A. Datasets

A.1. Data Preprocessing

For all of the datasets, frames consisted of 2500 samples and consecutive frames had no overlap with one another. Data splits were always performed at the patient-level.

PhysioNet 2020 (Perez Alday et al., 2020). Each ECG recording varied in duration from 6 seconds to 60 seconds with a sampling rate of 500Hz. Each ECG frame in our setup consisted of 2500 samples (5 seconds). We assign multiple labels to each ECG recording as provided by the original authors. These labels are: AF, I-AVB, LBBB, Normal, PAC, PVC, RBBB, STD, and STE. The ECG frames were normalized in amplitude between the values of 0 and 1.

Chapman (Zheng et al., 2020). Each ECG recording was originally 10 seconds with a sampling rate of 500Hz. We downsample the recording to 250Hz and therefore each ECG frame in our setup consisted of 2500 samples. We follow the labelling setup suggested by (Zheng et al., 2020) which resulted in four classes: Atrial Fibrillation, GSVT, Sudden Bradychardia, Sinus Rhythm. The ECG frames were normalized in amplitude between the values of 0 and 1.

Cardiology (Hannun et al., 2019). Each ECG recording was originally 30 seconds with a sampling rate of 200Hz. Each ECG frame in our setup consisted of 256 samples resampled to 2500 samples. Labels made by a group of physicians were used to assign classes to each ECG frame depending on whether that label coincided in time with the ECG frame. These labels are: AFIB, AVB, BIGEMINY, EAR, IVR, JUNCTIONAL, NOISE, NSR, SVT, TRIGEMINY, VT, and WENCKEBACH. Sudden bradycardia cases were excluded from the data as they were not included in the original formulation by the authors. The ECG frames were not normalized.

PhysioNet 2017 (Clifford et al., 2017). Each ECG recording originally varied in length between 9 and 30 seconds with a sampling rate of 300Hz. Each ECG frames in our setup consisted of 2500 samples. We use the original labels, resulting in four classes: Normal, AF, Other, and Noisy. The ECG frames were not normalized.

A.2. Data Samples

A.2.1. Self-supervised Pre-training

In this section, we outline the dimension of the inputs used for the various pre-training methods. They are expressed in the form of $N \times S \times L$ where N is the total number of instances, S is the frame length of each instance, and L (if applicable) is the number of leads used. Where L is not explicitly mentioned, we report values with four leads as this was primarily used for all experiments conducted.

Table 3. Dimension of the input data, $N \times S \times L$, used during the training and validation phases of the various self-supervised pre-training methods. $S = 2500$ is the number of samples in each instance fed to the network. L is the number of leads (projections) used during pre-training.

Dataset	Method	Train	Validation
PhysioNet 2020	BYOL	51,880 $\times S$	12,948 $\times S$
	SimCLR	51,880 $\times S$	12,948 $\times S$
	CMSC	24,080 $\times 2S$	6,076 $\times 2S$
	CMLC	24,080 $\times S \times L$	6,076 $\times S \times L$
	CMSMLC	6,020 $\times 2S \times L$	1,519 $\times 2S \times L$
Chapman	BYOL	25,543 $\times S$	8,512 $\times S$
	SimCLR	25,543 $\times S$	8,512 $\times S$
	CMSC	25,543 $\times 2S$	8,512 $\times 2S$
	CMLC	25,543 $\times S \times L$	8,512 $\times S \times L$
	CMSMLC	6,382 $\times 2S \times L$	2125 $\times 2S \times L$

A.2.2. Supervised Training

In this section, we outline the number of instances used during supervised training on the downstream tasks. For multi-lead datasets, we report these values having used four leads. A simple multiplicative factor can be used to deduce the number of instances used with a different number of leads.

Table 4. Number of instances (number of patients) used during the supervised training of the downstream tasks. For multi-lead datasets*, these represent sample sizes for the four leads (II, V2, aVL, aVR).

Dataset	Train	Validation	Test
PhysioNet 2020*	51,880 (4,402)	12,948 (1,100)	15,820 (1,375)
Chapman*	25,543 (6,387)	8,512 (2,129)	8,520 (2,130)
Cardiology	4,584 (201)	1,109 (50)	1,386 (62)
PhysioNet 2017	18,256 (5,459)	4,581 (1,364)	5,824 (1,705)

B. Visualization of Data Augmentations

In this section, we outline the various data augmentations applied to the time-series signals and provide exemplar visualizations for the reader. In Fig. 7a, We present a single, unperturbed ECG frame for illustration purposes. To that original frame, we apply the following transformations:

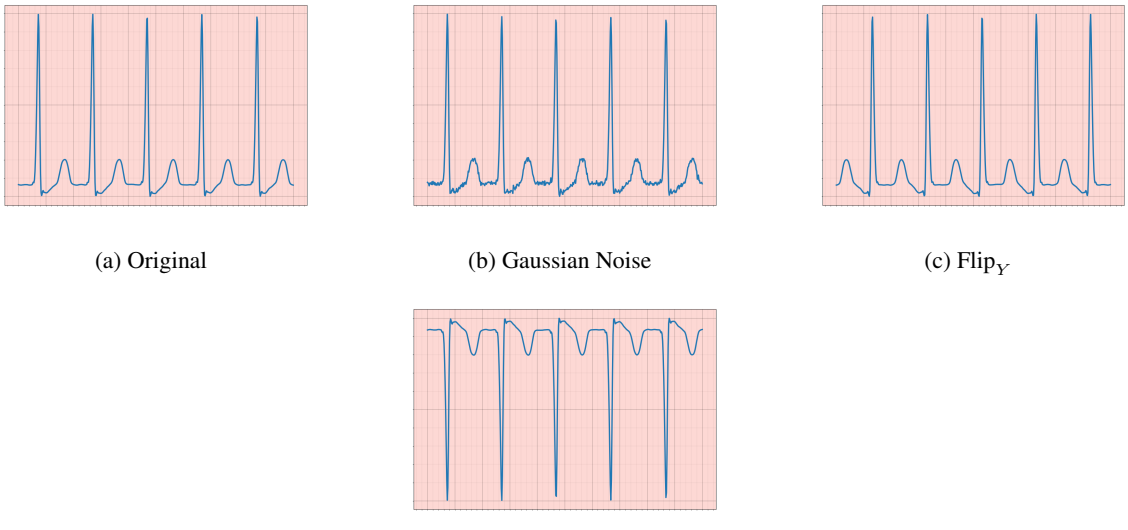
1. **Gaussian Noise**: Gaussian noise, $\epsilon \sim \mathcal{N}(0, \sigma)$ is added to the original frame. We chose the value of σ in order to preserve the class of the original frame. More concretely, for the Chapman dataset, $\sigma = 10$, whereas for the PTB-XL dataset, $\sigma = 0.01$. The difference in the magnitude of σ across datasets is attributed to the difference in the magnitude of the original signals from each dataset. Although it can be argued that such noise is trivial for a contrastive learning setup, recent work has shown the effect of additive noise on ECG signals (Han et al., 2020).

2. **Flip**$_Y$: we perturb the original frame by flipping it along the temporal dimension. In other words, the signal is read in reverse. In designing this perturbation, we were motivated by the self-supervision task proposed by () that revolves around reversing the 'arrow of time'.

3. **Flip**$_X$: we perturb the original frame by negating the magnitude of the signal. This perturbation was motivated by the fact that ECG recordings made by physical leads that are connected to the patient's body incorrectly could lead to such 'inverted' signals.

4. **SpecAugment**: we perturb the original frame by masking spectral or temporal components of the signal. To do so, we follow a similar setup to that introduced in SpecAugment (Park et al., 2019). We take the Short-time Fourier transform (STFT) of the signal, randomly choose the number *and* width of spectral or temporal bins to mask. As the resultant STFT is a matrix of complex numbers, we set masked values to zero. Lastly, we perform the Inverse STFT (ISTFT) to obtain the signal in the time-domain. We provide step-by-step instructions on how to apply this perturbation in the next section.

B.1. Perturbations



(a) Original (b) Gaussian Noise (c) Flip$_Y$

(d) Flip$_X$

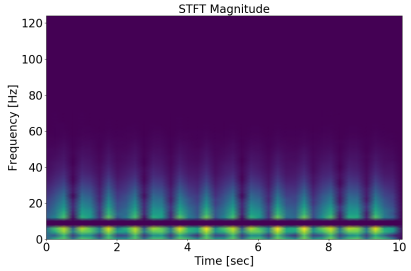
Figure 7. ECG segment a) without any perturbations, b) with additive Gaussian noise, c) after being flipped temporally, Flip$_Y$, and d) after being flipped along the x-axis, Flip$_X$.

B.2. SpecAugment

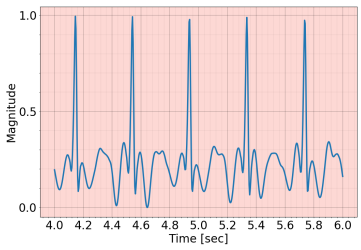
To apply the SpecAugment perturbations, we followed these steps:

1. Apply the Short-time Fourier transform (STFT) to the time-series signal. This splits the signal into N_f spectral and N_t temporal bins.
2. Depending on whether a spectral or temporal mask is desired, the bin width, $w \in [0, 1]$, defines the fraction of the total number of bins to mask. For example, $w = 0.5$ means that 50% of the bins are masked. The total number of bins to mask is thus $N_m = w \times N_f$ or $N_m = w \times N_t$.
3. Now that we have the number of bins to mask, we need to identify *which* bins to mask. We formulate this as identifying the bin to start the masking and do so by uniformly sampling a number, start, from 0 to $N_f - N_m$. The masked bins range from start to $\mathrm{start} + N_m$.
4. As the STFT of a time-series signal is a complex number, masking involves setting the complex-valued entries to zero, i.e., $0 + 0j$.
5. This process is repeated R times until all desired components are masked.
6. We convert the masked STFT back to the time-domain by taking its inverse (ISTFT).

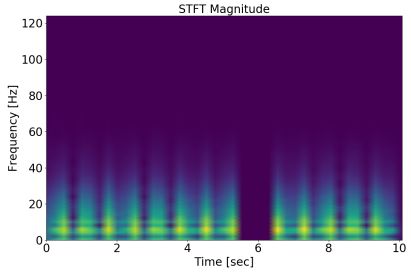
In following the aforementioned steps, several hyperparameters exist. In our implementation, we chose $w = 0.2$ and $R = 1$ to balance between masking too many components which might violate the assumption of a shared context and masking too few components which would make the contrastive learning task quite trivial.



