Contrastive Learning Is Not Optimal for Quasiperiodic Time Series

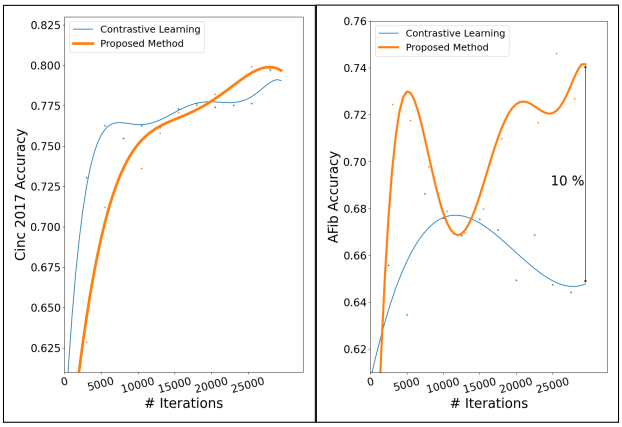
Adrian Atienza , Jakob Bardram , Sadasivan Puthusserypady
Department of Health Technology, Technical University of Denmark
{adar, jakba, sapu}@dtu.dk

Abstract

Despite recent advancements in Self-Supervised Learning (SSL) for time series analysis, a noticeable gap persists between the anticipated achievements and actual performance. While these methods have demonstrated formidable generalization capabilities with minimal labels in various domains, their effectiveness in distinguishing between different classes based on a limited number of annotated records is notably lacking. Our hypothesis attributes this bottleneck to the prevalent use of Contrastive Learning, a shared training objective in previous state-of-the-art (SOTA) methods. By mandating distinctiveness between representations for negative pairs drawn from separate records, this approach compels the model to encode unique record-based patterns but simultaneously neglects changes occurring across the entire record. To overcome this challenge, we introduce Distilled Embedding for Almost-Periodic Time Series (DEAPS) in this paper, offering a non-contrastive method tailored for quasiperiodic time series, such as electrocardiogram (ECG) data. By avoiding the use of negative pairs, we not only mitigate the model's blindness to temporal changes but also enable the integration of a "Gradual Loss ($\mathcal{L}_{gra}$)" function. This function guides the model to effectively capture dynamic patterns evolving throughout the record. The outcomes are promising, as DEAPS demonstrates a notable improvement of +10% over existing SOTA methods when just a few annotated records are presented to fit a Machine Learning (ML) model based on the learned representation.

1 Introduction

With the rise of Self-Supervised Learning (SSL), there have been many efforts to adapt different works from other fields to time series processing, particularly physiological time series ([Diamant *et al.*, 2022], [Wickstrøm *et al.*, 2022]). In this field, being able to optimize a model without the need for labels is highly valuable given that: (i) the model learns generic representations that are useful across many tasks, (ii) the labeling is costly, or the specific task is not known a priori, and (iii) the model learns representations that are more robust across different perturbations occurred during the recording of the data. Existing methods drive the model to learn representations that are refined throughout the optimization process. The learned representations are capable of performing a downstream task better and better throughout the training process, as shown in Figure 1a, when the Machine Learning (ML) model considers a high number (8500) of recordings when is fitted.



(a) Physionet Challenge 2017 (b) From MIT-ARR Training to MIT-AFIB Testing

Figure 1: Evolution of model performance across training procedures for best performing Contrastive Learning Method (blue) and proposed method (orange). While the downstream task displayed in Figure 1a considers multiple annotated recordings, the one shown in Figure 1b only considers a few of them. For an easier track of the evolution across the iterations, a polynomial has been fitted.

In certain other fields of Artificial Intelligence, existing SSL methods have shown promising results for a variety of applications by merely fitting a simple ML model on top of the representation using a limited amount of data [Caron *et al.*, 2021], [Grill *et al.*, 2020]. Achieving this level of performance is anticipated in a comparable time series analysis experiment, focusing on just a few records that encompass both normal and abnormal states. However, Figure 1b shows

that the results of this experiment are far from what was expected. It can also be seen how the performance of the model decreases throughout the optimization process.

While each method has its own uniqueness, they all share the use of Contrastive Learning [Chen *et al.*, 2020] as part of their optimization objective. This contrastive goal leads to the representations to be distant from the negative pairs which are drawn from distinct records. Quasiperiodic signals exhibit a combination of regular patterns with subtle variations. These quasiperiodic signals, such as electrocardiogram (ECG), exhibit much more visible differences between recordings from different subjects than between different classes within the same recording (See Figure 2).

Just by seeking similarities between positive pairs belonging to the same record and dissimilarities between different ones, Contrastive learning drives the model to capture unique record-based representations. However, in parallel, this contrastive objective leads the model to neglect the less visible, dynamic changes that occur across time. This clarifies why these methods struggle to guide the model in generalizing across various classes present in a recording when provided with a limited set of recordings.

