PCPs: Patient Cardiac Prototypes

Anonymous authors
Paper under double-blind review

Abstract

Existing deep learning methodologies within the medical domain are typically population-based and difficult to interpret. This limits their clinical utility as population-based findings may not generalize to the individual patient. To overcome these obstacles, we propose to learn patient-specific representations, entitled patient cardiac prototypes (PCPs), that efficiently summarize the cardiac state of a patient. We show that PCPs, learned in an end-to-end manner via contrastive learning, allow for the discovery of similar patients both within and across datasets, and can be exploited for dataset distillation as a compact substitute for the original dataset.

1 Introduction

Physiological signals such as the electrocardiogram (ECG) convey a significant amount of information about the functionality and potential abnormalities of an individual's heart. The automatic detection of these abnormalities, a subset of which are known as arrhythmias, through the use of deep learning has been quite successful in recent years (Galloway et al., 2019; Attia et al., 2019b;a; Ko et al., 2020). However, diagnoses generated by these algorithms remain population-based and difficult to interpret. Such properties arise due to the failure to incorporate patient-specific structure during model training or inference. As a result, medical practitioners are reluctant to integrate network-generated predictions into their clinical workflow.

Modern medical research is arguably anchored around the "gold standard" of evidence provided by randomized control trials (Cartwright, 2007). However, the conclusions derived from such trials are population-based and fail to capture nuances at the individual patient level. The complex structures that define a patient in terms of their demographics, physiological state, and treatment outcomes imply that population-based findings may not trivially extend to the level of an individual patient. Consequently, personalized medicine, the ability to deliver the right treatment to the right patient at the right time, is increasingly viewed as being an important aim of medical diagnosis (Hamburg & Collins, 2010).

Models that are specifically structured to exploit patient-specific information, already present within existing physiological datasets, stand to benefit from personalized medical diagnosis and improved clinical interpretability. Moreover, the ability to compare patients allows clinicians to validate network-generated diagnoses. It also allows them to explore previously unidentified patient relationships, thus paving the way for novel clinical hypotheses. Appreciating this potential, Pai & Bader (2018); Pai et al. (2019) propose patient similarity networks. These graph networks combine multi-modal clinical data to model connections between patients. Similarly, Zhu et al. (2016) and Suo et al. (2017) exploit methods from natural language processing to learn patient representations and similarities based on electronic health record data. In performing patient representation learning, existing approaches are relatively naive, do not concurrently optimize for a predictive task, and do not trivially extend to physiological signals.

Contributions. In this work, we propose a network capable of learning patient representations that can be exploited to perform personalized arrhythmia diagnosis based on a standard 12-lead ECG. More specifically, our contributions are the following:

1. **Patient Cardiac Prototypes (PCPs)** - we learn representations that efficiently summarize the cardiac state of a patient in an end-to-end manner via contrastive learning.
2. **Patient-specific Parameter Generation** - we feed PCPs to a hypernetwork, during inference, to generate patient-specific parameters for the final (linear) layer of a classification network.

3. **Patient Similarity Quantification** - we show that, by measuring the Euclidean distance between PCPs and representations, we can identify similar patients across different datasets.
4. **Dataset Distillation** - we show that PCPs can be used to train a network, in lieu of the original dataset, and maintain strong generalization performance.

2 Related Work

Contrastive learning is a self-supervised method that encourages representations of instances with commonalities to be similar to one another. This is performed for each instance and its perturbed counterpart (Chen et al., 2020) and for different visual modalities (views) of the same instance (Tian et al., 2019). Such approaches are overly-reliant on the choice of perturbations and necessitate a large number of comparisons. Instead, Caron et al. (2020) propose to learn cluster prototypes. Most similar to our work is that of Cheng et al. (2020) and CLOCS (Kiyasseh et al., 2020) which both show the benefit of encouraging patient-specific representations to be similar to one another. In contrast to existing methods, we learn patient-specific representations, PCPs, in an end-to-end manner

Meta-learning is focused on designing learning paradigms that allow for the fast adaptation of networks. Prototypical Networks (Snell et al., 2017) average representations to obtain class-specific prototypes. During inference, the similarity of representations to these prototypes determines the classification. Relational Networks (Sung et al., 2018) build on this idea by learning the similarity of representations to prototypes through a parametric function. Gidaris & Komodakis (2018) and Qiao et al. (2018) exploit hypernetworks (Ha et al., 2016) and propose to generate the parameters of the final linear layer of a network for few-shot learning on visual tasks. In contrast, during inference only, we compute the cosine similarity between representations and PCPs and use the latter as the input to a hypernetwork.