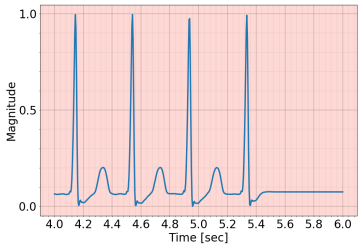
(a) Spectrogram with spectral mask, SA_f



(b) Frame with spectral mask



(c) Sepctrogram with temporal mask, SA_t



(d) Frame with temporal mask

Figure 8. Illustration of SpecAugment perturbations applied to the original ECG segment shown in Fig. 7. (a), (c) spectrograms with spectral and temporal masks, respectively. (b), (d) time-series representations of the masked spectrograms. Note that the time-series segments only span seconds 4-6.

C. Implementation Details

C.1. Network Architecture

In this section, we outline the architecture of the neural network used for all experiments. For pre-training, the final layer (Layer 5) was removed and representations with dimension E were learned. During training on the downstream tasks, the final layer was introduced.

Table 5. Network architecture used for all experiments. K, C_{in}, and C_{out} represent the kernel size, number of input channels, and number of output channels, respectively. A stride of 3 was used for all convolutional layers. E represents the dimension of the final representation.

Layer Number	Layer Components	Kernel Dimension
1	Conv 1D BatchNorm ReLU MaxPool(2) Dropout(0.1)	7 x 1 x 4 (K x C_{in} x C_{out})
2	Conv 1D BatchNorm ReLU MaxPool(2) Dropout(0.1)	7 x 4 x 16
3	Conv 1D BatchNorm ReLU MaxPool(2) Dropout(0.1)	7 x 16 x 32
4	Linear ReLU	320 x E
5	Linear	E x C (classes)

C.2. Experiment Details

Table 6. Batchsize and learning rates used for training with different datasets. The Adam optimizer was used for all experiments.

Dataset	Batchsize	Learning Rate
PhysioNet 2020	256	10^{-4}
Chapman	256	10^{-4}
Cardiology	16	10^{-4}
PhysioNet 2017	256	10^{-4}

C.3. Baseline Implementations

C.3.1. Supervised Pre-training

In this implementation, we pre-train on the specified dataset under the assumption that 100% of the data is labelled and available for training (i.e., $F = 1$). Given the presence of labels, pre-training involves solving a supervised classification task to diagnose the cardiac arrhythmia that corresponds to each ECG recording. In our context, supervised pre-training is expected to generate the best downstream generalization performance due to the availability of labels *and* the high similarity between the upstream and downstream tasks, namely cardiac arrhythmia classification.

C.3.2. MT-SSL

In this implementation, we introduce six different pre-text tasks that are used for pre-training a network. We follow the multi-task pre-training setup proposed by (Sarkar & Etemad, 2020) where six different classification heads are used to solve each of the six tasks. These tasks comprise binary classification where the network is asked to discriminate between ECG instances and their perturbed counterpart. Such perturbations take on the form of 1) Gaussian noise addition, 2) scaling, 3) negation, 4) temporal inversion, 5) permutation, and 6) time-warping. For the Chapman dataset, we only pre-train using scaling, negation, and temporal inversion since additional tasks prevented the network from converging. On the PhysioNet2020 dataset, however, we pre-train using all of the aforementioned tasks.

C.3.3. BYOL

In this implementation, an instance is perturbed by applying two stochastic transformations. In our setup, these transformations can include any of those outlined in Appendix B. This process results in two views of the same instance, each of which is passed through an online network and a target network. The target network is an exponential moving average of the online network, and is thus a delayed version of the online network. This delay is dictated by the decay rate, τ_d. We chose $\tau_d = 0.9$ with experiments to validate this decision in Appendix F. A key difference between the two networks is that they are *asymmetric*, with the online network consisting of an additional prediction head. The goal is for the representation from the online network to predict that from the target network. This is done by minimizing the mean squared error of the two representations. In our setup, we introduce asymmetry by repeating Layer 4 shown in Appendix C.1. This is similar to what was performed by (Grill et al., 2020).

C.3.4. SimCLR

In this implementation, an instance is perturbed by applying two stochastic transformations. In our setup, these transformations can include any of those outlined in Appendix B. This process results in two views of the same instance, each of which is passed through the same network. The InfoNCE loss is used to attract representations that are similar to one another and repel those that are different. Whether representations should be attracted to one another depends on whether they belong to the same original instance.

D. Linear Evaluation of Representations

In this section, we evaluate the utility of the representations learned as a result of self-supervised pre-training. We pre-train on two different datasets, freeze the network parameters, and transfer them to a downstream task whereby a linear multinomial logistic regression (MLR) model is trained. In doing so, we are evaluating the richness of the representations learned. We perform these experiments under two scenarios. The first involves pre-training and evaluating using 4 leads (II, V2, aVL, aVR) (see Sec. D.1). The second involves pre-training and evaluating using all 12 leads (see Sec. D.2). We chose these two scenarios to help determine whether our findings generalize to domains where a different number of leads is available.

D.1. Pre-training and Evaluating using 4 leads

We present Tables 7 - 10 which illustrate the test AUC of an MLR evaluated on Chapman and PhysioNet 2020 after having pre-trained on these two datasets using only 4 of the 12 leads, respectively, These are presented for a range of embedding dimensions, $E = (32, 64, 128, 256)$, and available labelled training data, $F = (0.25, 0.50, 0.75, 1)$.

D.1.1. Embedding Dimension, $E = 32$

We show that CMSMLC outperforms all other methods when evaluating on Chapman, regardless of the available labelled training data. This can be seen by the higher AUC achieved by this method relative to the remaining methods. For instance, at $F = 0.25$, CMSMLC achieves an $\mathrm{AUC} = 0.844$ compared to 0.665 for SimCLR. When evaluating on PhysioNet 2020, we find that CMSC consistently outperforms the remaining methods, as seen by its higher test AUC values.

Table 7. Comparison of self-supervised methods when using networks as feature extractors and performing linear evaluation on downstream datasets. Pre-training and evaluating multi-lead datasets* using 4 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.665 ± 0.014	0.564 ± 0.009
CMSC	0.831 ± 0.131	$\mathbf{0.701 \pm 0.046}$
CMLC	0.789 ± 0.020	0.563 ± 0.008
CMSMLC	$\mathbf{0.844 \pm 0.023}$	0.619 ± 0.019

(b) $F = 0.5$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.666 ± 0.015	0.587 ± 0.009
CMSC	0.831 ± 0.131	$\mathbf{0.707 \pm 0.038}$
CMLC	0.801 ± 0.016	0.572 ± 0.008
CMSMLC	$\mathbf{0.850 \pm 0.022}$	0.636 ± 0.020

(c) $F = 0.75$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.670 ± 0.013	0.591 ± 0.010
CMSC	0.833 ± 0.129	$\mathbf{0.709 \pm 0.039}$
CMLC	0.805 ± 0.018	0.585 ± 0.009
CMSMLC	$\mathbf{0.850 \pm 0.021}$	0.643 ± 0.020

(d) $F = 1$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.670 ± 0.013	0.594 ± 0.010
CMSC	0.831 ± 0.131	$\mathbf{0.709 \pm 0.038}$
CMLC	0.807 ± 0.017	0.593 ± 0.009
CMSMLC	$\mathbf{0.852 \pm 0.021}$	0.645 ± 0.021

D.1.2. EMBEDDING DIMENSION, $E = 64$

We find that the conclusions arrived at with $E = 32$ are similar to those in this scenario. Namely, CMSMLC outperforms all remaining methods when evaluating on Chapman. On the other hand, CMSC outperforms all methods when evaluating on PhysioNet 2020. This can be seen by the bold test AUC values in Table 8.

Table 8. Comparison of self-supervised methods when using networks as feature extractors and performing linear evaluation on downstream datasets. Pre-training and evaluating multi-lead datasets* using 4 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.709 ± 0.019	0.574 ± 0.005
CMSC	0.829 ± 0.130	$\mathbf{0.720 \pm 0.012}$
CMLC	0.842 ± 0.020	0.592 ± 0.019
CMSMLC	$\mathbf{0.856 \pm 0.022}$	0.641 ± 0.023

(b) $F = 0.5$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.722 ± 0.025	0.599 ± 0.010
CMSC	0.830 ± 0.132	$\mathbf{0.721 \pm 0.013}$
CMLC	0.850 ± 0.02	0.607 ± 0.018
CMSMLC	$\mathbf{0.861 \pm 0.02}$	0.662 ± 0.020

(c) $F = 0.75$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.726 ± 0.023	0.604 ± 0.010
CMSC	0.831 ± 0.126	$\mathbf{0.725 \pm 0.009}$
CMLC	0.854 ± 0.021	0.619 ± 0.017
CMSMLC	$\mathbf{0.861 \pm 0.021}$	0.671 ± 0.018

(d) $F = 1$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.727 ± 0.025	0.608 ± 0.010
CMSC	0.832 ± 0.126	$\mathbf{0.726 \pm 0.009}$
CMLC	0.855 ± 0.020	0.627 ± 0.015
CMSMLC	$\mathbf{0.862 \pm 0.020}$	0.673 ± 0.018

D.1.3. EMBEDDING DIMENSION = 128

In this scenario and in contrast to conclusions arrived at with $E = 32$ and 64, we find that CMSC outperforms all methods when evaluated on both datasets, Chapman and PhysioNet 2020. This can be seen by the bold test AUC values in Table 13. For instance, at $F = 0.25$, CMSC achieves an $\mathrm{AUC} = 0.895$ compared to 0.727 achieved by SimCLR. That is a 16.8% improvement relative to the state-of-the-art.