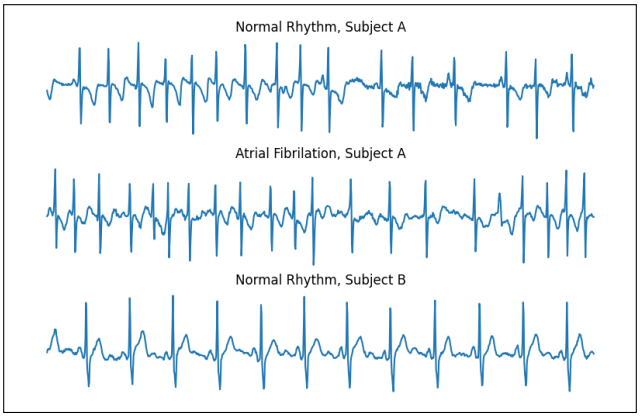


Figure 2: A pair time series representing a normal and an abnormal state, belonging to the same subject are shown. In addition, another time strip that belongs to a distinct subject is displayed. The ECG morphology observed across various states within subject A exhibits consistent patterns, whereas this is different between subject A and B. These patterns are hypothesized to enable the model to categorize between two subjects while not reflecting subject state information.

In this paper, we present Distilled Embedding for Almost-Periodic Time Series (DEAPS), a new SSL method designed for physiological time series. The underlying concept of this approach is based on the categorization of signal patterns into two types: (i) Static patterns that account for individual characteristics such as gender and age and, (ii) dynamic patterns that can reveal transitional states or events experienced by the subjects during the recording. DEAPS extends existing non-contrastive methods by not requiring the use of negative pairs during model optimization. This alleviates the model's blindness to dynamic patterns, while effectively continuing to encode the static ones.

In addition, avoiding the use of negative pairs allows DEAPS to introduce a novel loss function, "Gradual Loss ($\mathcal{L}_{gra}$)", that encourages the model to put more emphasis on representing the dynamic patterns. This loss function operates by following these steps: (i) Three time strips belonging to the same recording are given as inputs, (ii) Assuming that the dynamic patterns exhibit a smooth and linear behavior through the recording, $\mathcal{L}_{gra}$ ensures that the representation of the middle input can be interpolated from the other two representations, and (iii) A selective optimization is incorporated for only considering the most important features (i.e., the features for which their value changes the most over the recording interval).

We hypothesize that (i) the use of Contrastive Learning techniques, by introducing an objective that prioritizes that the representations of different recordings are far apart, causes the model to neglect the dynamic patterns, (ii) Dynamic patterns are much less evident than static ones, so avoiding the use of negative pairs is not enough. An additional objective function is needed for the model to capture them, and (iii) By effectively capturing these dynamic patterns, the model will acquire a better generalization capability by just considering a few labeled records.

As a result, DEAPS is able to not only encode static patterns but also to distinguish different classes in the same recording by encoding the dynamic patterns too. Therefore, the model generalizes these classes to perform up to 10% better when evaluated on another dataset, compared to SOTA contrastive methods (Figure 1b). It also performs better when the ML model is fitted using multiple recordings (Figure 1a). In addition to these two experiments, DEAPS is evaluated against SOTA methods in two other experiments. Overall, the evaluation involves up to 3 different tasks and 3 different datasets different from the one used during the optimization process.

In summary, the contributions of this paper are:

1. It is demonstrated, both logically and empirically, that Contrastive Learning objectives lead the model to neglect subtle changes within quasiperiodic time series data.

2. We introduce DEAPS, a novel SSL method that stands apart from previous SOTA Contrastive Learning methods. Dispensing with the negative pairs not only alleviates this model blindness to shifts within the record but also allows us to incorporate the novel $\mathcal{L}_{gra}$ function. This function, which emphasizes understanding changes within these quasiperiodic data, leads the model to encode dynamic patterns.

3. We show that by understanding and encoding these dynamic patterns, the model is able to generalize different classes given a small sample of recordings where these two appear. This is exemplified by the up to 10% improvement when compared with existing methods.