Patient similarity aims at discovering relationships between patient data Sharafoddini et al. (2017). To quantify these relationships, Pai & Bader (2018) and Pai et al. (2019) propose Patient Similarity Networks for cancer survival classification. Exploiting electronic health record data, Zhu et al. (2016) use Word2Vec to learn patient representations, and Suo et al. (2017) propose to exploit patient similarity to guide the re-training of models, an approach which is computationally expensive. Instead, our work naturally learns PCPs as efficient descriptors of the cardiac state of a patient.

3 Methods

3.1 Learning Patient Cardiac Prototypes

We assume the presence of an embedding network, $f_\theta : x \in \mathbb{R}^D \rightarrow h \in \mathbb{R}^E$, that is parameterized by θ and maps a D-dimensional input, x, to an E-dimensional representation, h. We also assume the presence of a distinct set of embeddings, $v \in V$, where $|V| = P$, representing each of the P patients in the training set.

We aim to learn embeddings that are efficient descriptors of the cardiac state of a patient. As a result, we refer to such embeddings as patient cardiac prototypes or PCPs. We propose to learn PCPs in an end-to-end manner via contrastive learning. More specifically, given an instance, x_i, that belongs to a particular patient, k, we encourage its representation, $h_i = f_\theta(x_i)$, to be similar to the same patient's PCP, v_k, and dissimilar to the remaining PCPs. We quantify this similarity, $s(h_i, v_k)$, by using the cosine similarity with a temperature parameter, τ. The intuition is that each PCP, in being attracted to a diverse set of representations that belong to the same patient, should become invariant to insidious intra-patient differences. For a mini-batch of size, B, the contrastive loss is as follows.

$$\mathcal{L}_{contrastive} = -\sum_{i}^{B} \log \left[\frac{e^{s(h_i, v_k)}}{\sum_j^P e^{s(h_i, v_j)}} \right] \tag{1}$$

$$s(h_i, v_j) = \frac{f_\theta(x_i) \cdot v_j}{\|f_\theta(x_i)\| \|v_j\|} \cdot \frac{1}{\tau} \tag{2}$$

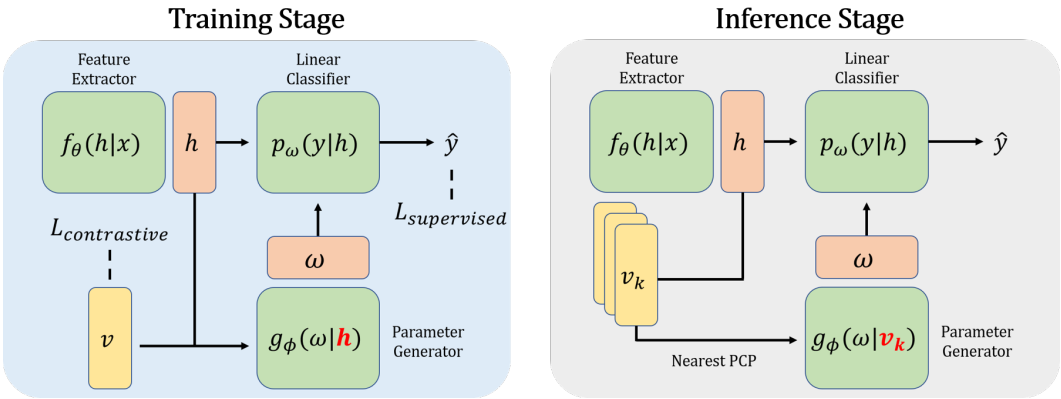


Figure 1: Training and inference stages of the personalized diagnosis pipeline. During training, instance representations, h, have a dual role. They are fed into a hypernetwork, g_ϕ, to generate parameters for a linear classification layer, p_ω, that outputs a prediction, $\hat{y}$. They are also encouraged to be similar to their corresponding patient cardiac prototype (PCP), v. During inference, the nearest PCP, v_k, to the representation is fed into the hypernetwork, thus generating patient-specific parameters for classification.

3.2 Generating Patient-specific Parameters

Neural network parameters are typically updated during training and fixed during inference. This allows the parameters to exploit population-based information in order to learn high-level features useful for solving the task at hand. Such an approach, however, means that all instances are exposed to the same set of parameters regardless of instance-specific information. In this section, we outline how to incorporate instance-specific information to generate patient-specific parameters.