Table 9. Comparison of self-supervised methods when using networks as feature extractors and performing linear evaluation on downstream datasets. Pre-training and evaluating multi-lead datasets* using 4 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Dataset	Chapman*	PhysioNet 2020*
BYOL	0.671 ± 0.042	0.587 ± 0.021
SimCLR	0.727 ± 0.032	0.585 ± 0.016
CMSC	$\mathbf{0.895 \pm 0.004}$	$\mathbf{0.713 \pm 0.032}$
CMLC	0.863 ± 0.026	0.580 ± 0.007
CMSMLC	0.842 ± 0.021	0.661 ± 0.010

(b) $F = 0.5$

Dataset	Chapman*	PhysioNet 2020*
BYOL	0.643 ± 0.043	0.595 ± 0.018
SimCLR	0.738 ± 0.034	0.615 ± 0.014
CMSC	$\mathbf{0.896 \pm 0.005}$	$\mathbf{0.715 \pm 0.033}$
CMLC	0.870 ± 0.022	0.596 ± 0.008
CMSMLC	0.847 ± 0.024	0.680 ± 0.008

(c) $F = 0.75$

Dataset	Chapman*	PhysioNet 2020*
BYOL	0.666 ± 0.032	0.598 ± 0.022
SimCLR	0.742 ± 0.033	0.620 ± 0.015
CMSC	$\mathbf{0.898 \pm 0.002}$	$\mathbf{0.717 \pm 0.033}$
CMLC	0.872 ± 0.022	0.606 ± 0.008
CMSMLC	0.848 ± 0.023	0.685 ± 0.008

(d) $F = 1$

Dataset	Chapman*	PhysioNet 2020*
BYOL	0.653 ± 0.026	0.602 ± 0.015
SimCLR	0.742 ± 0.033	0.623 ± 0.014
CMSC	$\mathbf{0.897 \pm 0.003}$	$\mathbf{0.718 \pm 0.033}$
CMLC	0.873 ± 0.021	0.612 ± 0.010
CMSMLC	0.849 ± 0.022	0.686 ± 0.008

D.1.4. EMBEDDING DIMENSION, $E = 256$

In this scenario, we find that CMLC consistently outperforms all methods when evaluated on Chapman. Similar to findings at lower embedding dimensions, CMSC outperforms all methods on PhysioNet 2020. These claims are supported by the bold test AUC values in Table 10.

Table 10. Comparison of self-supervised methods when using networks as feature extractors and performing linear evaluation on downstream datasets. Pre-training and evaluating multi-lead datasets* using 4 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.742 ± 0.031	0.591 ± 0.007
CMSC	0.832 ± 0.128	$\mathbf{0.721 \pm 0.016}$
CMLC	$\mathbf{0.883 \pm 0.009}$	0.607 ± 0.027
CMSMLC	0.828 ± 0.040	0.652 ± 0.023

(b) $F = 0.5$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.749 ± 0.032	0.615 ± 0.010
CMSC	0.833 ± 0.130	$\mathbf{0.722 \pm 0.017}$
CMLC	$\mathbf{0.887 \pm 0.008}$	0.619 ± 0.026
CMSMLC	0.831 ± 0.042	0.670 ± 0.018

(c) $F = 0.75$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.752 ± 0.033	0.619 ± 0.010
CMSC	0.833 ± 0.130	$\mathbf{0.723 \pm 0.017}$
CMLC	$\mathbf{0.889 \pm 0.007}$	0.626 ± 0.026
CMSMLC	0.831 ± 0.040	0.675 ± 0.018

(d) $F = 1$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.753 ± 0.033	0.621 ± 0.010
CMSC	0.833 ± 0.132	$\mathbf{0.724 \pm 0.017}$
CMLC	$\mathbf{0.890 \pm 0.006}$	0.633 ± 0.026
CMSMLC	0.832 ± 0.039	0.677 ± 0.018

D.2. Pre-training and Evaluating using 12 leads

We present Tables 11 - 14 which illustrate the test AUC of an MLR evaluated on Chapman and PhysioNet 2020 after having pre-trained on these two datasets using all 12 leads, respectively, These are presented for a range of embedding dimensions, $E = (32, 64, 128, 256)$, and available labelled training data, $F = (0.25, 0.50, 0.75, 1)$. Overall, we find that pre-training and evaluating with all 12 leads results in a clearly superior self-supervised method, CMSC. We support this claim with the results presented in the subsequent sections. This finding is in contrast to what we observed when pre-training and evaluating on only 4 of the 12 leads. In that scenario, although our proposed methods outperform SimCLR, CMSC does not consistently outperform the other methods.

D.2.1. Embedding Dimension, $E = 32$

We show that CMSC consistently outperforms all other pre-training methods when evaluated on both the Chapman and PhysioNet 2020 dataset. This can be seen by the higher AUC achieved by this method relative to the remaining methods. For instance, when evaluating on the Chapman dataset using only 25% of the labels ($F = 0.25$) during training, CMSC achieves an $\mathrm{AUC} = 0.899$ compared to 0.667 for SimCLR.

Table 11. Comparison of self-supervised methods when using networks as feature extractors and performing linear evaluation on downstream datasets. Pre-training and evaluating multi-lead datasets* using all 12 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.667 ± 0.019	0.585 ± 0.013
CMSC	$\mathbf{0.899 \pm 0.003}$	$\mathbf{0.744 \pm 0.011}$
CMLC	0.728 ± 0.021	0.627 ± 0.037
CMSMLC	0.838 ± 0.015	0.644 ± 0.026

(b) $F = 0.5$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.667 ± 0.021	0.585 ± 0.015
CMSC	$\mathbf{0.898 \pm 0.004}$	$\mathbf{0.741 \pm 0.011}$
CMLC	0.729 ± 0.019	0.627 ± 0.038
CMSMLC	0.838 ± 0.018	0.641 ± 0.030

(c) $F = 0.75$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.667 ± 0.020	0.589 ± 0.014
CMSC	$\mathbf{0.897 \pm 0.004}$	$\mathbf{0.747 \pm 0.012}$
CMLC	0.730 ± 0.019	0.635 ± 0.034
CMSMLC	0.843 ± 0.018	0.652 ± 0.025

(d) $F = 1$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.667 ± 0.019	0.589 ± 0.013
CMSC	$\mathbf{0.897 \pm 0.004}$	$\mathbf{0.745 \pm 0.013}$
CMLC	0.726 ± 0.019	0.632 ± 0.038
CMSMLC	0.844 ± 0.015	0.649 ± 0.027

D.2.2. EMBEDDING DIMENSION, $E = 64$

We find that the conclusions arrived at with $E = 32$ are similar to those in this scenario. Namely, CMSC outperforms all remaining methods when evaluating on both Chapman and PhysioNet 2020. Moreover, our other proposed pre-training methods (CMLC and CMSMLC) also outperform the state-of-the-art method, SimCLR. This can be seen by the bold test AUC values in Table 12.

Table 12. Comparison of self-supervised methods when using networks as feature extractors and performing linear evaluation on downstream datasets. Pre-training and evaluating multi-lead datasets* using all 12 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.752 ± 0.045	0.611 ± 0.009
CMSC	$\mathbf{0.904 \pm 0.005}$	$\mathbf{0.764 \pm 0.022}$
CMLC	0.734 ± 0.021	0.650 ± 0.039
CMSMLC	0.852 ± 0.024	0.669 ± 0.016

(b) $F = 0.5$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.753 ± 0.046	0.613 ± 0.011
CMSC	$\mathbf{0.905 \pm 0.005}$	$\mathbf{0.759 \pm 0.024}$
CMLC	0.735 ± 0.020	0.650 ± 0.039
CMSMLC	0.851 ± 0.023	0.665 ± 0.016

(c) $F = 0.75$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.753 ± 0.046	0.617 ± 0.008
CMSC	$\mathbf{0.904 \pm 0.005}$	$\mathbf{0.770 \pm 0.019}$
CMLC	0.735 ± 0.020	0.661 ± 0.035
CMSMLC	0.854 ± 0.018	0.675 ± 0.016

(d) $F = 1$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.754 ± 0.045	0.616 ± 0.009
CMSC	$\mathbf{0.905 \pm 0.005}$	$\mathbf{0.770 \pm 0.019}$
CMLC	0.735 ± 0.020	0.654 ± 0.040
CMSMLC	0.853 ± 0.017	0.674 ± 0.016

D.2.3. EMBEDDING DIMENSION, $E = 128$

In this scenario, the same conclusions as those arrived at with smaller embedding dimensions still hold. Although CMSC continues to outperform all other methods, the performance gap between such methods decreases when compared to results obtained at smaller embedding dimensions. For instance, when evaluating on the Chapman dataset at $F = 0.25$ with $E = 128$, CMSC achieves an $\mathrm{AUC} = 0.903$ whereas SimCLR achieves an $\mathrm{AUC} = 0.771$, a performance gap of 13.2%. In contrast, at $E = 64$, the performance gap between these two methods was 15.2%.

Table 13. Comparison of self-supervised methods when using networks as feature extractors and performing linear evaluation on downstream datasets. Pre-training and evaluating multi-lead datasets* using all 12 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.771 ± 0.012	0.605 ± 0.013
CMSC	$\mathbf{0.903 \pm 0.002}$	$\mathbf{0.7600 \pm 0.019}$
CMLC	0.779 ± 0.018	0.667 ± 0.030
CMSMLC	0.846 ± 0.024	0.659 ± 0.016

(b) $F = 0.5$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.773 ± 0.012	0.606 ± 0.013
CMSC	$\mathbf{0.902 \pm 0.003}$	$\mathbf{0.758 \pm 0.019}$
CMLC	0.783 ± 0.020	0.665 ± 0.032
CMSMLC	0.850 ± 0.022	0.659 ± 0.016

(c) $F = 0.75$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.774 ± 0.012	0.611 ± 0.012
CMSC	$\mathbf{0.902 \pm 0.003}$	$\mathbf{0.763 \pm 0.019}$
CMLC	0.788 ± 0.018	0.671 ± 0.032
CMSMLC	0.851 ± 0.019	0.669 ± 0.013

(d) $F = 1$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.775 ± 0.012	0.610 ± 0.013
CMSC	$\mathbf{0.902 \pm 0.003}$	$\mathbf{0.761 \pm 0.019}$
CMLC	0.787 ± 0.020	0.672 ± 0.028
CMSMLC	0.853 ± 0.017	0.669 ± 0.013

D.2.4. EMBEDDING DIMENSION, $E = 256$

In this scenario, and similar to findings at lower embedding dimensions, CMSC outperforms all methods on both the Chapman and PhysioNet 2020 datasets. Pre-training and evaluating multi-lead datasets* using all 12 leads. These claims are supported by the bold test AUC values in Table 14.