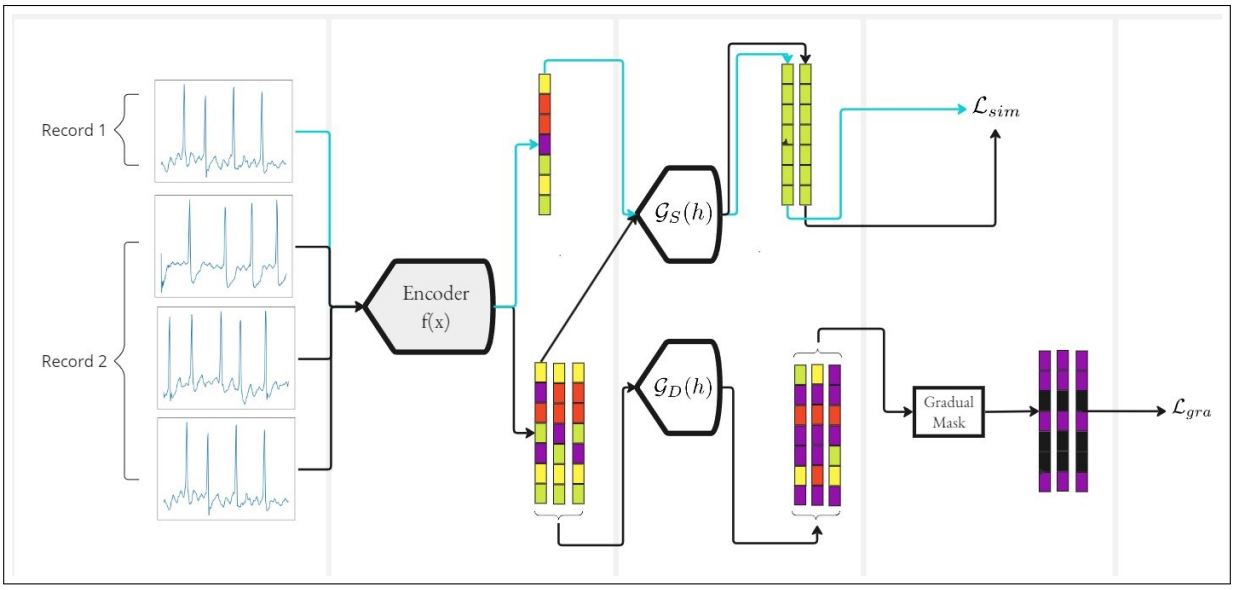


Figure 3: DEAPS illustrated. Different colors represent different temporal patterns: Green (static patterns), and purple (dynamics patterns). The different colors of the arrows indicate inputs from different recordings. While $\mathcal{L}_{sim}$ is computed between the time series representations belonging to different records, $\mathcal{L}_{gra}$ is computed using the time series belonging to the same record. The four inputs belong to the same subject. The encoder is the only component that is saved at the end of the training procedure. The other components are dismissed.

2 Related Work

Unsupervised Learning in Time Series

Interpreting physiological signals demands a level of expertise, making the labeling of such data costly. It represents a bottleneck within the deep learning context, particularly given the substantial volume of data required for training deep learning models. Consequently, various studies have been directed towards optimizing the model using a limited amount of labels [Fan *et al.*, 2021] or directly without relying on these labels ([Tonekaboni *et al.*, 2021], [Franceschi *et al.*, 2020]).

Contrastive Learning in Time Series

With the rise of SSL, different methods have recently been designed in the field of time series analysis, exploiting the characteristics of the time series data. Contrastive Learning of Cardiac Signals Across Space (CLOCS) [Kiyasseh *et al.*, 2021] utilizes two consecutive ECG time strips as positive pairs. The mixing-up method [Wickstrøm *et al.*, 2022] utilizes the temporal characteristics of the data for a more tailored data augmentation by creating a time series, which is a product of two time series from the same subject. An alternative approach is proposed in the Time-Frequency Consistency (TF-C) [Zhang *et al.*, 2022] method, which produces two variations of the time-domain and frequency-domain pairs associated with the input signal. SSL method is developed to identify similarities inherent in these paired representations. The last example is Patient Contrastive Learning (PCLR) [Diamant *et al.*, 2022] which considers two inputs from the same subject but different recordings as positive pairs.

Although each of these methods has its particularity in defining the positive pairs, they all share the use of the Contrastive Learning objective. PCLR obtains the best results in the benchmark used in this work. Accordingly, DEAPS considers the same selection of positive pairs criteria, i.e., between pairs from different recordings.

Non-Contrastive Learning

In other fields of artificial intelligence, such as Computer Vision, non-Contrastive Methods have emerged as an alternative to Contrastive Methods. The major challenge of these methods is to avoid the collapse mode in the representations. This collapse occurs when the model computes the same representations independently of the given input. Given that non-contrastive methods do not force representations of negative pairs to be distinct, other strategies have to be considered to avoid this issue.

Methods such as Bootstrap Your Own Latent (BYOL) [Grill *et al.*, 2020] or Self-Distillation with no Labels (DINO) [Caron *et al.*, 2021] rely on a teacher-student architecture. The cost functions are computed between the corresponding teacher-student. While the student network is optimized with Stochastic Gradient Descent (SGD), the weights of the teacher network are updated via exponential moving average (EMA) of the student weights. Other methods such as Variance-Invariance-Covariance Regularization (VIC-REG) [Bardes *et al.*, 2022] maximize the batch variance of each feature within the representations. The method presented in this work utilizes BYOL framework as the baseline.

3 Distilled Embedding for Almost-Periodic Time Series (DEAPS)

DEAPS is designed to capture both the static and the dynamic patterns of the time series. It is illustrated in Figure 3. The model takes up to four different time strips belonging to two different records from the same subject. The encoder computes the representations of these distinct inputs. The

static projector (labeled $\mathcal{G}_S$) takes a pair of representations from the two different recordings which are displayed in a static space. The similarity loss function ($\mathcal{L}_{sim}$) is computed between these pairs of projections, ensuring that the unique patterns of each subject are represented consistently even between different records. The dynamic projector (labeled $\mathcal{G}_D$) processes the three representations from the same recording. DEAPS only considers the most important features for each triplet of inputs, i.e., the ones that are encoding dynamic patterns that are evolving between this time interval. The rest of the dynamic projection is masked. The unmasked projections are used for computing the gradual loss function ($\mathcal{L}_{gra}$). The overall loss function proposed by DEAPS can be expressed as the following:

$$\mathcal{L}_{DEAPS} = \mathcal{L}_{sim} + \mathcal{L}_{gra} + \alpha \cdot \mathcal{L}_c. \tag{1}$$

This section provides detailed explanations for each of its different components.