We assume the presence of a hypernetwork, $g_\phi : h \in \mathbb{R}^E \rightarrow \omega \in \mathbb{R}^{E \times C}$, parameterized by ϕ that maps an E-dimensional representation, h, to a matrix of classification parameters, ω, where C is the number of class labels. During training, we feed a representation, h_i, to the hypernetwork and generate instance-specific parameters, ω_i. During inference, however, we use a representation, h_i, to retrieve the most similar PCP (according to some similarity metric, s). We decide to feed this retrieved PCP to the hypernetwork, after experimenting with various strategies (see Sec. 5.2). By performing this retrieval, our approach exploits the similarity between patients in the training and inference set. The hypernetwork then generates parameters that parameterize the linear classifier, $p_\omega : h \in \mathbb{R}^E \rightarrow y \in \mathbb{R}^C$, that maps a representation, h, to a posterior class distribution, y. A schematic of our proposed pipeline can be found in Fig. 1. We train the entire network in an end-to-end manner using a combined supervised and contrastive loss.

$$\mathcal{L}_{supervised} = -\sum_{i}^{B} \log p_{\omega_i}(y_i = c|h_i) \tag{3}$$

$$\omega_i = \begin{cases} g_\phi(h_i) & \text{for training} \\ g_\phi(v_k) & \text{for inference, } v_k = \arg\max_{v_j} s(h_i, v_j) \end{cases} \tag{4}$$

$$\mathcal{L}_{combined} = \mathcal{L}_{contrastive} + \mathcal{L}_{supervised} \tag{5}$$

4 Experimental Design

4.1 Datasets

We conduct experiments using PyTorch (Paszke et al., 2019) on three large-scale ECG datasets that contain a significant number of patients. **PhysioNet 2020 ECG** consists of 12-lead ECG recordings from 6,877 patients alongside labels corresponding to 9 different classes of cardiac arrhythmia. Each recording can be associated with multiple labels. **Chapman ECG** (Zheng et al., 2020) consists

of 12-lead ECG recordings from 10,646 patients alongside labels corresponding to 11 different classes of cardiac arrhythmia. As is suggested by Zheng et al. (2020), we group these labels into 4 major classes. **PTB-XL ECG** (Wagner et al., 2020) consists of 12-lead ECG recordings from 18,885 patients alongside 71 different types of annotations provided by two cardiologists. We follow the training and evaluation protocol presented by Strodthoff et al. (2020) where we leverage the 5 diagnostic class labels. We alter the original setup only considering ECG segments with one label assigned to them and converting the task into a binary classification problem. Further details can be found in Appendix A.1.

Unless otherwise mentioned, datasets were split into training, validation, and test sets according to patient ID using a 60, 20, 20 configuration. In other words, patients appeared in only one of the sets. Further details about the dataset splits can be found in Appendix A.2.

4.2 Hyperparameters

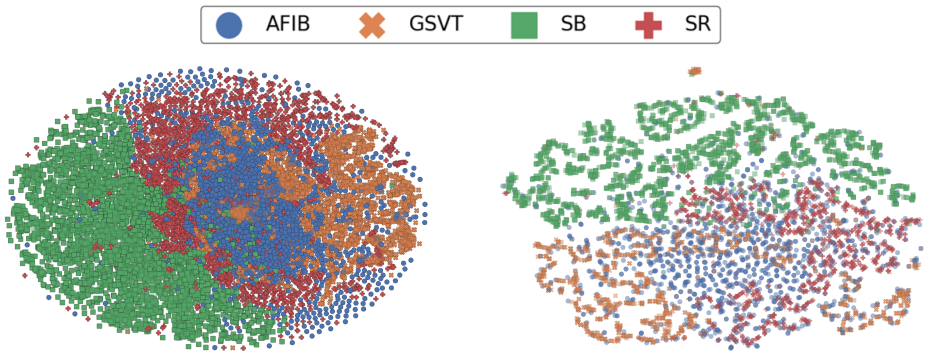
When calculating the contrastive loss, we chose $\tau = 0.1$ as in Kiyasseh et al. (2020). We also use the same neural network architecture for all experiments. Further implementation details can be found in Appendix B.