Table 14. Comparison of self-supervised methods when using networks as feature extractors and performing linear evaluation on downstream datasets. Pre-training and evaluating multi-lead datasets* using all 12 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.769 ± 0.028	0.614 ± 0.007
CMSC	$\mathbf{0.904 \pm 0.002}$	$\mathbf{0.761 \pm 0.011}$
CMLC	0.784 ± 0.013	0.672 ± 0.033
CMSMLC	0.852 ± 0.013	0.672 ± 0.013

(b) $F = 0.5$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.770 ± 0.027	0.617 ± 0.008
CMSC	$\mathbf{0.906 \pm 0.002}$	$\mathbf{0.756 \pm 0.010}$
CMLC	0.790 ± 0.016	0.668 ± 0.036
CMSMLC	0.852 ± 0.010	0.672 ± 0.012

(c) $F = 0.75$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.770 ± 0.026	0.619 ± 0.008
CMSC	$\mathbf{0.905 \pm 0.003}$	$\mathbf{0.764 \pm 0.011}$
CMLC	0.793 ± 0.019	0.680 ± 0.029
CMSMLC	0.854 ± 0.012	0.680 ± 0.012

(d) $F = 1$

Dataset	Chapman*	PhysioNet 2020*
SimCLR	0.771 ± 0.027	0.619 ± 0.008
CMSC	$\mathbf{0.906 \pm 0.003}$	$\mathbf{0.764 \pm 0.010}$
CMLC	0.797 ± 0.016	0.677 ± 0.029
CMSMLC	0.858 ± 0.011	0.679 ± 0.011

E. Transfer Capabilities of Representations

In this section, we evaluate the utility of self-supervised pre-training in generating a favourable parameter initialization for a downstream task. After pre-training, we transfer the parameters to a downstream task and allow all parameters to be updated. In doing so, we are evaluating the benefit brought about by the inductive bias of self-supervised pre-training.

We perform these experiments under two scenarios. The first involves pre-training, fine-tuning, and evaluating using 4 leads (II, V2, aVL, aVR) (see Sec. E.1). The second involves pre-training, fine-tuning, and evaluating using all 12 leads (see Sec. E.2). We chose these two scenarios for several reasons. Firstly, they will help determine whether our findings generalize to domains where a different number of leads is available. For example, expensive hospital equipment may record all 12 leads of an ECG, whereas low-cost wearable sensors may only collect data from a subset of leads. Secondly, we wanted to evaluate whether or not contrastive-learning with more views (leads) would improve generalization performance on the downstream task. Previous studies in computer vision have shown this to be the case.

E.1. Pre-training, Fine-Tuning, and Evaluating using 4 Leads

We present Tables 15 - 18 which illustrate the test AUC on downstream datasets after having pre-trained on Chapman or PhysioNet 2020. These are shown for a range of embedding dimensions, $E = (32, 64, 128, 256)$, and available labelled training data, $F = (0.25, 0.50, 0.75, 1.00)$.

E.1.1. Embedding Dimension, $E = 32$

In this section, we show that in 18/24 (75%) of all experiments conducted, our family of contrastive learning methods outperforms the state-of-the-art method, SimCLR. This can be seen by the bold test AUC results in Table 15. The majority of these positive results can be attributed to CMSC. Such a finding illustrates the robustness of our methods to the pre-training and downstream dataset used for evaluation, especially given the diversity of the tasks at hand.

Table 15. Comparison of self-supervised methods when used as parameter initializations before fine-tuning on downstream datasets. Pre-training, fine-tuning, and evaluating multi-lead datasets* using 4 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.631 ± 0.006	0.738 ± 0.014	0.766 ± 0.005	0.631 ± 0.006	0.738 ± 0.014	0.898 ± 0.002
SimCLR	0.649 ± 0.012	0.731 ± 0.017	0.790± 0.008	0.642 ± 0.020	0.738 ± 0.009	0.907 ± 0.013
CMSC	0.661 ± 0.018	**0.770 ± 0.012**	**0.801 ± 0.013**	**0.658 ± 0.018**	0.748 ± 0.027	**0.908 ± 0.011**
CMLC	0.652 ± 0.014	0.767 ± 0.012	0.768 ± 0.004	0.635 ± 0.017	**0.753 ± 0.013**	0.906 ± 0.009
CMSMLC	**0.669 ± 0.020**	0.758 ± 0.008	0.761 ± 0.015	0.652 ± 0.013	0.733 ± 0.004	0.900 ± 0.009

(b) $F = 0.5$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.669 ± 0.007	0.782 ± 0.011	0.811 ± 0.011	0.669 ± 0.007	0.782 ± 0.011	0.907 ± 0.011
SimCLR	0.691 ± 0.008	0.748 ± 0.018	**0.829 ± 0.003**	0.679 ± 0.012	0.767 ± 0.012	**0.933 ± 0.010**
CMSC	0.687 ± 0.018	**0.771 ± 0.030**	0.822 ± 0.011	**0.689 ± 0.025**	**0.769 ± 0.010**	0.926 ± 0.010
CMLC	0.680 ± 0.003	0.772 ± 0.007	0.812 ± 0.013	0.677 ± 0.013	0.764 ± 0.027	0.918 ± 0.008
CMSMLC	**0.708 ± 0.017**	0.769 ± 0.015	0.799 ± 0.011	0.684 ± 0.011	0.761 ± 0.022	0.923 ± 0.012

(c) $F = 0.75$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.682 ± 0.016	0.764 ± 0.011	0.824 ± 0.013	0.682 ± 0.016	0.764 ± 0.011	0.925 ± 0.009
SimCLR	0.699 ± 0.010	0.782 ± 0.015	**0.839 ± 0.003**	0.691 ± 0.009	**0.795 ± 0.017**	0.938 ± 0.012
CMSC	0.712 ± 0.011	0.760 ± 0.032	0.835 ± 0.006	**0.704 ± 0.024**	0.780 ± 0.015	**0.941 ± 0.006**
CMLC	0.682 ± 0.011	0.769 ± 0.020	0.826 ± 0.014	0.665 ± 0.016	0.764 ± 0.020	0.930 ± 0.013
CMSMLC	**0.715 ± 0.009**	**0.789 ± 0.014**	0.820 ± 0.004	0.703 ± 0.010	0.778 ± 0.019	0.936 ± 0.007

(d) $F = 1$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.700 ± 0.019	0.771 ± 0.018	0.832 ± 0.006	0.700 ± 0.019	0.771 ± 0.018	0.937 ± 0.005
SimCLR	0.715 ± 0.005	**0.804 ± 0.020**	**0.844 ± 0.001**	0.704 ± 0.009	0.785 ± 0.025	0.938 ± 0.011
CMSC	**0.723 ± 0.004**	0.803 ± 0.033	0.841 ± 0.007	**0.724 ± 0.015**	**0.795 ± 0.009**	**0.945 ± 0.004**
CMLC	0.699 ± 0.025	0.778 ± 0.015	0.837 ± 0.006	0.707 ± 0.014	0.780 ± 0.023	0.933 ± 0.014
CMSMLC	0.719 ± 0.016	0.798 ± 0.018	0.830 ± 0.006	0.715 ± 0.014	0.775 ± 0.009	0.940 ± 0.006

E.1.2. EMBEDDING DIMENSION, $E = 64$

In this section, we show that in 20/24 (83%) of all experiments conducted, our family of contrastive learning methods outperforms the state-of-the-art method, SimCLR. This can be seen by the bold test AUC results in Table 16.

Table 16. Comparison of self-supervised methods when used as parameter initializations before fine-tuning on downstream datasets. Pre-training, fine-tuning, and evaluating multi-lead datasets* using 4 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.632 ± 0.018	0.746 ± 0.009	0.775 ± 0.010	0.632 ± 0.018	0.746 ± 0.009	0.895 ± 0.001
SimCLR	0.652 ± 0.010	0.744 ± 0.013	0.784 ± 0.022	0.641 ± 0.006	0.739 ± 0.019	$\mathbf{0.911 \pm 0.007}$
CMSC	0.663 ± 0.019	$\mathbf{0.765 \pm 0.023}$	$\mathbf{0.794 \pm 0.022}$	0.659 ± 0.032	0.755 ± 0.018	0.910 ± 0.007
CMLC	0.650 ± 0.007	0.753 ± 0.012	0.786 ± 0.008	0.644 ± 0.014	$\mathbf{0.762 \pm 0.009}$	0.904 ± 0.007
CMSMLC	$\mathbf{0.675 \pm 0.010}$	0.755 ± 0.009	0.767 ± 0.008	$\mathbf{0.660 \pm 0.012}$	0.743 ± 0.016	0.901 ± 0.003

(b) $F = 0.5$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.685 ± 0.004	0.768 ± 0.010	0.817 ± 0.009	0.685 ± 0.004	0.768 ± 0.010	0.906 ± 0.003
SimCLR	0.676 ± 0.019	0.778 ± 0.008	0.822 ± 0.011	0.678 ± 0.009	0.771 ± 0.018	0.927 ± 0.009
CMSC	0.695 ± 0.011	$\mathbf{0.786 \pm 0.017}$	0.816 ± 0.016	$\mathbf{0.701 \pm 0.023}$	0.772 ± 0.009	0.928 ± 0.004
CMLC	0.679 ± 0.016	$\mathbf{0.775 \pm 0.010}$	$\mathbf{0.824 \pm 0.004}$	0.677 ± 0.021	$\mathbf{0.775 \pm 0.010}$	0.918 ± 0.014
CMSMLC	$\mathbf{0.717 \pm 0.005}$	0.773 ± 0.011	0.808 ± 0.009	0.699 ± 0.011	0.769 ± 0.009	$\mathbf{0.934 \pm 0.004}$