3.1 Non-Contrastive Learning. Revisiting BYOL

BYOL [Grill *et al.*, 2020] was the first SSL non-contrastive learning method that dispensed with negative pairs. DEAPS utilizes it as baseline.

In other to avoid the mode collapse when the use of negative pairs is removed, BYOL features both a teacher network and a student network, each with an encoder and a projector initialized with identical parameters. The student network is additionally equipped with a predictor, which acts on the views computed by the projector. The optimization involves SGD for the student network, projector, and predictor. Both the teacher network and the teacher projector serve as an EMA of the student network. The loss function only measures the similarities between the student/teacher output pairs and it is described as the following:

$$\mathcal{L}_{sim}(\mathbf{z}_2^s, \mathbf{q}(\zeta_1)^s) = 1 - \frac{\mathbf{z}_2^s \cdot \mathbf{q}(\zeta_1)^s}{\max\left(\|\mathbf{z}_2^s\|_2 \cdot \|\mathbf{q}(\zeta_1)^s\|_2, \epsilon\right)}, \tag{2}$$

where $\mathbf{z}_2^s$ and $\mathbf{q}(\zeta_1)^s$ are the teacher similarity projection and student similarity prediction, respectively, for inputs belonging to the first and second records.

Both the projector and the predictor are discarded after the optimization process. Therefore, the representations used for the downstream tasks are obtained directly from the student network. The student-teacher framework and the characteristic predictor are excluded from Figure 3 for clarity. In practice, the different loss functions are computed between projection/prediction pairs obtained by the teacher and student networks, denoted as $\mathbf{z}$ and $\mathbf{q}(\zeta)$, respectively, where ζ represents the student projection.

3.2 Gradual Loss

Avoiding the use of a Contrastive objective alleviates the model's blindness to changes in the record. However, it is not sufficient to lead the model to capture dynamic patterns. Non-contrastive methods do not compare each representation with the rest of the representations in the batch. It allows us to incorporate a metric that compares just the representations within the same record.

Capturing the Dynamic Patterns

We introduce the "Gradual Loss ($\mathcal{L}_{gra}$)" as a part of the training objective. This loss function just considers X_{t-i}, X_t, and X_{t+j}, three that belong to the same record and have been drawn over a fixed window size. DEAPS enforces the representation of the dynamic patterns to exhibit a linear behavior by mandating an accurate interpolation of the representation of an inter- mediate input from the representations of the other two. In other words, if a subject's state evolves from X_{t-i} to X_{t+j}, the representation of X_t, i.e., z_t, should approximate an intermediate point between z_{t-i} to z_{t+j}. This objective ensures that the temporal evolution is properly captured within the representations. $\mathcal{L}_{gra}$ is described as:

$$\mathcal{L}_{gra}(\mathbf{q}_s(\mathbf{z}_t^s), \mathbf{z}_{t-i}^t, \mathbf{z}_{t+j}^t) = 1 - \frac{\mathbf{q}_s(\mathbf{z}_t^s) \cdot \mathcal{PAR}(\mathbf{z}_{t-i}^t, \mathbf{z}_{t+j}^t)}{\max\left(\|\mathbf{q}_s(\mathbf{z}_t^s)\|_2 \cdot \|\mathcal{PAR}(\mathbf{z}_{t-i}^t, \mathbf{z}_{t+j}^t)\|_2, \epsilon\right)}, \tag{3}$$

where Pondered Average Representation (PAR) is the approximation of $\mathbf{z}_t$, drawn from $\mathbf{z}_{t-i}$ and $\mathbf{z}_{t+j}$. We do not force $\mathbf{z}_t$ to be equally distant from both $\mathbf{z}_{t-i}$ and $\mathbf{z}_{t+j}$, therefore, we calculate PAR as,

$$\mathcal{PAR}(\mathbf{z}_{t-i}^t, \mathbf{z}_{t+j}^t) = \frac{\mathbf{z}_{t-i}^t \cdot j + \mathbf{z}_{t+j}^t \cdot i}{i+j}. \tag{4}$$

Note that the incorporation of $\mathcal{L}_{gra}$ into the method is only possible since DEAPS is a non-contrastive method. $\mathcal{L}_{gra}$ only makes sense when taking into account only the $\mathbf{z}_t$ and the interpolation of this representation given the $\mathbf{z}_{t-i}$ and $\mathbf{z}_{t+j}$ pairs, and not the rest of the representations in the batch considered as negative-pairs.

