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 (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.

PTB-XL (Wagner et al., 2020). Each ECG recording was originally 10 seconds with a sampling rate of 500Hz. We extract 5-second non-overlapping segments of each recording generating frames of length 2500 samples. We follow the diagnostic class labelling setup suggested by Wagner et al. (2020) which resulted in five classes: Conduction Disturbance (CD), Hypertrophy (HYP), Myocardial Infarction (MI), Normal (NORM), and Ischemic ST-T Changes (STTC). We alter the original setup in two main ways. Firstly, we only consider ECG segments with one label assigned to them. Secondly, we convert the task into a binary classification problem of NORM vs. (CD, HYP, MI, STTC) from above. The ECG frames were normalized in amplitude between the values of 0 and 1.

A.2 DATA SAMPLES

In this section, we outline the number of instances used during training.

Table 3: Number of instances (number of patients) used during training. These represent sample sizes for all 12 leads.

Dataset	Train	Validation	Test
PhysioNet 2020	157,188 (4,402)	37,296 (1,100)	47,460 (1,375)
Chapman	76,614 (6,387)	25,524 (2,129)	25,558 (2,130)
PTB-XL	286,632 (10,807)	36,816 (1,411)	37,008 (1,383)

B IMPLEMENTATION DETAILS

B.1 NETWORK ARCHITECTURE

In this section, we outline the architecture of the neural network used for all experiments.

Table 4: 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)

B.2 EXPERIMENT DETAILS

Table 5: 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}
PTB-XL	256	10^{-3}

C PATIENT CARDIAC PROTOTYPES ARE PATIENT-SPECIFIC

We claim that the patient cardiac prototypes that we learn in an end-to-end manner are also *patient-specific*. In the main manuscript, we provided evidence for this in the form of Euclidean distances between representations. We showed that PCPs are much closer to representations of instances that correspond to the same patient than they are to representations of instances from different patients. To determine the generalizability of these claims to other datasets, we reproduce Fig. 4 (in the main manuscript) for two additional datasets, PTB-XL and PhysioNet 2020. In Figs. 7 and 8, we illustrate the distribution of Euclidean distances between the aforementioned representations for the two datasets, respectively.

C.1 DISTRIBUTION OF PAIRWISE EUCLIDEAN DISTANCES

C.1.1 PTB-XL

Patient cardiac prototypes are indeed patient-specific. In Fig. 7, this is supported by the smaller average Euclidean distances between PCPs and representations of instances of the same patient than between PCPs and representations of instances from different patients (average Euclidean distance ≈ 7 vs. 10, respectively. The same conclusion was arrived at in the main manuscript. Furthermore, PCPs have the potential to be used for the detection of out-of-distribution data. The high overlap between the **PCP to Validation Patients** and **PCP to Different Training Patients** has a twofold implication. First, it suggests that instances in the validation set belong to patients not found in the training set (by design). Second, that patients in the validation set, on average, belong to the same overall distribution of patients.

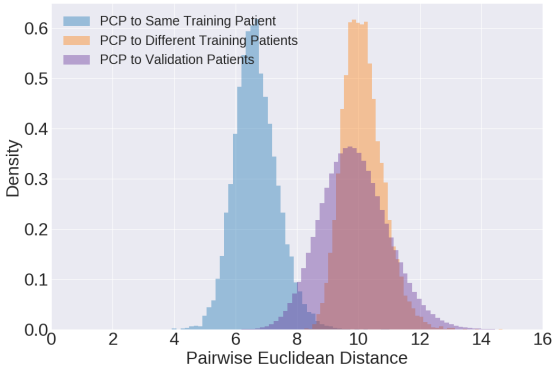


Figure 7: Distribution of pairwise Euclidean distance from the learned PCPs on the PTB-XL dataset to three sets of representations: those in the training set that belong to the same patient (blue), those in the training set that belong to different patients (orange), and those in the validation set (purple). PCPs are patient-specific since they are closer to representations belonging to the same patient than they are to representations belonging to different patients.

C.1.2 PHYSIONET 2020

As shown for the Chapman and PTB-XL datasets, we also show that PCPs are patient-specific when trained on the PhysioNet 2020 dataset. This is emphasized by the high degree of separability between the **PCP to Same Training Patient** and **PCP to Different Training Patients** distributions. We claim that the instances in the PTB-XL dataset exhibit a higher degree of diversity relative to those in other datasets. We see this by comparing the **PCP to Same Training Patient** distribution which has a larger mean (≈ 8) compared to ≈ 4 (Fig. 4) and 7 (Fig. 7) for the Chapman and PTB-XL datasets, respectively. Such a finding implies that the PCPs had a more difficult time summarizing the representations that belong to the same patient.