5 Experimental Results

5.1 Patient Cardiac Prototypes are Discriminative

During training, we optimize an objective function that consists of both a supervised and contrastive loss term (see Eq. 5). Based on the former, we expect representations to exhibit discriminative behaviour for the task at hand. The latter term encourages these representations to be similar to PCPs, and thus we also expect PCPs to be discriminative. In Fig. 2, we illustrate the representations of instances in the training set and the PCPs after being projected to a 2-dimensional subspace using t-SNE and colour-coded according to their class label.

We find that both training representations, h, and PCPs, v, are class-discriminative. This can be seen by the high separability of the projected representations along class boundaries. Based on this finding alone, one could make the argument that PCPs are trivially detecting class label differences between patients. However, we see that the PCPs exhibit tighter clusters than those found with training representations. Such a finding suggests that PCPs are encoding information beyond class labels and that which could be patient-specific (a claim we validate in Sec. 5.3).



(a) Training representations

(b) PCPs

Figure 2: t-SNE projection of (a) representations, $h \in \mathbb{R}^{128}$, of instances in the training set of the Chapman dataset and (b) PCPs, $v \in \mathbb{R}^{128}$, learned on the training set, colour-coded according to the arrhythmia label assigned to each patient. Since PCPs are trained to be similar to representations, they are also class-discriminative. However, the tight clustering exhibited by the PCPs suggests that they may also capture patient-specific information.

5.2 EFFECT OF HYPERNETWORK INPUT STRATEGY ON PERFORMANCE

As described, our pipeline uses the PCP nearest to each representation as input to the hypernetwork. This approach places a substantial dependency on that single chosen PCP. Therefore, we explore three additional input strategies that incorporate PCPs differently. **Nearest 10** searches for, and takes the average of, the 10 PCPs that are nearest to the representation. **Mean** simply takes the average of all PCPs. **Similarity-Weighted Mean** takes a linear combination of all PCPs, weighted according to their cosine similarity to the representation. In Fig. 3, we show the effect of these strategies on the test set AUC as the embedding dimension, E, is changed.

We find that exploiting the similarity of representations during inference benefits the generalization performance of the network. This is shown by the inferiority of the mean strategy relative to the remaining strategies. For example, at $E = 256$, the Mean strategy achieves an $\text{AUC} \approx 0.50$, equivalent to a random guess. However, simply weighting those PCPs according to their similarity to representations, as exemplified by $\text{Similarity Weighted Mean}$ achieves an $\text{AUC} \approx 0.65$. This implies that representations are capturing patient-specific information.

We also find that it is more advantageous to exploit similarity to identify the nearest PCPs than to weight many PCPs. In Fig. 3, the Nearest and Nearest 10 input strategies perform best, with the latter achieving an $\text{AUC} \approx 0.90$, regardless of the embedding dimension. We hypothesize that such behaviour can be attributed to the notion that fewer PCPs are less likely to overwhelm the hypernetwork. This, in turn, allows the hypernetwork to generate reasonable parameters for the linear classification layer. Moreover, the strong performance of these strategies despite their high dependence on so few PCPs reaffirms the utility of the learned representations.

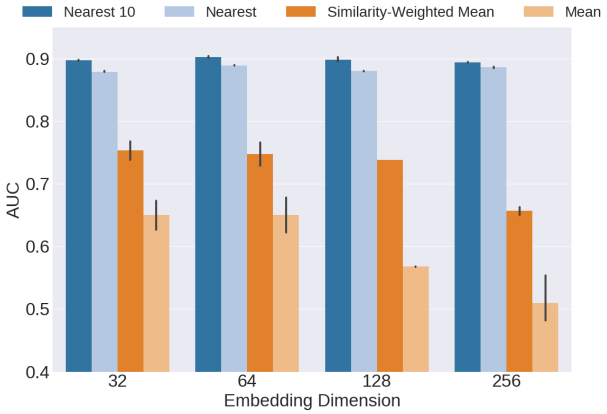


Figure 3: AUC on test set of Chapman dataset as a function of hypernetwork input strategies and embedding dimension, E. Bars are averaged across five seeds and the error bars illustrate one standard deviation. The Nearest 10 input strategy outperforms its counterparts and is unaffected by changes to the embedding dimension.