(c) $F = 0.75$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.680 ± 0.015	0.7600 ± 0.011	0.830 ± 0.007	0.680 ± 0.015	0.7600 ± 0.011	0.916 ± 0.011
SimCLR	0.698 ± 0.008	0.790 ± 0.010	0.832 ± 0.007	0.689 ± 0.016	$\mathbf{0.783 \pm 0.015}$	0.934 ± 0.006
CMSC	0.708 ± 0.006	0.790 ± 0.027	0.834 ± 0.004	$\mathbf{0.715 \pm 0.008}$	0.779 ± 0.013	$\mathbf{0.940 \pm 0.007}$
CMLC	0.688 ± 0.016	0.777 ± 0.017	$\mathbf{0.837 \pm 0.003}$	0.678 ± 0.019	0.777 ± 0.011	0.926 ± 0.016
CMSMLC	$\mathbf{0.720 \pm 0.004}$	$\mathbf{0.795 \pm 0.008}$	0.829 ± 0.009	0.704 ± 0.007	0.775 ± 0.013	0.932 ± 0.005

(d) $F = 1$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.692 ± 0.023	0.778 ± 0.017	0.840 ± 0.004	0.692 ± 0.023	0.778 ± 0.017	0.932 ± 0.008
SimCLR	0.718 ± 0.010	0.797 ± 0.014	0.837 ± 0.009	0.706 ± 0.008	0.789 ± 0.023	$\mathbf{0.944 \pm 0.004}$
CMSC	0.708 ± 0.022	0.800 ± 0.020	$\mathbf{0.842 \pm 0.003}$	$\mathbf{0.726 \pm 0.007}$	$\mathbf{0.797 \pm 0.009}$	0.941 ± 0.005
CMLC	0.691 ± 0.007	0.780 ± 0.013	0.841 ± 0.003	0.696 ± 0.022	0.786 ± 0.016	0.931 ± 0.018
CMSMLC	$\mathbf{0.732 \pm 0.003}$	$\mathbf{0.810 \pm 0.012}$	0.837 ± 0.010	0.722 ± 0.010	0.792 ± 0.006	0.940 ± 0.007

E.1.3. EMBEDDING DIMENSION, $E = 128$

In this section, we show that in 21/24 (88%) of all experiments conducted, our family of contrastive learning methods outperforms the state-of-the-art method, SimCLR. This can be seen by the bold test AUC results in Table 17.

Table 17. Comparison of self-supervised methods when used as parameter initializations before fine-tuning on downstream datasets. Pre-training, fine-tuning, and evaluating multi-lead datasets* using 4 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.625 ± 0.015	0.746 ± 0.006	0.764 ± 0.016	0.625 ± 0.015	0.746 ± 0.006	0.894 ± 0.002
Supervised	0.671 ± 0.009	0.786 ± 0.012	0.804 ± 0.005	0.679 ± 0.011	0.805 ± 0.005	0.942 ± 0.011
Self-supervised Pre-training						
BYOL	0.620 ± 0.013	0.726 ± 0.013	0.764 ± 0.013	0.624 ± 0.021	0.752 ± 0.011	0.904 ± 0.006
SimCLR	0.634 ± 0.014	0.738 ± 0.006	0.777 ± 0.015	0.631 ± 0.022	0.727 ± 0.014	0.903 ± 0.007
CMSC	$\mathbf{0.691 \pm 0.015}$	$\mathbf{0.768 \pm 0.005}$	$\mathbf{0.813 \pm 0.007}$	$\mathbf{0.671 \pm 0.018}$	$\mathbf{0.756 \pm 0.009}$	$\mathbf{0.911 \pm 0.016}$
CMLC	0.639 ± 0.010	0.745 ± 0.012	0.770 ± 0.006	0.641 ± 0.014	0.746 ± 0.014	0.897 ± 0.003
CMSMLC	0.671 ± 0.016	0.755 ± 0.011	0.781 ± 0.012	0.668 ± 0.011	0.751 ± 0.007	0.903 ± 0.009

(b) $F = 0.5$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.678 ± 0.011	0.763 ± 0.005	0.803 ± 0.008	0.678 ± 0.011	0.763 ± 0.005	0.907 ± 0.006
Supervised	0.684 ± 0.015	0.799 ± 0.008	0.827 ± 0.001	0.730 ± 0.002	0.810 ± 0.009	0.954 ± 0.003
Self-supervised Pre-training						
BYOL	0.678 ± 0.021	0.748 ± 0.014	0.802 ± 0.013	0.674 ± 0.022	0.757 ± 0.01	0.916 ± 0.009
SimCLR	0.676 ± 0.011	0.772 ± 0.010	0.823 ± 0.011	0.658 ± 0.027	0.762 ± 0.009	0.923 ± 0.010
CMSC	0.695 ± 0.024	0.773 ± 0.013	$\mathbf{0.830 \pm 0.002}$	$\mathbf{0.714 \pm 0.014}$	0.760 ± 0.013	$\mathbf{0.932 \pm 0.008}$
CMLC	0.665 ± 0.016	0.767 ± 0.013	0.810 ± 0.011	0.675 ± 0.013	0.762 ± 0.007	0.910 ± 0.012
CMSMLC	$\mathbf{0.717 \pm 0.006}$	$\mathbf{0.774 \pm 0.004}$	0.814 ± 0.009	0.698 ± 0.011	$\mathbf{0.774 \pm 0.012}$	0.930 ± 0.012

(c) $F = 0.75$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.675 ± 0.020	0.775 ± 0.005	0.831 ± 0.011	0.675 ± 0.020	0.775 ± 0.005	0.937 ± 0.008
Supervised	0.712 ± 0.017	0.799 ± 0.014	0.837 ± 0.005	0.731 ± 0.007	0.815 ± 0.007	0.958 ± 0.004
Self-supervised Pre-training						
BYOL	0.671 ± 0.022	0.754 ± 0.009	0.825 ± 0.009	0.700 ± 0.02	0.751 ± 0.033	0.930 ± 0.005
SimCLR	0.694 ± 0.019	0.776 ± 0.013	0.834 ± 0.009	0.686 ± 0.019	$\mathbf{0.785 \pm 0.011}$	0.931 ± 0.013
CMSC	0.700 ± 0.012	$\mathbf{0.801 \pm 0.013}$	$\mathbf{0.840 \pm 0.004}$	0.707 ± 0.015	0.777 ± 0.016	$\mathbf{0.942 \pm 0.012}$
CMLC	0.670 ± 0.019	0.771 ± 0.010	0.831 ± 0.004	0.682 ± 0.005	0.772 ± 0.009	0.917 ± 0.011
CMSMLC	$\mathbf{0.719 \pm 0.011}$	0.792 ± 0.014	0.837 ± 0.008	$\mathbf{0.711 \pm 0.011}$	0.777 ± 0.017	0.938 ± 0.010

(d) $F = 1$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.702 ± 0.016	0.773 ± 0.010	0.843 ± 0.002	0.702 ± 0.016	0.773 ± 0.010	0.930 ± 0.013
Supervised	0.712 ± 0.017	0.799 ± 0.011	0.844 ± 0.003	0.732 ± 0.008	0.821 ± 0.006	0.961 ± 0.004
Self-supervised Pre-training						
BYOL	0.697 ± 0.006	0.774 ± 0.017	0.834 ± 0.011	0.709 ± 0.017	0.771 ± 0.022	0.935 ± 0.008
SimCLR	0.705 ± 0.008	$\mathbf{0.810 \pm 0.016}$	0.844 ± 0.005	0.700 ± 0.012	$\mathbf{0.795 \pm 0.021}$	0.941 ± 0.006
CMSC	0.715 ± 0.018	0.804 ± 0.018	$\mathbf{0.846 \pm 0.002}$	$\mathbf{0.725 \pm 0.020}$	0.779 ± 0.024	0.942 ± 0.009
CMLC	0.698 ± 0.007	0.781 ± 0.014	0.836 ± 0.003	0.681 ± 0.005	0.785 ± 0.011	0.933 ± 0.014
CMSMLC	$\mathbf{0.732 \pm 0.003}$	0.793 ± 0.012	0.844 ± 0.005	0.716 ± 0.010	0.778 ± 0.025	$\mathbf{0.945 \pm 0.005}$

E.1.4. EMBEDDING DIMENSION, $E = 256$

In this section, we show that in 16/24 (66%) of all experiments conducted, our family of contrastive learning methods outperforms the state-of-the-art method, SimCLR. This can be seen by the bold test AUC results in Table 18.

Table 18. Comparison of self-supervised methods when used as parameter initializations before fine-tuning on downstream datasets. Pre-training, fine-tuning, and evaluating multi-lead datasets* using 4 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.630 ± 0.014	0.737 ± 0.008	0.765 ± 0.004	0.630 ± 0.014	0.737 ± 0.008	0.896 ± 0.002
SimCLR	0.647 ± 0.014	0.727 ± 0.007	$\mathbf{0.791 \pm 0.014}$	0.636 ± 0.009	0.736 ± 0.008	0.902 ± 0.006
CMSC	0.656 ± 0.031	$\mathbf{0.756 \pm 0.011}$	0.789 ± 0.019	$\mathbf{0.682 \pm 0.024}$	0.750 ± 0.014	$\mathbf{0.905 \pm 0.009}$
CMLC	0.649 ± 0.012	0.743 ± 0.005	0.784 ± 0.009	0.645 ± 0.017	0.741 ± 0.008	0.898 ± 0.004
CMSMLC	$\mathbf{0.686 \pm 0.008}$	0.752 ± 0.010	0.768 ± 0.017	0.652 ± 0.023	$\mathbf{0.758 \pm 0.014}$	0.896 ± 0.002

(b) $F = 0.5$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.659 ± 0.012	0.758 ± 0.021	0.817 ± 0.008	0.659 ± 0.012	0.758 ± 0.021	0.901 ± 0.003
SimCLR	0.667 ± 0.019	0.758 ± 0.002	0.825 ± 0.014	0.659 ± 0.010	$\mathbf{0.769 \pm 0.017}$	$\mathbf{0.924 \pm 0.012}$
CMSC	0.667 ± 0.030	0.765 ± 0.003	0.819 ± 0.002	$\mathbf{0.709 \pm 0.028}$	0.762 ± 0.015	0.914 ± 0.011
CMLC	0.679 ± 0.014	0.768 ± 0.006	$\mathbf{0.826 \pm 0.005}$	0.669 ± 0.027	0.768 ± 0.012	0.906 ± 0.007
CMSMLC	$\mathbf{0.702 \pm 0.017}$	$\mathbf{0.776 \pm 0.011}$	0.812 ± 0.014	0.694 ± 0.011	0.762 ± 0.009	0.917 ± 0.011