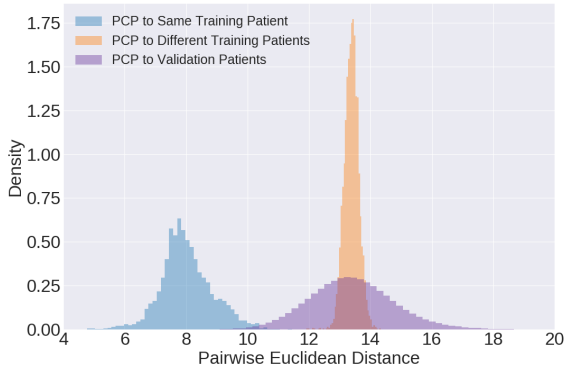


Figure 8: Distribution of pairwise Euclidean distance from the learned PCPs on the PhysioNet 2020 dataset to three sets of representations: those in the training set that belong to the same patient (blue), those in the training set that belong to different patients (orange), and those in the validation set (purple). PCPs are patient-specific since they are closer to representations belonging to the same patient than they are to representations belonging to different patients.

D DISCOVERING (DIS)SIMILAR PATIENTS

In the main manuscript, we showed that patients identified as being similar, based on the pairwise Euclidean distance between PCPs and representations of instances in the validation set, are indeed similar. In this section, we use additional datasets to further validate the role of PCPs as tools for quantifying patient similarity (Secs. D.1 and D.2) and dissimilarity (Secs. D.3 and D.4).

We quantify patient similarity by calculating the pairwise Euclidean distance between their representations. More specifically, when computing the similarity between patients in the training set, we calculate the Euclidean distance between the PCPs. In contrast, when comparing patients in the validation set to those in the training set, we calculate the Euclidean distance between the latter's PCP and the former's representations. Such pairwise distances are averaged across the multiple instances that may exist for the same patient in the validation set.

D.1 Quantification of Patient Similarity Within Same Dataset (PhysioNet 2020)