5.3 PATIENT CARDIAC PROTOTYPES ARE PATIENT-SPECIFIC

So far, we have shown that PCPs are class-discriminative and can assist in achieving strong generalization performance. In this section, we aim to validate our initial claim that PCPs are patient-specific. To do so, we calculate the Euclidean distance between each PCP and two sets of representations. The first includes representations corresponding to the same patient as that of the PCP (**PCP to Same Training Patient**). The second includes representations that correspond to all remaining patients (**PCP to Different Training Patients**). In Fig. 4, we illustrate the distribution of these distances.

We find that PCPs are indeed patient-specific. This can be seen by the smaller average distance between PCPs and representations of the same patient (≈ 4.5) than between PCPs and representations of different patients (≈ 9.5). Such a finding implies that PCPs are, on average, a factor of two more similar to representations from the same patient than they are to those from other patients.

Having established that PCPs are patient-specific, we set out to investigate whether computing their similarity to representations of instances in the validation set (as is done in Fig. 1) was appropriate. In Fig. 4, we overlay the distribution of distances between the PCPs and representations of instances from the validation set (**PCP to Validation Patients**).

We find that comparing PCPs to representations of instances in the validation set is appropriate. This is emphasized by how the average distance between these two entities (≈ 9) is on the same order of the average distance between PCPs and representations of instances in the *training* set (≈ 4). Based on the distributions in Fig. 4, we can also confirm that patients in the validation set are not present in the training set, as was enforced by design. This can be seen by the minimal overlap between the blue and purple distributions. Such a finding suggests that PCPs can be deployed to detect out-of-distribution patients or distribution shift. We also illustrate the generalizability of these findings by arriving at similar conclusions on the PTB-XL dataset (see Appendix C).

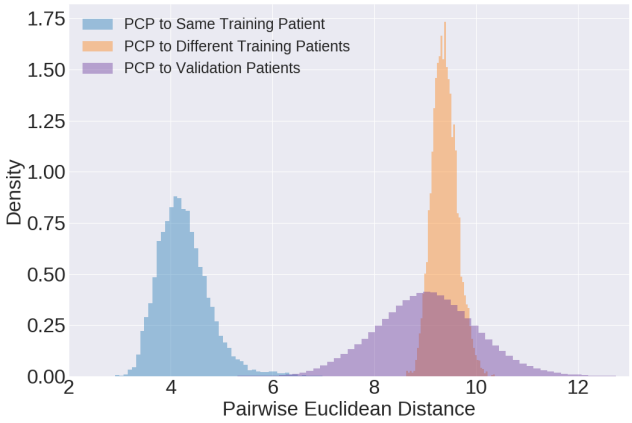


Figure 4: Distribution of pairwise Euclidean distance from the learned PCPs on the Chapman dataset to three sets of representations: those in the training set that belong to the same patient (blue), those in the training set that belong to different patients (orange), and those in the validation set (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.

5.4 Discovery of Similar Patients via Patient Cardiac Prototypes

PCPs, with patient-specific and class-discriminative properties, may allow for the quantification of patient similarity. In this section, we exploit PCPs to identify similar and dissimilar patients. However, validating our findings is non-trivial since similarity can be based on demographics, physiology, or treatment outcomes. With this in mind, we follow these two steps: 1) use PCPs to discover similar and dissimilar patients within the same dataset and 2) validate these findings by comparing the patients' 12-lead ECG recordings. To achieve the first goal, we compute the pairwise Euclidean distance between each PCP and each representation in the validation set. The distribution of these distances can be found in Fig. 5 (top). We average these distances across representations that belong to the same patient, and generate a matrix of distances between pairs of patients (see Fig. 5 centre for a subset of that matrix). By locating the cell with the lowest Euclidean distance, we identify the pair of patients that are most similar to one another and visualize their corresponding 12-lead ECG recordings (Fig. 5 bottom).

We find that PCPs are able to sufficiently distinguish between unseen patients and thus act as reasonable patient-similarity tools. In Fig. 5 (centre), we see that there exists a large range of distance values for any chosen PCP (row). In other words, it is closer to some representations than to others, implying that a chosen PCP is not trivially equidistant to all other representations. However, distinguishing between patients is not sufficient for a patient-similarity tool. We show that PCPs can also correctly capture the relative similarity to these patients. In Fig. 5 (bottom), we show that the two patients identified as being closest to one another, using our method, have ECG recordings with a similar morphology and arrhythmia label, supra-ventricular tachycardia. We hypothesize that this

behaviour arises due to the ability of PCPs to efficiently summarize the cardiac state of a patient. Such a finding reaffirms the potential of PCPs as patient-similarity tools. We also repeat the above procedure in attempt to discover similar and dissimilar patients across *different* datasets. In doing so, we arrive at positive results and similar conclusions to those above (see Appendix D). Cross-dataset patient-similarity quantification can be consequential in exploiting clinical knowledge located in different institutions.

