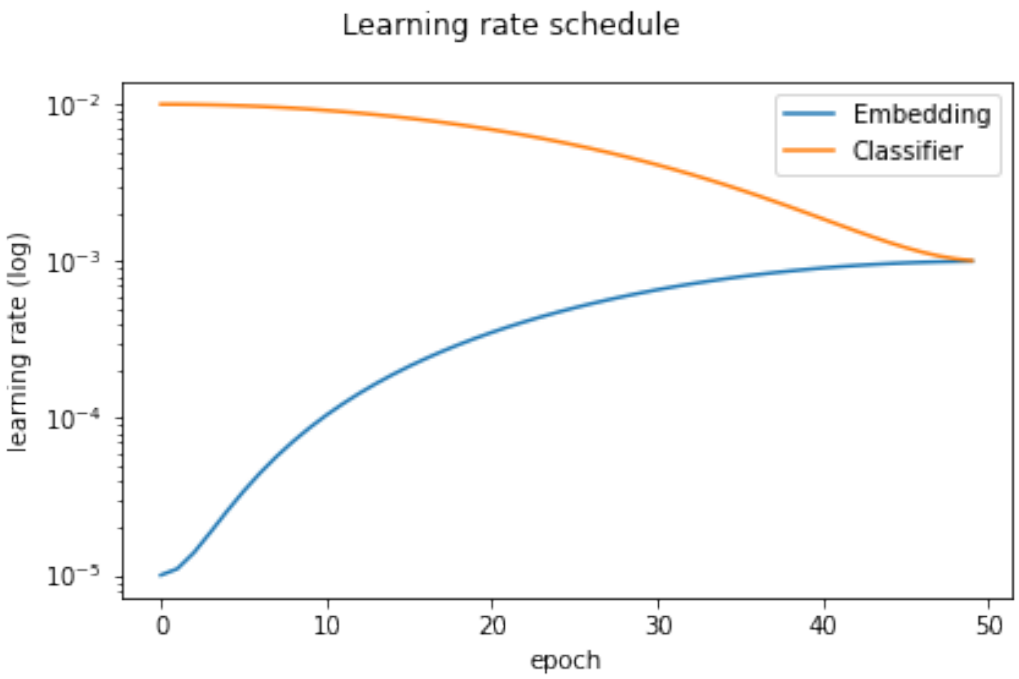

Figure 4: Cosine learning rate schedule for the downstream task in the 2lr approach. The learning rate is adjusted at each epoch end.

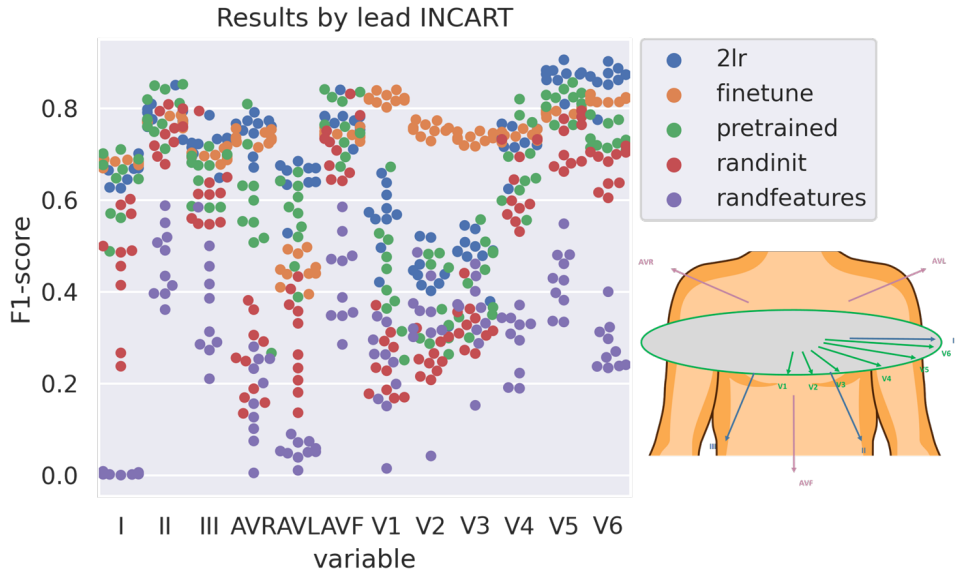

Figure 5: Lead by lead F1 scores on the INCART database.