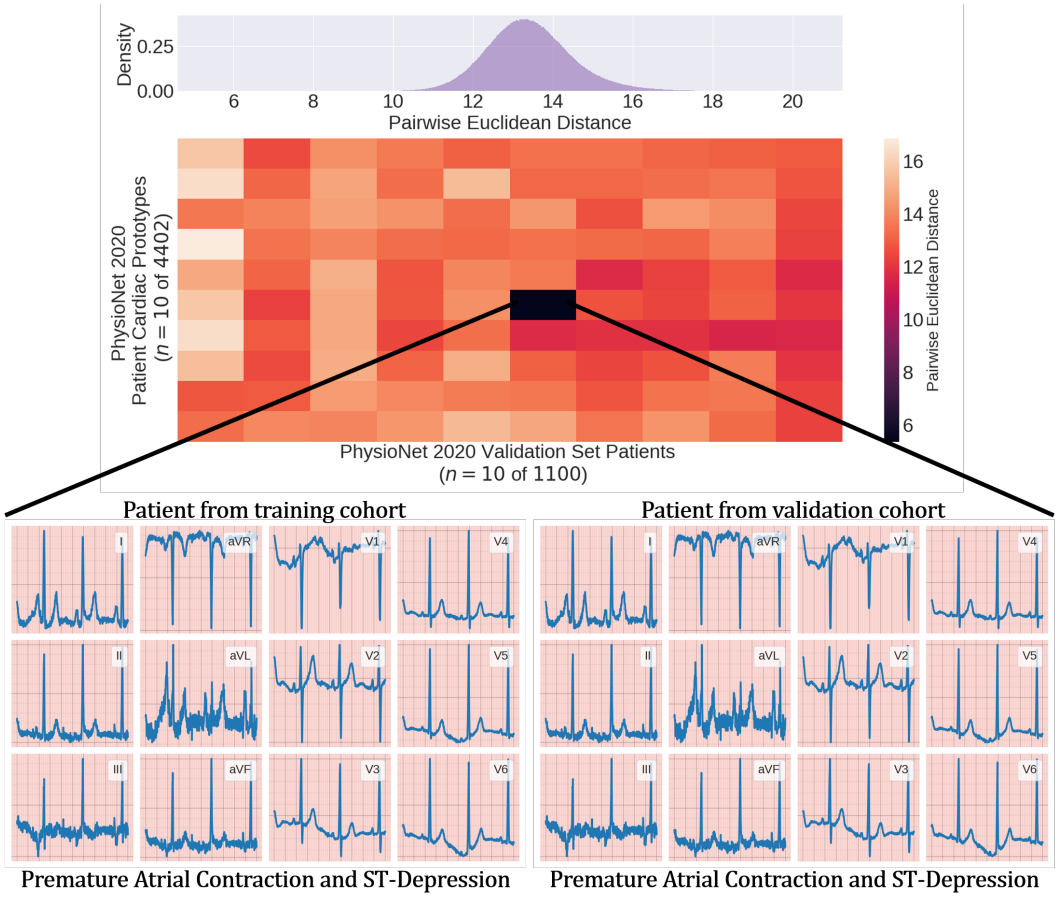


Figure 9: Identification of the two most similar patients in the training and validation sets of the PhysioNet 2020 dataset. (top) Distribution of all pairwise Euclidean distances between PCPs and representations in the validation set. (centre) Matrix illustrating pairwise distances between a subset of PCPs and representations of patients in the validation set. (bottom) Visualization of the 12-Lead ECG recordings of the two most similar patients. Both recordings are similar and correspond to the same morphology, sinus rhythm, which is considered normal.

D.2 Quantification of Patient Similarity Across Datasets (Chapman → PTB-XL)

In this section, we attempt to discover similar patients *across* datasets. To do so, we compute the pairwise Euclidean distance between the PCPs pf each dataset. The distribution of all these distances are shown in Fig. 10 (top). We also illustrate the pairwise distances for a subset of the PCPs in the Chapman and PTB-XL dataset Fig. 10 (centre). From a clinical perspective, such a matrix provides physicians with the ability to identify patients that are most similar to the one they are currently diagnosing or treating. This information can help guide future clinical intervention. By locating the pair of PCPs with the lowest Euclidean distance, we identify the pair of patients from each dataset that are most similar to one another. We visualize their corresponding 12-Lead ECG recordings in Fig. 10 (bottom).