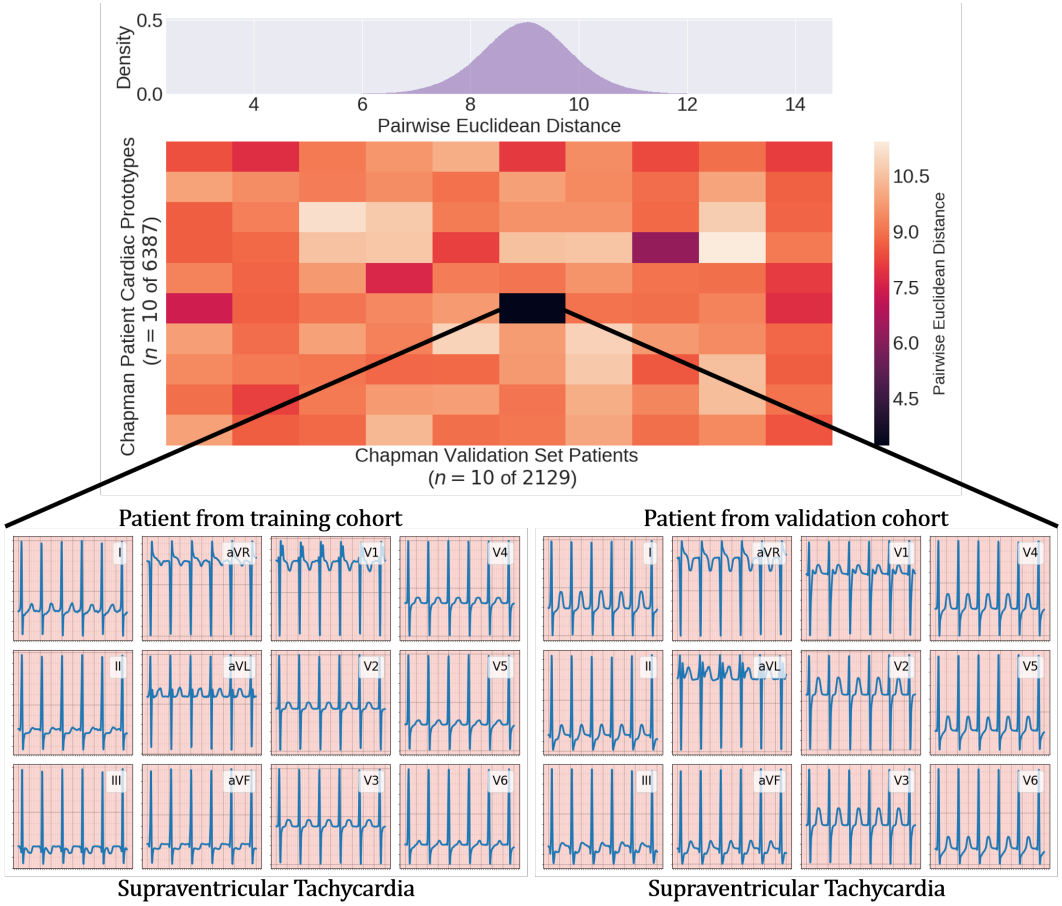


Figure 5: (**Top**) Distribution of all pairwise Euclidean distances between PCPs and representations in the validation set. (**Centre**) Matrix illustrating average pairwise distances between a subset of PCPs and representations of patients in the validation set. (**Bottom**) Visualization of the 12-lead ECG recordings of the two patients identified as being most similar by our method. Both recordings are similar and correspond to the same arrhythmia, supra-ventricular tachycardia, thus lending support to PCPs as a reliable patient-similarity tool.

5.5 Dataset Distillation via Patient Cardiac Prototypes

Our interpretation and the growing evidence we have presented in support of PCPs as efficient descriptors of the cardiac state of a patient led us to investigate the following question: could PCPs be sufficient for training classification tasks, in lieu of the original dataset? This idea is similar to that of dataset distillation which focuses on obtaining a core-set of instances that do not compromise the generalization capabilities of a model (Wang et al., 2018). To investigate the role of PCPs as dataset distillers, we train a Support Vector Machine (SVM) on them and evaluate the model on representations of held-out instances. In Fig. 6, we illustrate the generalization performance of these models after having trained on a different fraction of the available PCPs. For comparison's sake, we also show the AUC when our neural network is trained on all instances in the training set **(Full Training Set)**, which is several folds larger than the number of PCPs.