Intuitions of the Use of Two Projectors

$\mathcal{L}_{sim}$ and $\mathcal{L}_{gra}$ have conflicting goals. While the first one prioritizes that representations between positive pairs are invariant, the purpose of $\mathcal{L}_{gra}$ is encouraging changes within the recording to be captured in the representations. To facilitate the disentanglement between static and dynamic features, two projectors (represented as $\mathcal{G}_s$ and $\mathcal{G}_d$ in Figure 3) are incorporated into both the student and teacher network instead of the single projector of the original BYOL framework. In this way, only static features will be projected by $\mathcal{G}_s$ and thus taken into account by $\mathcal{L}_{sim}$. Analogously, only dynamic features will be considered when computing $\mathcal{L}_{gra}$.

Intuitions of the Windows Size

An essential consideration in implementing the method is determining the appropriate spacing window size between $\mathbf{X}_{t-i}$ and $\mathbf{X}_{t+j}$, i.e., how much these inputs may be separated in time. This spacing window size must be large enough to accommodate signal changes, yet narrow enough just allow different patterns that encode distinct changes to move in one single direction each. If successive changes occur within this window, it can lead to conflicting directions in which these changes are reflected in the representations, thereby $\mathcal{PAR}(\mathbf{z}_{t-i}^t, \mathbf{z}_{t+j}^t)$ may not be aligned with $\mathbf{q}_s(\mathbf{z}_t^s)$. The optimal value of this window size may vary depending on the quasiperiodic data modality. The effect of this size is studied in Section 5.

3.3 Selective Optimization

It is assumed that a set of features contained in the representation encodes the dynamic patterns of the time series. However, it is not reasonable to think that all these patterns will evolve in every fixed time window. To address this issue, DEAPS introduces a Selective Optimization. This ensures that only the features that encode changing patterns within a particular time window are considered when computing the loss function.

Masking the Output The process of determining which features of each dynamic projection to include in each loss function follows these steps: (i) It calculates the absolute difference between the values dynamic projections of each input pair belonging to the $\mathbf{X}_{t-i}$ and $\mathbf{X}_{t+j}$ batches, (ii) the *N* values that exhibit a larger difference are considered when computing $\mathcal{L}_{gra}$, and (iii) any remaining features in each projection are masked and their gradients are not considered. This process is represented in Figure 3. The effect of this selective optimization is discussed in section 5.

Covariance Term To enhance the efficacy of our selective optimization approach, we strive to ensure that each feature within the representations encapsulates a distinctive and individual pattern. To achieve this objective, we incorporate the Covariance Loss function ($\mathcal{L}_c$) as a regularization factor, effectively penalizing redundancy within the representations. This particular cost function, which has also found application in prior studies like Barlow Twins [Zbontar *et al.*, 2021] and VIC-REG [Bardes *et al.*, 2022], operates as follows:

$$\mathcal{L}_c(\zeta) = \frac{1}{d}\sum_{i \neq j}[C(\zeta)]_{i,j}^2, \tag{5}$$

where $C(\zeta)$ is the covariance matrix computed on the student projection, ζ. Its effect is studied in Section 5.

3.4 Implementation Details

In order to ensure the reproducibility of the method, this subsection details the configuration of the hyperparameters.

Model Architecture We use an adaptation of the Vision Transformer (ViT) [Dosovitskiy *et al.*, 2021] model for processing physiological signals. The input data is a time series of 1000 samples, which correspond to 10 seconds-length signal sampled at 100Hz. This input is split into segments of a length of 20 samples. The model counts with 6 regular transformer blocks with 4 heads each. The model dimension is set to 128, for a total of 1,192,616 trainable parameters.

DEAPS Implementation The projectors and predictors in our approach are implemented as a two-layer Multilayer Perceptron (MLP) with a dimensionality of 512 and 256, respectively. Batch normalization and rectified linear unit (ReLU) operations are incorporated between the two layers of each structure. The EMA updating factor (τ) is set to 0.995. The window size is set to 2 minutes. We weigh the covariance loss with a factor of 0.1. We optimize the most important 32 features during the selective optimization. The effect of both the window size and the number of features to be considered during the selective optimization is discussed in Section 5.

Optimization The model is trained with 10 second-length signals belonging to the Sleep Heart Health Study (SHHS) dataset [Zhang *et al.*, 2018], [Quan *et al.*, 1998]. The training procedure consists of 30,000 iterations. We use a batch size of 256, and Adam [Kingma and Ba, 2017] with a learning rate of $3e-4$ and a weight decay of $1.5e-6$ as the optimizer. The training procedure and the evaluations are performed on a local computer, with a Nvidia GeForce RTX 3070 GPU.

4 Evaluation

The proposed model has undergone extensive simulation studies to evaluate its performance. Initially DEAPS was assessed by comparing it to four SOTA methods across three distinct downstream tasks, each using four different databases. These four databases are: MIT-BIH Atrial Fibrillation Database (MIT-AFIB)[Moody and Mark, 1983], MIT-BIH Arrhythmia Database (MIT-ARR), [Moody and Mark, 1983], Physionet Challenge 2017 (Cinc2017) [Clifford *et al.*, 2017] and SHHS [Zhang *et al.*, 2018]. All used databases are publicly available in Physionet [Goldberger *et al.*, 2000] and National Sleep Research Resource (NSRR).