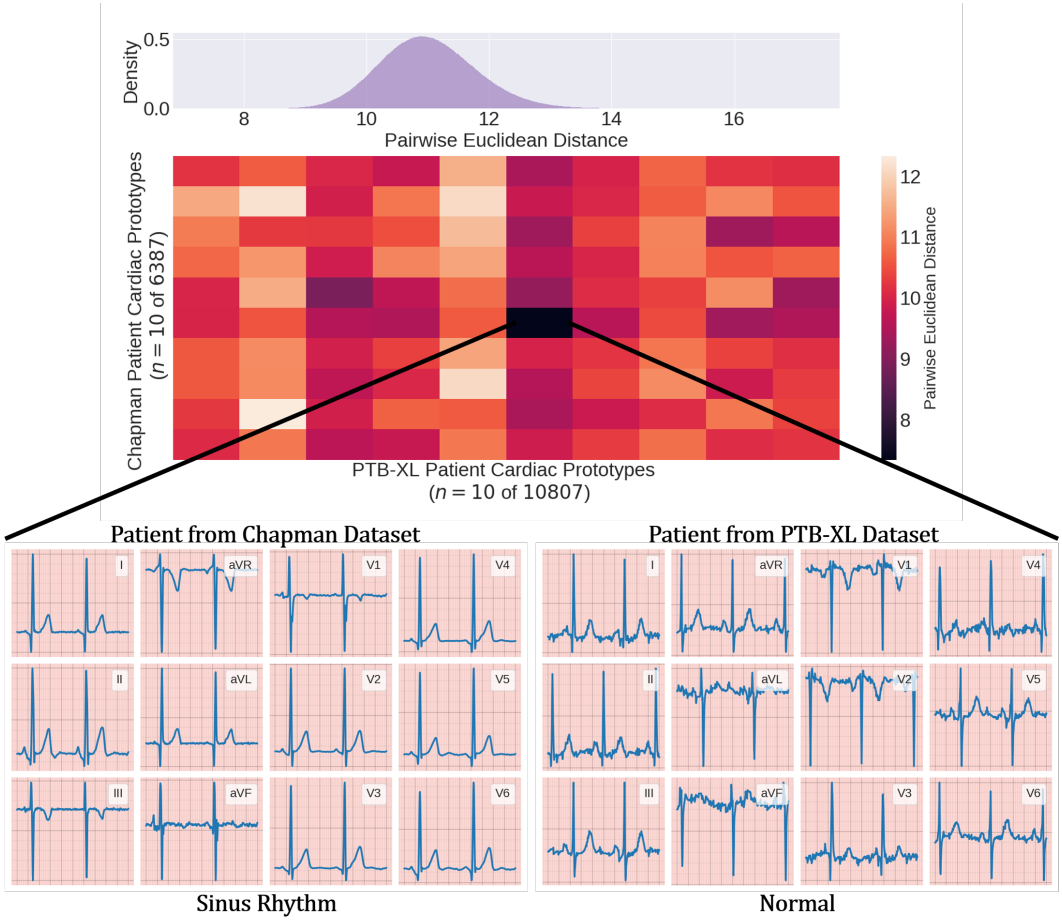


Figure 10: Identification of the two most similar patients in the training sets of the Chapman and PTB-XL dataset. (top) Distribution of all pairwise Euclidean distances between PCPs and representations in the validation set. (centre) Matrix illustrating pairwise distances between a subset of PCPs and representations of patients in the validation set. (bottom) Visualization of the 12-Lead ECG recordings of the two most similar patients. Both recordings are similar and correspond to the same morphology, sinus rhythm, which is considered normal.

D.3 VISUALIZATION OF DISSIMILAR PATIENTS WITHIN SAME DATASET

D.3.1 CHAPMAN

To further validate the PCPs and their ability to discern between patients, we use the complete version of Fig. 5 to identify two patients deemed dissimilar according to their Euclidean distance. In Fig. 11, we illustrate the 12-Lead ECG segments corresponding to these two patients. The different morphology of the ECG segments between patients and the different arrhythmia labels (Sinus Rhythm vs. Atrial Fibrillation) show that these patients are indeed dissimilar. Such a finding reaffirms the notion that PCPs both capture the cardiac state of the patient and allow for reliably patient similarity quantification.

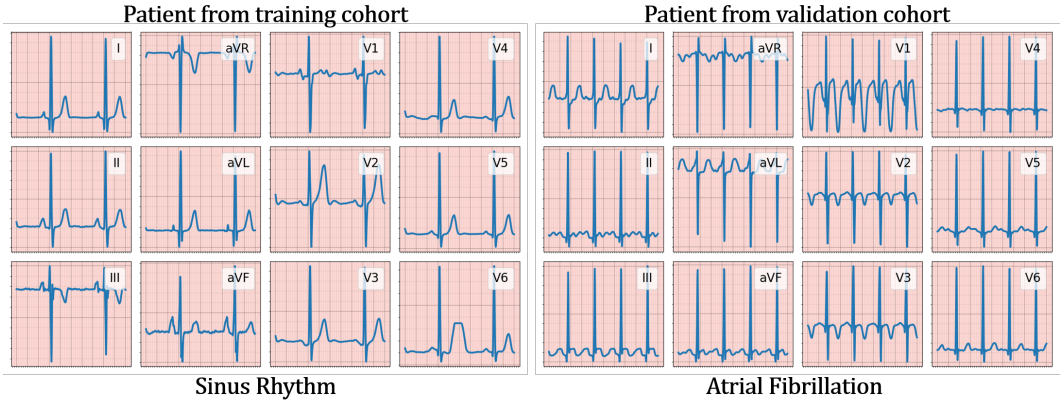


Figure 11: 12-Lead ECG segments corresponding to two dissimilar patients in the training and validation set of the Chapman dataset. Similarity is defined as low Euclidean distance between patient cardiac prototypes (PCPs) and representations of instances in the validation set. ECG segments between patients exhibit different morphology and correspond to the different arrhythmia labels (Sinus Rhythm vs. Atrial Fibrillation).