We find that PCPs do indeed act as effective dataset distillers. In Fig. 6, we show that training on 100% of the PCPs ($n = 6,387$) achieves an $\text{AUC} \approx 0.89$ which is similar to that achieved when training on the full training set ($n = 76,614$). In other words, we achieve similar generalization performance despite a *12-fold* decrease in the number of training instances. We also show that further reducing the number of PCPs, by selecting a random subset for training, does not significantly hurt performance. For example, training with only 5% of available PCPs ($n = 319$) achieves an $\text{AUC} \approx 0.82$. Concisely, this corresponds to a 7% reduction in performance despite a *240-fold* decrease in the number of training instances relative to that found in the standard training procedure. We arrive at similar conclusions when changing the embedding dimension, E (see Appendix F). Such a finding supports the potential of PCPs at dataset distillers. We hypothesize that this behaviour arises due to our patient-centric contrastive learning approach. By encouraging each PCP to be similar to several representations of instances that belong to the same patient, it is able to capture the most pertinent information and shed that which is least useful.

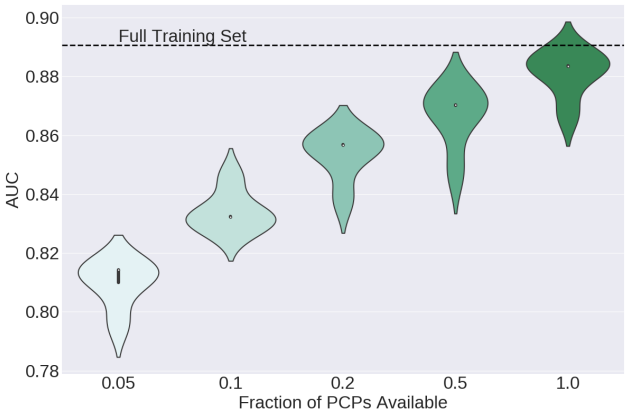


Figure 6: Validation AUC after training an SVM on a different fraction of PCPs ($E = 32$) available. The generalization performance when training on the full training set is also shown. Results are shown across 5 seeds. This illustrates the dataset distillation capabilities of PCPs. By training on 100% of PCPs, we achieve similar performance to when training on the full training set.

5.6 Discussion and Future Work

In this paper, we proposed to learn efficient descriptors of the cardiac state of a patient, which we refer to as patient cardiac prototypes, using a combination of supervised and contrastive learning. We showed that patient cardiac prototypes are both patient-specific and discriminative for the task at hand. We successfully deployed PCPs for the discovery of similar patients within the same dataset and across different datasets. This opens the door to leveraging clinical information that is present in disparate healthcare institutions. Lastly, we illustrated the potential of PCPs as dataset distillers, whereby they can be used to train models in lieu of larger datasets without compromising generalization performance. We now elucidate several avenues worth exploring.

Obtaining Multi-modal Summary of Cardiac State of Patient. Although our approach was validated on multiple, large, time-series datasets, it was limited to a single modality, the ECG. Incorporating additional modalities to our approach such as coronary angiograms, cardiac MRI, and cardiac CT, may provide a more reliable summary of the cardiac state of the patient. This could ultimately lead to more reliable patient similarity quantification.

Guiding Design of Graph Neural Networks. Arriving at ground-truth values for the similarity of a pair of patients is non-trivial. Recently, graph neural networks have been relatively successful at discovering and quantifying the similarity of instances, but most necessitate a pre-defined graph structure, which may be difficult to design in the context of physiological signals. We believe that designing this graph structure can be facilitated by our patient-similarity scores, for instance, by using them as an initialization of the edge weights between nodes.

REFERENCES

Erick A Perez Alday, Annie Gu, Amit Shah, Chad Robichaux, An-Kwok Ian Wong, Chengyu Liu, Feifei Liu, Ali Bahrami Rad, Andoni Elola, Salman Seyedi, et al. Classification of 12-lead ecgs: the physionet/computing in cardiology challenge 2020. *medRxiv*, 2020.