(c) $F = 0.75$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.680 ± 0.018	0.764 ± 0.006	0.834 ± 0.004	0.680 ± 0.018	0.764 ± 0.006	0.916 ± 0.015
SimCLR	0.677 ± 0.016	$\mathbf{0.790 \pm 0.015}$	0.834 ± 0.011	0.684 ± 0.008	$\mathbf{0.787 \pm 0.015}$	0.933 ± 0.013
CMSC	0.698 ± 0.015	0.784 ± 0.015	0.827 ± 0.014	$\mathbf{0.717 \pm 0.010}$	0.780 ± 0.018	$\mathbf{0.935 \pm 0.007}$
CMLC	0.677 ± 0.023	0.773 ± 0.006	$\mathbf{0.841 \pm 0.001}$	0.681 ± 0.012	0.779 ± 0.012	0.917 ± 0.012
CMSMLC	$\mathbf{0.715 \pm 0.008}$	0.785 ± 0.004	0.827 ± 0.015	0.697 ± 0.016	0.784 ± 0.010	0.930 ± 0.004

(d) $F = 1$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.696 ± 0.015	0.763 ± 0.012	0.842 ± 0.005	0.696 ± 0.015	0.763 ± 0.012	0.918 ± 0.015
SimCLR	0.711 ± 0.008	$\mathbf{0.798 \pm 0.014}$	0.841 ± 0.006	0.703 ± 0.007	$\mathbf{0.806 \pm 0.012}$	0.943 ± 0.004
CMSC	0.704 ± 0.023	0.794 ± 0.018	0.840 ± 0.007	$\mathbf{0.718 \pm 0.012}$	0.792 ± 0.016	$\mathbf{0.944 \pm 0.007}$
CMLC	0.705 ± 0.009	0.781 ± 0.005	$\mathbf{0.844 \pm 0.002}$	0.690 ± 0.020	0.779 ± 0.007	0.926 ± 0.015
CMSMLC	$\mathbf{0.731 \pm 0.007}$	0.789 ± 0.020	0.839 ± 0.007	0.709 ± 0.013	0.791 ± 0.007	0.943 ± 0.004

E.2. Pre-training, Fine-tuning, and Evaluating using 12 Leads

We present Tables 19 - 22 which illustrate the test AUC on downstream datasets after having pre-trained on Chapman or PhysioNet 2020 using all 12 leads. These are shown for a range of embedding dimensions, $E = (32, 64, 128, 256)$, and available labelled training data, $F = (0.25, 0.50, 0.75, 1.00)$. Overall, we find that encouraging the representations of a large and diverse set of leads to be similar to one another might be detrimental. This is shown in the subsequent sections by the consistently poorer performance ($\downarrow$ AUC) of CMLC and CMSMLC relative to CMSC where the latter method does not enforce the aforementioned similarity.

E.2.1. Embedding Dimension, $E = 32$

In this section, we show that in 22/24 (92%) of all experiments conducted, CMSC outperforms the state-of-the-art method, SimCLR. This can be seen by the bold test AUC results in Table 19. Such a finding illustrates the robustness of our methods to the pre-training and downstream dataset used for evaluation, especially given the diversity of the tasks at hand.

The performance gap between CMSC and SimCLR widens as the fraction of available labelled training data decreases. For instance, when evaluating on the Cardiology dataset, as $F = 1 \rightarrow 0.25$, CMSC's $\text{AUC} = 0.723 \rightarrow 0.689$ whereas SimCLR's $\text{AUC} = 0.694 \rightarrow 0.636$. Therefore, the performance gap widens by almost a factor of 2 from 2.9% to 5.3%. This suggests that CMSC is better equipped to deal with downstream tasks that lack a sufficient amount of labelled data.

Table 19. Comparison of self-supervised methods when used as parameter initializations before fine-tuning on downstream datasets. Pre-training, fine-tuning, and evaluating multi-lead datasets* using all 12 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.631 ± 0.006	0.738 ± 0.014	0.823 ± 0.007	0.631 ± 0.006	0.738 ± 0.014	0.907 ± 0.006
SimCLR	0.636 ± 0.019	0.724 ± 0.016	0.826 ± 0.011	0.616 ± 0.011	0.727 ± 0.020	0.921 ± 0.011
CMSC	$\mathbf{0.689 \pm 0.017}$	$\mathbf{0.782 \pm 0.005}$	$\mathbf{0.833 \pm 0.002}$	$\mathbf{0.681 \pm 0.017}$	$\mathbf{0.769 \pm 0.015}$	$\mathbf{0.936 \pm 0.011}$
CMLC	0.639 ± 0.023	0.744 ± 0.018	0.827 ± 0.003	0.630 ± 0.022	0.744 ± 0.022	0.912 ± 0.007
CMSMLC	0.644 ± 0.026	0.740 ± 0.019	0.818 ± 0.015	0.647 ± 0.022	0.745 ± 0.015	0.920 ± 0.011

(b) $F = 0.5$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.669 ± 0.007	0.782 ± 0.011	0.814 ± 0.009	0.669 ± 0.007	0.782 ± 0.011	0.938 ± 0.009
SimCLR	0.659 ± 0.011	0.764 ± 0.003	0.820 ± 0.032	0.669 ± 0.023	0.766 ± 0.015	0.936 ± 0.014
CMSC	0.686 ± 0.024	$\mathbf{0.800 \pm 0.013}$	$\mathbf{0.836 \pm 0.004}$	$\mathbf{0.719 \pm 0.014}$	$\mathbf{0.778 \pm 0.019}$	$\mathbf{0.951 \pm 0.003}$
CMLC	0.674 ± 0.012	0.773 ± 0.018	0.831 ± 0.002	0.667 ± 0.011	0.758 ± 0.017	0.933 ± 0.008
CMSMLC	$\mathbf{0.691 \pm 0.007}$	0.759 ± 0.020	0.831 ± 0.009	0.684 ± 0.028	0.763 ± 0.024	0.942 ± 0.005

(c) $F = 0.75$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.682 ± 0.016	0.764 ± 0.011	0.845 ± 0.001	0.682 ± 0.016	0.764 ± 0.011	0.937 ± 0.016
SimCLR	0.690 ± 0.023	0.786 ± 0.023	0.840 ± 0.006	0.668 ± 0.013	0.782 ± 0.007	0.945 ± 0.009
CMSC	0.702 ± 0.013	$\mathbf{0.809 \pm 0.009}$	$\mathbf{0.847 \pm 0.001}$	$\mathbf{0.709 \pm 0.010}$	$\mathbf{0.806 \pm 0.005}$	$\mathbf{0.952 \pm 0.010}$
CMLC	0.684 ± 0.027	0.774 ± 0.019	0.841 ± 0.015	0.680 ± 0.021	0.783 ± 0.018	0.933 ± 0.014
CMSMLC	$\mathbf{0.719 \pm 0.007}$	0.757 ± 0.025	0.843 ± 0.005	0.708 ± 0.011	0.787 ± 0.015	0.944 ± 0.006

(d) $F = 1$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.700 ± 0.019	0.771 ± 0.018	0.825 ± 0.016	0.700 ± 0.019	0.771 ± 0.018	0.945 ± 0.003
SimCLR	0.694 ± 0.010	0.790 ± 0.022	0.839 ± 0.008	0.691 ± 0.009	0.790 ± 0.020	0.942 ± 0.014
CMSC	$\mathbf{0.723 \pm 0.011}$	$\mathbf{0.821 \pm 0.013}$	$\mathbf{0.845 \pm 0.003}$	$\mathbf{0.725 \pm 0.017}$	$\mathbf{0.798 \pm 0.008}$	$\mathbf{0.954 \pm 0.007}$
CMLC	0.702 ± 0.011	0.762 ± 0.014	0.844 ± 0.003	0.708 ± 0.026	0.777 ± 0.019	0.948 ± 0.005
CMSMLC	0.722 ± 0.007	0.782 ± 0.013	0.845 ± 0.005	0.710 ± 0.020	0.768 ± 0.033	0.946 ± 0.005

E.2.2. EMBEDDING DIMENSION, $E = 64$

In this section, we show that in 21/24 (88%) of all experiments conducted, CMSC outperforms the state-of-the-art method, SimCLR. This can be seen by the bold test AUC results in Table 20.

Table 20. Comparison of self-supervised methods when used as parameter initializations before fine-tuning on downstream datasets. Pre-training, fine-tuning, and evaluating multi-lead datasets* using all 12 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.632 ± 0.018	0.746 ± 0.009	0.822 ± 0.011	0.632 ± 0.018	0.746 ± 0.009	0.901 ± 0.004
SimCLR	0.632 ± 0.021	0.736 ± 0.019	0.833 ± 0.008	0.626 ± 0.008	0.734 ± 0.018	0.925 ± 0.013
CMSC	$\mathbf{0.681 \pm 0.024}$	$\mathbf{0.798 \pm 0.008}$	$\mathbf{0.834 \pm 0.006}$	$\mathbf{0.658 \pm 0.026}$	$\mathbf{0.779 \pm 0.012}$	$\mathbf{0.942 \pm 0.011}$
CMLC	0.626 ± 0.025	0.735 ± 0.011	0.825 ± 0.004	0.627 ± 0.016	0.739 ± 0.014	0.910 ± 0.007
CMSMLC	0.659 ± 0.024	0.738 ± 0.013	0.820 ± 0.016	0.647 ± 0.023	0.743 ± 0.012	0.912 ± 0.009

(b) $F = 0.5$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.685 ± 0.004	0.768 ± 0.010	0.831 ± 0.007	0.685 ± 0.004	0.768 ± 0.01	0.931 ± 0.016
SimCLR	0.672 ± 0.023	0.762 ± 0.021	0.833 ± 0.011	0.681 ± 0.011	0.767 ± 0.012	0.943 ± 0.006
CMSC	$\mathbf{0.708 \pm 0.010}$	$\mathbf{0.804 \pm 0.011}$	$\mathbf{0.834 \pm 0.010}$	$\mathbf{0.709 \pm 0.013}$	$\mathbf{0.792 \pm 0.015}$	$\mathbf{0.954 \pm 0.005}$
CMLC	0.680 ± 0.017	0.763 ± 0.010	0.832 ± 0.005	0.694 ± 0.019	0.748 ± 0.023	0.933 ± 0.009
CMSMLC	0.706 ± 0.007	0.759 ± 0.014	0.815 ± 0.025	0.699 ± 0.023	0.753 ± 0.017	0.940 ± 0.008