Furthermore, we conducted a Principal Component Analysis (PCA) analysis to substantiate the core concept of our approach, which revolves around disentangling static and dynamic patterns within time-series data with the aim of demonstrating DEAPS' capability to encode both of them.

4.1 Comparison against SOTA Methods

In order to evaluate the performance of DEAPS, three experiments of different nature have been carried out: (i) Atrial Fibrillation (AFib) identification to assess the effectiveness of generalising different classes within the same recording (ii) Gender Identification, to assess the effectiveness of static patterns and (iii) Physionet Challenge 2017 to assess the scalability of the representations when multiple recordings are given. DEAPS' performance has been compared against the four most relevant SOTA methods, namely; (i) PCLR [Diamant *et al.*, 2022], (ii) CLOCS [Kiyasseh *et al.*, 2021], (iii) Mixing-Up [Wickstrøm *et al.*, 2022], and (iv) TF-C [Zhang *et al.*, 2022]. Moreover, we have included the standard BYOL [Grill *et al.*, 2020] framework as a baseline.

To ensure fairness in evaluation, we have optimized the same model used in this work, under the same configuration (optimizer, data, batch size, and number of iterations), except for the TF-C method, where their proposed model has been used, since it requires the use of two encoders instead of one. The TF-C model contains approximately 32 million parameters, which is 30x more parameters than the proposed DEAPS model. Therefore, it has been optimized over 75K iterations, instead of the 30K iterations proposed in this work.

Atrial Fibrillation (AFib) Identification

In order to assess the ability of the method to generalise different classes within the same record, we have conducted two experiments where there is no subject overlap between the training and validation sets. This overlapping would significantly simplify AFib identification. In the first experiment,

we employed 10-second ECG time series strips originating from the eight subjects affected by AFib, belonging to the MIT-ARR database for fitting a Support Vector Classificatier (SVC) [Platt, 2000] on top of the representation. We subsequently evaluated this model on the complete MIT-AFIB database. In the second, we have conducted a Leave-One-Out (LOO) validation across the 23 MIT-AFIB subjects. The outcomes of this experiment are tabulated in Table 1. DEAPS demonstrates a significantly superior performance.

	MIT-ARR → MIT-AFIB			MIT-AFIB LOO
SSL METHOD	ACCURACY (%)	SENSITIVITY (%)	SPECIFICITY (%)	ACCURACY (%)
Mixing-Up	65.6	60.6	67.4	73.4±17.1
TF-C	71.8	64.8	76.5	77.2±20.5
PCLR	65.5	59.8	68.1	75.1±18.2
BYOL	67.9	62.6	69.9	75.6±21.1
CLOCS	62.8	55.0	66.2	75.4±19.1
DEAPS	**75.5**	**74.6**	**75.9**	**79.4±18.7**

Table 1: AFib Identification given a few, non-overlapping annotated records. In MIT-ARR → MIT-AFIB experiment. a SVC is fitted and evaluated on these datasets, respectively. In MIT-AFIB LOO experiment, a Leave-One-Out cross-validation is carried out.

Gender Classification

We randomly selected 1549 ECG time series strips of length 10 seconds, each associated with a distinct subject, from the SHHS database. We conducted a five-fold cross-validation to evaluate the performance of the downstream tasks. The ML model used is the SVC and the outcomes of these experiments are presented in Table 2, where it can be seen how DEAPS obtains comparable results with the best performing method.

	GENDER IDENTIFICATION		
SSL METHOD	TOTAL ACCURACY (%)	MALE ACCURACY (%)	FEMALE ACCURACY (%)
Mixing-Up	70.4±1.5	68.0±3.3	71.9±2.3
TF-C	65.8±2.9	62.8±5.5	67.4±2.1
PCLR	76.1±1.7	74.7±2.6	77.2±3.1
BYOL	**76.3±3.3**	**75.3±4.0**	**77.6±3.9**
CLOCS	70.4±1.11	68.1±3.1	71.9±2.2
DEAPS	75.8±1.1	74.4±2.9	76.6±2.5

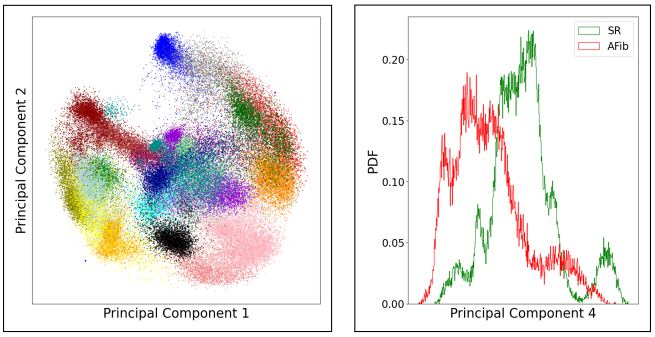
Table 2: Evaluation of gender classification task. Five-cross evaluation for the SVC model fitted on top of the representations.