Zachi I Attia, Suraj Kapa, Francisco Lopez-Jimenez, Paul M McKie, Dorothy J Ladewig, Gaurav Satam, Patricia A Pellikka, Maurice Enriquez-Sarano, Peter A Noseworthy, Thomas M Munger, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nature Medicine*, 25(1):70–74, 2019a.

Zachi I Attia, Peter A Noseworthy, Francisco Lopez-Jimenez, Samuel J Asirvatham, Abhishek J Deshmukh, Bernard J Gersh, Rickey E Carter, Xiaoxi Yao, Alejandro A Rabinstein, Brad J Erickson, et al. An artificial intelligence-enabled ecg algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet*, 394(10201):861–867, 2019b.

Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *arXiv preprint arXiv:2006.09882*, 2020.

Nancy Cartwright. Are RCTs the gold standard? *BioSocieties*, 2(1):11–20, 2007.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020.

Joseph Y Cheng, Hanlin Goh, Kaan Dogrusoz, Oncel Tuzel, and Erdrin Azemi. Subject-aware contrastive learning for biosignals. *arXiv preprint arXiv:2007.04871*, 2020.

Conner D Galloway, Alexander V Valys, Jacqueline B Shreibati, Daniel L Treiman, Frank L Petterson, Vivek P Gundotra, David E Albert, Zachi I Attia, Rickey E Carter, Samuel J Asirvatham, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiology*, 4(5):428–436, 2019.

Spyros Gidaris and Nikos Komodakis. Dynamic few-shot visual learning without forgetting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4367–4375, 2018.

David Ha, Andrew Dai, and Quoc V Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016.

Margaret A Hamburg and Francis S Collins. The path to personalized medicine. *New England Journal of Medicine*, 363(4):301–304, 2010.

Dani Kiyasseh, Tingting Zhu, and David A Clifton. CLOCS: Contrastive learning of cardiac signals. *arXiv preprint arXiv:2005.13249*, 2020.

Wei-Yin Ko, Konstantinos C Siontis, Zachi I Attia, Rickey E Carter, Suraj Kapa, Steve R Ommen, Steven J Demuth, Michael J Ackerman, Bernard J Gersh, Adelaide M Arruda-Olson, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *Journal of the American College of Cardiology*, 75(7):722–733, 2020.

Shraddha Pai and Gary D Bader. Patient similarity networks for precision medicine. *Journal of Molecular Biology*, 430(18):2924–2938, 2018.

Shraddha Pai, Shirley Hui, Ruth Isserlin, Muhammad A Shah, Hussam Kaka, and Gary D Bader. netdx: Interpretable patient classification using integrated patient similarity networks. *Molecular Systems Biology*, 15(3), 2019.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pp. 8024–8035, 2019.

Siyuan Qiao, Chenxi Liu, Wei Shen, and Alan L Yuille. Few-shot image recognition by predicting parameters from activations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7229–7238, 2018.

Anis Sharafoddini, Joel A Dubin, and Joon Lee. Patient similarity in prediction models based on health data: a scoping review. *JMIR Medical Informatics*, 5(1):e7, 2017.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pp. 4077–4087, 2017.

Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, and Wojciech Samek. Deep learning for ecg analysis: Benchmarks and insights from ptb-xl. *arXiv preprint arXiv:2004.13701*, 2020.

Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1199–1208, 2018.

Qiuling Suo, Fenglong Ma, Ye Yuan, Mengdi Huai, Weida Zhong, Aidong Zhang, and Jing Gao. Personalized disease prediction using a cnn-based similarity learning method. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 811–816. IEEE, 2017.

Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. *arXiv preprint arXiv:1906.05849*, 2019.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Wojciech Samek, and Tobias Schaeffter. PTB-XL, a large publicly available electrocardiography dataset, 2020. URL `https://physionet.org/content/ptb-xl/1.0.1/`.

Tongzhou Wang, Jun-Yan Zhu, Antonio Torralba, and Alexei A Efros. Dataset distillation. *arXiv preprint arXiv:1811.10959*, 2018.

Jianwei Zheng, Jianming Zhang, Sidy Danioko, Hai Yao, Hangyuan Guo, and Cyril Rakovski. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. *Scientific Data*, 7(1):1–8, 2020.

Zihao Zhu, Changchang Yin, Buyue Qian, Yu Cheng, Jishang Wei, and Fei Wang. Measuring patient similarities via a deep architecture with medical concept embedding. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 749–758. IEEE, 2016.