(c) $F = 0.75$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.68 ± 0.015	0.76 ± 0.011	0.841 ± 0.008	0.680 ± 0.015	0.760 ± 0.011	0.937 ± 0.009
SimCLR	0.695 ± 0.023	0.779 ± 0.012	0.844 ± 0.007	0.674 ± 0.017	0.775 ± 0.011	0.948 ± 0.009
CMSC	0.709 ± 0.014	$\mathbf{0.809 \pm 0.014}$	0.844 ± 0.007	$\mathbf{0.714 \pm 0.017}$	$\mathbf{0.802 \pm 0.012}$	$\mathbf{0.953 \pm 0.006}$
CMLC	0.690 ± 0.007	0.778 ± 0.010	$\mathbf{0.844 \pm 0.002}$	0.704 ± 0.021	0.768 ± 0.018	0.946 ± 0.003
CMSMLC	$\mathbf{0.711 \pm 0.011}$	0.763 ± 0.016	0.838 ± 0.006	0.689 ± 0.022	0.762 ± 0.019	0.946 ± 0.008

(d) $F = 1$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.692 ± 0.023	0.778 ± 0.017	0.846 ± 0.003	0.692 ± 0.023	0.778 ± 0.017	0.946 ± 0.005
SimCLR	0.715 ± 0.011	0.808 ± 0.009	0.842 ± 0.007	0.703 ± 0.006	0.797 ± 0.018	0.952 ± 0.008
CMSC	$\mathbf{0.736 \pm 0.016}$	$\mathbf{0.810 \pm 0.005}$	0.843 ± 0.005	$\mathbf{0.731 \pm 0.010}$	$\mathbf{0.810 \pm 0.015}$	$\mathbf{0.958 \pm 0.007}$
CMLC	0.706 ± 0.012	0.777 ± 0.017	$\mathbf{0.846 \pm 0.002}$	0.709 ± 0.012	0.779 ± 0.018	0.947 ± 0.005
CMSMLC	0.722 ± 0.008	0.780 ± 0.015	0.842 ± 0.008	0.701 ± 0.023	0.779 ± 0.015	0.943 ± 0.009

E.2.3. EMBEDDING DIMENSION, $E = 128$

In this section, we show that in 24/24 (100%) of all experiments conducted, CMSC outperforms the the state-of-the-art method, SimCLR. This can be seen by the bold test AUC results in Table 21.

Table 21. Comparison of self-supervised methods when used as parameter initializations before fine-tuning on downstream datasets. Pre-training, fine-tuning, and evaluating multi-lead datasets* using all 12 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.625 ± 0.015	0.746 ± 0.006	0.819 ± 0.008	0.625 ± 0.015	0.746 ± 0.006	0.909 ± 0.006
SimCLR	0.630 ± 0.011	0.735 ± 0.012	0.833 ± 0.008	0.624 ± 0.007	0.729 ± 0.018	0.918 ± 0.015
CMSC	$\mathbf{0.678 \pm 0.010}$	$\mathbf{0.790 \pm 0.012}$	$\mathbf{0.833 \pm 0.008}$	$\mathbf{0.680 \pm 0.011}$	$\mathbf{0.777 \pm 0.027}$	$\mathbf{0.940 \pm 0.007}$
CMLC	0.639 ± 0.012	0.740 ± 0.007	0.831 ± 0.003	0.639 ± 0.019	0.743 ± 0.016	0.913 ± 0.012
CMSMLC	0.661 ± 0.029	0.748 ± 0.005	0.813 ± 0.024	0.646 ± 0.023	0.736 ± 0.007	0.918 ± 0.012

(b) $F = 0.5$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.678 ± 0.011	0.763 ± 0.005	0.832 ± 0.003	0.678 ± 0.011	0.763 ± 0.005	0.931 ± 0.014
SimCLR	0.667 ± 0.021	0.768 ± 0.012	0.835 ± 0.010	0.659 ± 0.012	0.754 ± 0.024	0.939 ± 0.007
CMSC	$\mathbf{0.716 \pm 0.010}$	$\mathbf{0.802 \pm 0.007}$	$\mathbf{0.840 \pm 0.003}$	$\mathbf{0.718 \pm 0.005}$	$\mathbf{0.791 \pm 0.025}$	$\mathbf{0.944 \pm 0.008}$
CMLC	0.690 ± 0.012	0.763 ± 0.009	0.840 ± 0.003	0.663 ± 0.040	0.752 ± 0.016	0.927 ± 0.013
CMSMLC	0.699 ± 0.013	0.751 ± 0.013	0.815 ± 0.014	0.695 ± 0.020	0.748 ± 0.013	0.931 ± 0.011

(c) $F = 0.75$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.675 ± 0.020	0.775 ± 0.005	0.844 ± 0.006	0.675 ± 0.020	0.775 ± 0.005	0.945 ± 0.004
SimCLR	0.682 ± 0.023	0.775 ± 0.009	0.843 ± 0.007	0.681 ± 0.020	0.764 ± 0.019	0.946 ± 0.010
CMSC	$\mathbf{0.719 \pm 0.008}$	$\mathbf{0.813 \pm 0.006}$	$\mathbf{0.847 \pm 0.002}$	$\mathbf{0.711 \pm 0.004}$	$\mathbf{0.810 \pm 0.020}$	$\mathbf{0.955 \pm 0.005}$
CMLC	0.684 ± 0.008	0.777 ± 0.021	0.846 ± 0.001	0.700 ± 0.016	0.755 ± 0.016	0.942 ± 0.005
CMSMLC	0.711 ± 0.011	0.782 ± 0.006	0.839 ± 0.007	0.694 ± 0.028	0.769 ± 0.014	0.941 ± 0.007

(d) $F = 1$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.702 ± 0.016	0.773 ± 0.01	0.842 ± 0.008	0.702 ± 0.016	0.773 ± 0.01	0.942 ± 0.006
SimCLR	0.703 ± 0.020	0.801 ± 0.014	0.845 ± 0.009	0.703 ± 0.014	0.784 ± 0.009	0.948 ± 0.008
CMSC	$\mathbf{0.731 \pm 0.022}$	$\mathbf{0.819 \pm 0.004}$	$\mathbf{0.847 \pm 0.003}$	$\mathbf{0.718 \pm 0.012}$	$\mathbf{0.809 \pm 0.021}$	$\mathbf{0.959 \pm 0.004}$
CMLC	0.705 ± 0.010	0.777 ± 0.011	0.845 ± 0.002	0.713 ± 0.023	0.789 ± 0.012	0.946 ± 0.005
CMSMLC	0.719 ± 0.005	0.764 ± 0.010	0.837 ± 0.007	0.711 ± 0.013	0.779 ± 0.013	0.947 ± 0.003

E.2.4. EMBEDDING DIMENSION, $E = 256$

In this section, we show that in 22/24 (92%) of all experiments conducted, CMSC outperforms the state-of-the-art method, SimCLR. This can be seen by the bold test AUC results in Table 22.

Table 22. Comparison of self-supervised methods when used as parameter initializations before fine-tuning on downstream datasets. Pre-training, fine-tuning, and evaluating multi-lead datasets* using all 12 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.630 ± 0.014	0.737 ± 0.008	0.809 ± 0.023	0.630 ± 0.014	0.737 ± 0.008	0.903 ± 0.005
SimCLR	0.620 ± 0.028	0.729 ± 0.013	0.830 ± 0.007	0.621 ± 0.016	0.726 ± 0.008	0.933 ± 0.007
CMSC	$\mathbf{0.692 \pm 0.007}$	$\mathbf{0.792 \pm 0.014}$	$\mathbf{0.832 \pm 0.009}$	$\mathbf{0.689 \pm 0.013}$	$\mathbf{0.782 \pm 0.010}$	$\mathbf{0.940 \pm 0.010}$
CMLC	0.618 ± 0.004	0.733 ± 0.006	0.831 ± 0.009	0.648 ± 0.018	0.743 ± 0.010	0.912 ± 0.006
CMSMLC	0.666 ± 0.012	0.741 ± 0.010	0.820 ± 0.013	0.666 ± 0.008	0.736 ± 0.012	0.922 ± 0.011

(b) $F = 0.5$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.659 ± 0.012	0.758 ± 0.021	0.831 ± 0.011	0.659 ± 0.012	0.758 ± 0.021	0.929 ± 0.010
SimCLR	0.670 ± 0.021	0.764 ± 0.008	0.830 ± 0.011	0.663 ± 0.007	0.762 ± 0.009	0.942 ± 0.005
CMSC	$\mathbf{0.706 \pm 0.024}$	$\mathbf{0.809 \pm 0.004}$	0.835 ± 0.009	$\mathbf{0.714 \pm 0.006}$	$\mathbf{0.798 \pm 0.009}$	$\mathbf{0.953 \pm 0.007}$
CMLC	0.668 ± 0.006	0.762 ± 0.005	$\mathbf{0.837 \pm 0.007}$	0.700 ± 0.013	0.768 ± 0.011	0.935 ± 0.010
CMSMLC	0.704 ± 0.012	0.763 ± 0.009	0.829 ± 0.009	0.713 ± 0.006	0.748 ± 0.011	0.940 ± 0.003

(c) $F = 0.75$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.680 ± 0.018	0.764 ± 0.006	0.844 ± 0.004	0.68 ± 0.018	0.764 ± 0.006	0.936 ± 0.014
SimCLR	0.675 ± 0.016	0.782 ± 0.014	0.842 ± 0.007	0.678 ± 0.007	0.786 ± 0.010	0.953 ± 0.002
CMSC	$\mathbf{0.714 \pm 0.013}$	$\mathbf{0.816 \pm 0.003}$	$\mathbf{0.843 \pm 0.006}$	$\mathbf{0.722 \pm 0.015}$	$\mathbf{0.805 \pm 0.011}$	$\mathbf{0.958 \pm 0.003}$
CMLC	0.678 ± 0.011	0.777 ± 0.009	0.841 ± 0.006	0.705 ± 0.013	0.775 ± 0.013	0.936 ± 0.008
CMSMLC	0.701 ± 0.014	0.775 ± 0.009	0.842 ± 0.007	0.705 ± 0.004	0.763 ± 0.009	0.946 ± 0.005