Physionet Challenge 2017

In this final experiment, we utilized the Cinc2017 database, which categorizes ECG time series strips as either Sinus Rhythm (SR), AFib, or Others. We used the dataset's predefined partitioning of "train" and "validation" sets for evaluating the SVC model fitted on top of the representations. The findings of this experiment are summarized in Table 3, wherein DEAPS achieves superior performance compared with the rest of the methods.

	PHYSIONET CHALLENGE 2017			
SSL METHOD	TOTAL ACCURACY (%)	SR ACCURACY (%)	AFIB ACCURACY (%)	OTHER ACCURACY (%)
Mixing-Up	74.9	74.3	85.4	71.0
TF-C	60.2	62.0	**100**	44.3
PCLR	77.5	75.0	88.3	79.5
BYOL	78.0	76.5	84.5	79.0
CLOCS	74.8	72.7	84.7	77.1
DEAPS	**80.1**	**77.4**	86.0	**85.2**

Table 3: Evaluation on Physionet Challenge 2017. The original train-validation split has been used when fitting the SVC model.



(a) Subject-Based Analysis (b) Episode-Based Analysis

Figure 4: A PCA has been fitted on top of the MIT-AFIB representations. The values of the different components have been studied for disentangling the static and dynamic features.

4.2 PCA Analysis on Representations

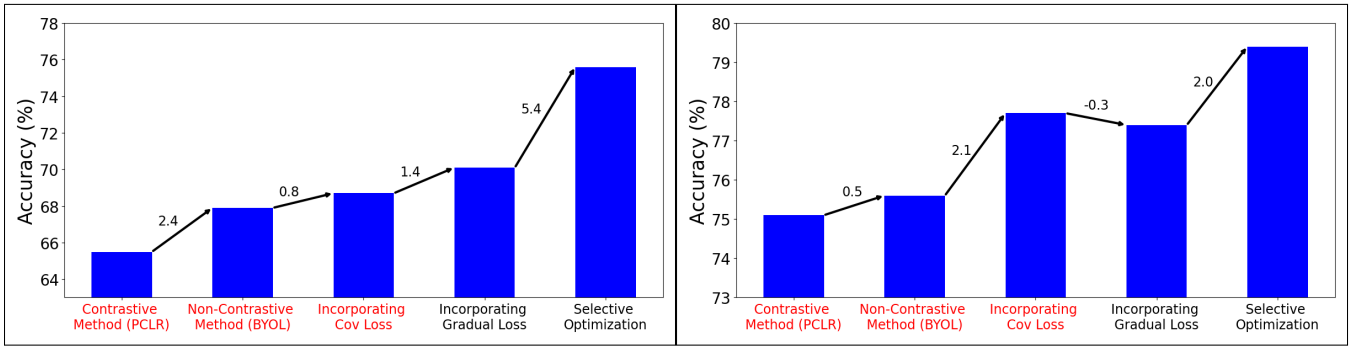
The fundamental concept underpinning this approach revolves around the integration of dissimilarities to facilitate the encoding of both static and dynamic features within the model. To ascertain the validity of this premise, a Principal Component Analysis (PCA) [Tipping and Bishop, 1999] is executed on the representations from MIT-AFIB database.

As depicted in Figure 4a where different patients are displayed in different colors, temporal segments originating from the same individual exhibit analogous positions in Principal Components 1 and 2, irrespective of the individual's status. Conversely, Principal Component 5 yields distinct values between AFib and normal rhythm (NR) classifications, regardless of the individual, as shown in Figure 4b. This second behavior is not seen in Contrastive Learning methods (See Appendix).

4.3 Discussion of the Results

This study has demonstrated that by following a non-contrastive methodology in addition to introducing a training objective ($\mathcal{L}_{gra}$) to focus on the dynamic patterns, both static and dynamic natures of time series data are encoded by the model (Figure 4). It leads to a performing model, which can be used in a wide spectrum of downstream tasks, using different types of datasets (Tables 1, 2 and 3). Compared to related methods, DEAPS shows a superior performance.

These results show substantial support for the hypothesis: (i) the use of Contrastive Learning techniques, by introducing an objective that prioritizes that the representations of different recordings are far apart, causes the model to neglect the dynamic patterns, (ii) Dynamic patterns are much less evident than static ones, so avoiding the use of negative pairs is not enough. An additional objective function is needed to lead the model to capture them. (iii) By effectively capturing these dynamic patterns, the model will acquire a better generalization capability by just considering a few labeled records.



(a) MIT-ARR → MIT-AFIB

(b) LOO MIT-AFIB

Figure 5: Effect of incorporating different elements to the training procedure. Components inherited from other existing methods are labeled in red. The impact of adding each component is displayed by the difference in the performance of each downstream task.

5 Ablation Study

This ablation study evaluates the effect of incorporating the different parts of the proposed method. In addition, the effect of the effect of the configuration of the two hyperparameters of the method i.e. window size and number of features that are considered in the Selective Optimization.