D.3.2 PHYSIONET 2020

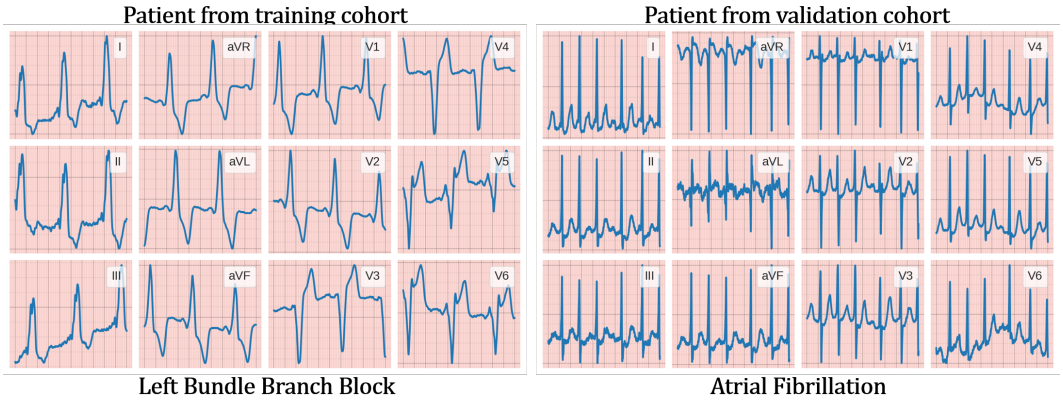


Figure 12: 12-Lead ECG segments corresponding to two dissimilar patients in the training and validation set of the Chapman dataset. Similarity is defined as low Euclidean distance between patient cardiac prototypes (PCPs) and representations of instances in the validation set. ECG segments between patients exhibit different morphology and correspond to the different arrhythmia labels (Sinus Rhythm vs. Atrial Fibrillation).

D.4 VISUALIZATION OF DISSIMILAR PATIENTS ACROSS DIFFERENT DATASETS (CHAPMAN → PTB-XL)

In this section, we leverage the distance matrix visualized in Fig. 10 to identify the two most dissimilar patients across the Chapman and PTB-XL datasets. In Fig. 13, we visualize the 12-Lead ECG recordings of this pair of patients. PCPs are able to reliably identify two patients that are dissimilar. This can be seen the drastically different ECG morphology present in Fig. 13. The patient in the Chapman dataset is suffering from sudden bradycardia, a decrease in the rate at which the heart beats. In contrast, the patient from the PTB-XL dataset is suffering from changes to the ST segment of the ECG recording. Such a finding reaffirms our interpretation of PCPs as reliable descriptors of the cardiac state of a patient.

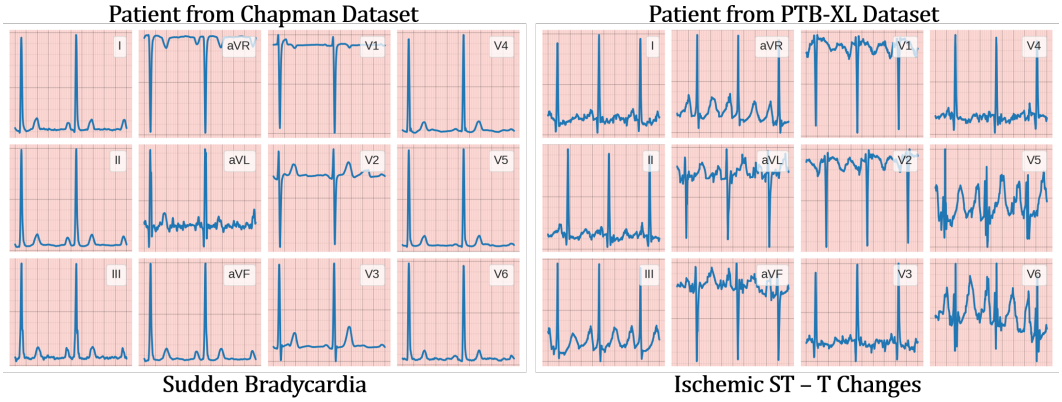


Figure 13: 12-Lead ECG segments corresponding to two dissimilar patients in the training set of the Chapman and PTB-XL datasets, respectively. Similarity is defined as low Euclidean distance between patient cardiac prototypes (PCPs) and representations of instances in the validation set. ECG segments between patients exhibit different morphology and correspond to the different arrhythmia labels (Sudden Bradycardia vs. Ischemic ST-T Changes).

E EFFECT OF NUMBER OF LEADS ON PERFORMANCE

In the main manuscript, we conducted experiments with all 12 leads of an ECG. However, the availability of all 12 leads for training is not always guaranteed. This can be the case, for instance, in low-resource clinical settings where medical infrastructure is lacking or in the context of home-monitoring where wearable sensors are used. To investigate the robustness of our method to such scenarios, we repeat a subset of the experiments in the presence of only 4 leads (II, V2, aVR, aVL), whose results can be found in Table. 6.