(d) $F = 1$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
Random Init.	0.696 ± 0.015	0.763 ± 0.012	0.839 ± 0.009	0.696 ± 0.015	0.763 ± 0.012	0.943 ± 0.002
SimCLR	0.708 ± 0.019	0.789 ± 0.006	$\mathbf{0.844 \pm 0.007}$	0.707 ± 0.009	0.792 ± 0.008	0.951 ± 0.005
CMSC	$\mathbf{0.735 \pm 0.006}$	$\mathbf{0.822 \pm 0.004}$	0.843 ± 0.006	$\mathbf{0.729 \pm 0.010}$	$\mathbf{0.807 \pm 0.012}$	$\mathbf{0.957 \pm 0.004}$
CMLC	0.705 ± 0.006	0.795 ± 0.014	0.843 ± 0.007	0.719 ± 0.003	0.793 ± 0.016	0.942 ± 0.006
CMSMLC	0.722 ± 0.008	0.778 ± 0.013	0.842 ± 0.008	0.722 ± 0.003	0.767 ± 0.013	0.946 ± 0.003

F. Effect of τ_d on BYOL Implementation

In the BYOL implementation, two networks exist; an online network and a target network. The latter is a delayed version of the former where its parameters are an exponential moving average of the parameters of the online network. This exponential moving average is a function of the hyperparameter, τ_d. In this section, we outline the effect of τ_d on the downstream generalization performance of networks both in the linear and transfer evaluation scenarios. This can be found in Figs 23 and 24, respectively. We find that the results associated with $\tau_d = 0.900$ lead to the best performance, and are thus quoted in the main manuscript.

F.1. Linear Evaluation of Representations

Table 23. Effect of the value of τ_d during BYOL pre-training on the downstream generalization performance of a linear evaluation scenario. Pre-training, fine-tuning, and evaluating multi-lead datasets* using 4 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Dataset	Chapman*	PhysioNet 2020*
$\tau_d = 0.500$	0.602 ± 0.072	0.581 ± 0.010
$\tau_d = 0.900$	$\mathbf{0.671 \pm 0.042}$	$\mathbf{0.587 \pm 0.021}$
$\tau_d = 0.990$	0.597 ± 0.068	0.571 ± 0.028

(b) $F = 0.5$

Dataset	Chapman*	PhysioNet 2020*
$\tau_d = 0.500$	0.618 ± 0.087	0.590 ± 0.010
$\tau_d = 0.900$	$\mathbf{0.643 \pm 0.043}$	$\mathbf{0.595 \pm 0.018}$
$\tau_d = 0.990$	0.604 ± 0.079	0.578 ± 0.033

(c) $F = 0.75$

Dataset	Chapman*	PhysioNet 2020*
$\tau_d = 0.500$	0.635 ± 0.075	0.597 ± 0.008
$\tau_d = 0.900$	$\mathbf{0.666 \pm 0.032}$	$\mathbf{0.598 \pm 0.022}$
$\tau_d = 0.990$	0.613 ± 0.085	0.586 ± 0.026

(d) $F = 1$

Dataset	Chapman*	PhysioNet 2020*
$\tau_d = 0.500$	0.637 ± 0.082	0.601 ± 0.008
$\tau_d = 0.900$	$\mathbf{0.653 \pm 0.026}$	$\mathbf{0.602 \pm 0.015}$
$\tau_d = 0.990$	0.619 ± 0.088	0.592 ± 0.026

F.2. Transfer Capabilities of Representations

Table 24. Effect of the value of τ_d during BYOL pre-training on the downstream generalization performance in the fine-tuning evaluation scenario. Pre-training, fine-tuning, and evaluating multi-lead datasets* using 4 leads. Mean and standard deviation are shown across 5 seeds.

(a) $F = 0.25$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
$\tau_d = 0.500$	0.614 ± 0.026	0.738 ± 0.023	0.765 ± 0.015	0.609 ± 0.015	0.724 ± 0.027	0.900 ± 0.003
$\tau_d = 0.900$	0.620 ± 0.013	0.726 ± 0.013	0.764 ± 0.013	0.624 ± 0.021	0.752 ± 0.011	0.904 ± 0.006
$\tau_d = 0.990$	0.612 ± 0.009	0.732 ± 0.022	0.767 ± 0.018	0.617 ± 0.022	0.729 ± 0.015	0.901 ± 0.003

(b) $F = 0.5$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
$\tau_d = 0.500$	0.685 ± 0.015	0.763 ± 0.011	0.797 ± 0.019	0.658 ± 0.046	0.739 ± 0.027	0.913 ± 0.009
$\tau_d = 0.900$	0.678 ± 0.021	0.748 ± 0.014	0.802 ± 0.013	0.674 ± 0.022	0.757 ± 0.010	0.916 ± 0.009
$\tau_d = 0.990$	0.671 ± 0.013	0.748 ± 0.014	0.802 ± 0.017	0.658 ± 0.017	0.755 ± 0.021	0.910 ± 0.009

(c) $F = 0.75$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
$\tau_d = 0.500$	0.689 ± 0.007	0.766 ± 0.014	0.824 ± 0.014	0.693 ± 0.015	0.758 ± 0.030	0.919 ± 0.013
$\tau_d = 0.900$	0.671 ± 0.022	0.754 ± 0.009	0.825 ± 0.009	0.700 ± 0.020	0.751 ± 0.033	0.930 ± 0.005
$\tau_d = 0.990$	0.678 ± 0.019	0.764 ± 0.009	0.822 ± 0.011	0.662 ± 0.026	0.763 ± 0.01	0.925 ± 0.010

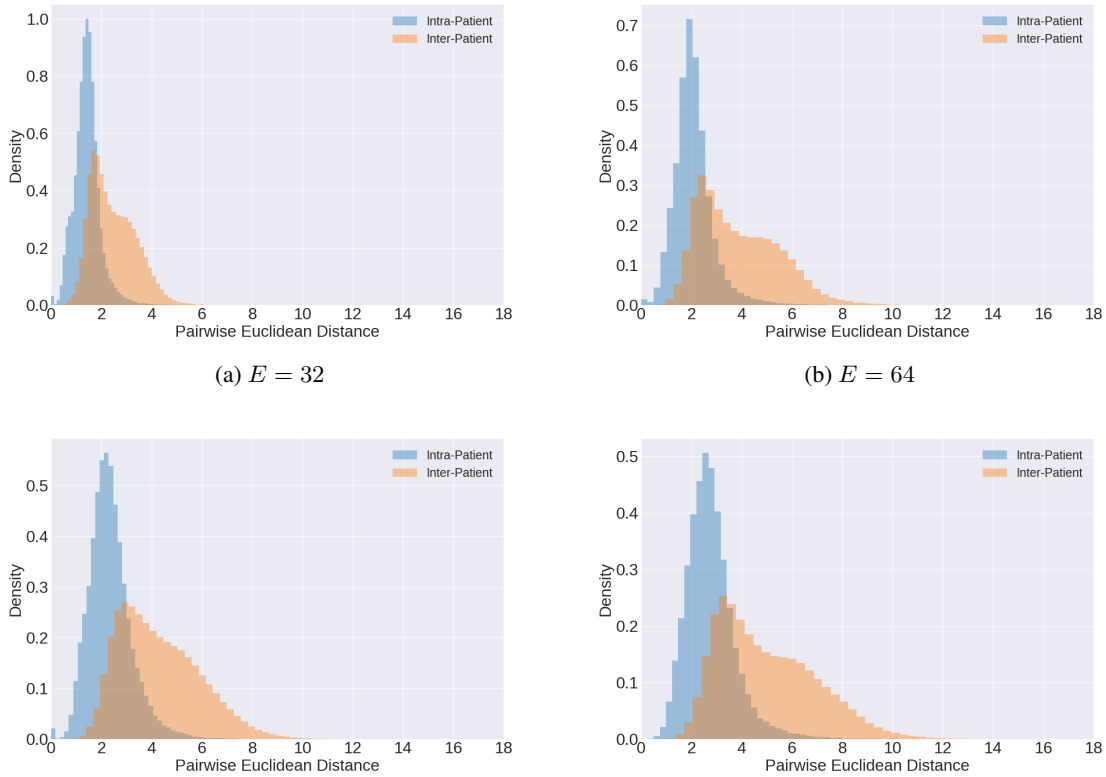
(d) $F = 1$

Pre-training Dataset	Chapman*			PhysioNet 2020*		
Downstream Dataset	Cardiology	PhysioNet 2017	PhysioNet 2020*	Cardiology	PhysioNet 2017	Chapman*
$\tau_d = 0.500$	0.709 ± 0.013	0.754 ± 0.008	0.826 ± 0.015	0.691 ± 0.038	0.770 ± 0.017	0.931 ± 0.007
$\tau_d = 0.900$	0.697 ± 0.006	0.774 ± 0.017	0.834 ± 0.011	0.709 ± 0.017	0.771 ± 0.022	0.935 ± 0.008
$\tau_d = 0.990$	0.701 ± 0.014	0.761 ± 0.020	0.833 ± 0.008	0.679 ± 0.042	0.756 ± 0.013	0.936 ± 0.011

G. Intra and Inter-Patient Representation Distances

G.1. Effect of Embedding Dimension, E, on Learning Patient-specific Representations

In Fig. 9, we show that, when using a low embedding dimension ($E = 32$), the intra-patient distances are the lowest with a mean of around 1. As $E = 32 \rightarrow 256$, the distributions begin to shift to higher values. Such high pairwise distances imply that maintaining similar representations at higher dimensions is more difficult. Moreover, we clearly see two distinct distributions belonging to intra-patient and inter-patient distances. This suggests that the training procedure worked as expected, leading to representations that are more similar within patients than across patients.



(a) $E = 32$ (b) $E = 64$

(c) $E = 128$ (d) $E = 256$

Figure 9. Distribution of pairwise Euclidean distance between representations belonging to the same patient (Intra-Patient) and those belonging to different patients (Inter-Patient). Representations are of instances present in the validation set of PhysioNet 2020. Self-supervision was performed with CMSC on PhysioNet 2020 using 4 leads.