Dataset # of Leads	Chapman 12	Chapman 4
AUC	0.891 ± 0.003	0.886 ± 0.002

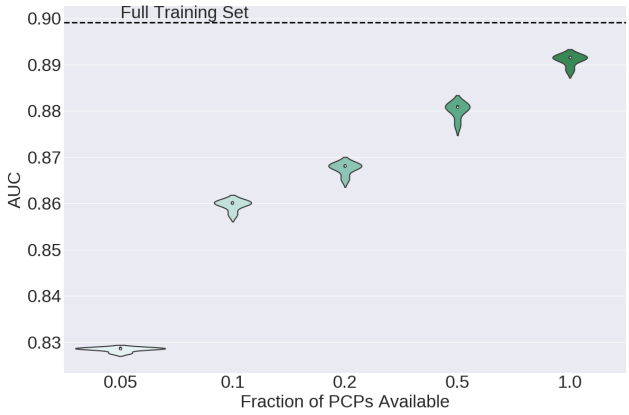
Table 6: AUC on test set of Chapman dataset in the presence of a different number of leads. The inference strategy involves using the nearest PCP ($E = 128$) as input to the hypernetwork. Results are shown for five seeds.

F DATASET DISTILLATION - TRAINING WITH PATIENT CARDIAC PROTOTYPES

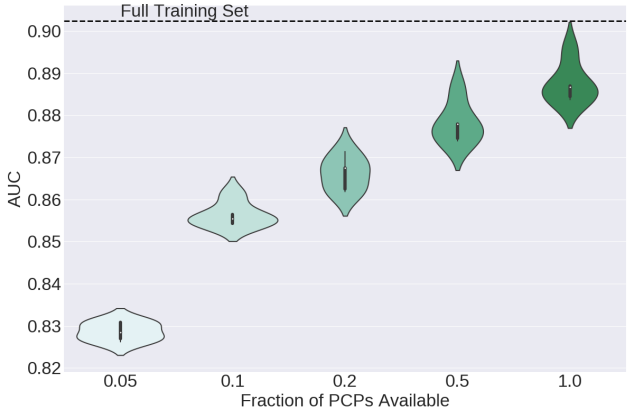
In the main manuscript, we illustrated the potential applicability of PCPs as dataset distillers. Namely, PCPs can act as a compact core-set that does not compromise the generalization performance of a model. In this section, we build on those results and illustrate the utility of PCPs as dataset distillers as we change the dimension of the representation, $E = [64, 128, 256]$.

F.1 CHAPMAN

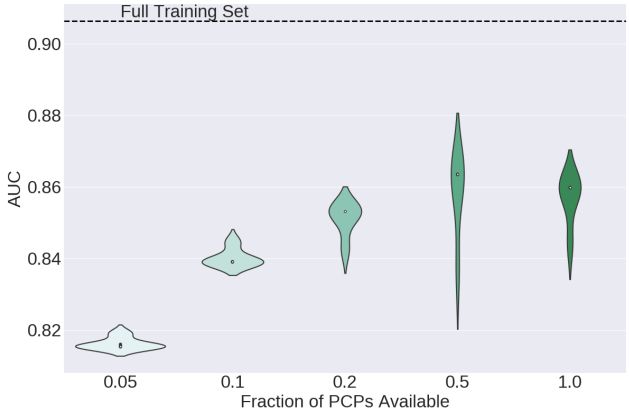
We show that the embedding dimension, E, has a significant effect on the dataset distillation capabilities of PCPs. In Fig. 14b, at an embedding dimension, $E = 128$, the performance drop due to training with 100% of PCPs relative to training with the full training set is minimal, $\Delta\text{AUC} \approx 0.90 - 0.89 = 0.01$. In contrast, at $E = 256$, this performance drop is more substantial $\Delta\text{AUC} \approx 0.905 - 0.86 = 0.045$. Such a finding suggests that more attention should be given to the embedding dimension when designing dataset distillation methods.



(a) $E = 64$



(b) $E = 128$



(c) $E = 256$

Figure 14: Validation AUC after training an SVM on a different fraction of PCPs available. The generalization performance when training on the full training set is also shown. (a)-(c) illustrate the effect of changing the embedding dimension, E, on the generalization performance. Results are shown across 5 seeds.

F.2 PTB-XL

Figure 15: Validation AUC after training an SVM on a different fraction of PCPs ($E - 126$) available. The generalization performance when training on the full training set is also shown. Results are shown across 5 seeds.