Role of the Different Components

Starting with contrastive methods, DEAPS changes this contrastive goal to the student-teacher architecture proposed by BYOL work. In addition, it incorporates the covariance term from methods such as VIC-REG. In Figure 5 it can be seen how these two components, which are not part of the contributions of this study, slightly improve these Contrastive Methods. In addition, it shows how the incorporation of the gradual loss, together with the selective optimization, plays a decisive role in obtaining the results with which DEAPS significantly improves existing SOTA methods.

Role of the Different Hyperparameters

As it is described in Section 3.2 The size of the spacing window must be sufficiently broad to capture signal variations but sufficiently limited to ensure that the direction of these variations is constant. Table 4 shows the impact of this window size. Although the proposed configurations improve existing methods, the best performance level is achieved by considering inputs belonging to a two-minute window size.

Window Size (Seconds)	MIT-ARR → MIT-AFIB Accuracy(%)	MIT-AFIB LOO Accuracy (%)
90	71.3	77.3
120 (Proposed)	**75.5**	**79.4**
150	69.7	79.1

Table 4: Ablation Study on the length of the window size for the $\mathcal{L}_{gra}$ computation.

The last aspect to take into consideration is the number of features that are taken into account in each iteration, as described in Section 3.3. Empirically we have determined that they should be 32. Table 5b shows the results of different configurations, where it can be seen that the proposal obtains the best performance. It can be seen that all configurations achieve better performance than existing methods.

# Features	MIT ARR → MIT AFIB Accuracy(%)	MIT-AFIB LOO Accuracy (%)
16	69.3	79.2
32 (Proposed)	**75.5**	**79.4**
48	72.4	78.9

Table 5: Ablation Study on the number of features to be considered during the Selective Optimization

6 Contributions

This study presents compelling evidence elucidating the tendency of Contrastive Learning objectives to induce neglect of subtle changes within quasiperiodic time series data. Addressing this limitation, we introduce DEAPS as a novel SSL method, distinct from conventional SOTA Contrastive Learning approaches. By dispensing with negative pairs, our method mitigates the model's susceptibility to capture shifts and integrates the novel $\mathcal{L}_{gra}$ function. This function guides the model to effectively encode dynamic patterns. Consequently, the model's capacity to generalize across different classes is demonstrated. The practical impact of this advancement is highlighted through a notable improvement of up to 10% compared to existing methods.

We believe that this characterization of physiological signals in static and dynamic patterns as well as the incorporation of a specific training objective for dynamic ones will provide inspiration for future Time Series SSL methods.

Limitations

Only one database (SHHS) has been used to pre-train the model. Using time strips from different recordings for the computation of $\mathcal{L}_{sim}$ is found to be the best strategy. SHHS is the only database we have found with multiple recordings from the same subject. However, we believe we have mitigated this lack of distinct datasets for pre-training by involving multiple datasets in the evaluation.

References

[Bardes *et al.*, 2022] Adrien Bardes, Jean Ponce, and Yann LeCun. Vicreg: Variance-invariance-covariance regularization for self-supervised learning, 2022.

[Caron *et al.*, 2021] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers, 2021.

[Chen *et al.*, 2020] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations, 2020.

[Clifford *et al.*, 2017] Gari D Clifford, Chengyu Liu, Benjamin Moody, Li-wei H. Lehman, Ikaro Silva, Qiao Li, A E Johnson, and Roger G. Mark. Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017. In *2017 Computing in Cardiology (CinC)*, pages 1–4, 2017.

[Diamant *et al.*, 2022] Nathaniel Diamant, Erik Reinertsen, Steven Song, Aaron D. Aguirre, Collin M. Stultz, and Puneet Batra. Patient contrastive learning: A performant, expressive, and practical approach to electrocardiogram modeling. *PLOS Computational Biology*, 18(2):1–16, 02 2022.

[Dosovitskiy *et al.*, 2021] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

[Fan *et al.*, 2021] Fan, Haoyi, Zhang, Fengbin, Wang, Ruidong, Huang, Xunhua, and Zuoyong Li. Semi-supervised time series classification by temporal relation prediction. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3545–3549, 2021.

[Franceschi *et al.*, 2020] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. Unsupervised scalable representation learning for multivariate time series, 2020.

[Goldberger *et al.*, 2000] Ary Goldberger, Luís Amaral, L. Glass, Shlomo Havlin, J. Hausdorg, Plamen Ivanov, R. Mark, J. Mietus, G. Moody, Chung-Kang Peng, H. Stanley, and Physiotoolkit Physiobank. Components of a new research resource for complex physiologic signals. *PhysioNet*, 101, 01 2000.

[Grill *et al.*, 2020] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning, 2020.

[Kingma and Ba, 2017] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[Kiyasseh *et al.*, 2021] Dani Kiyasseh, Tingting Zhu, and David A. Clifton. Clocs: Contrastive learning of cardiac signals across space, time, and patients, 2021.

[Moody and Mark, 1983] G.B. Moody and R.G. Mark. A new method for detecting atrial fibrillation using r-r intervals. *Computers in Cardiology*, pages 227–230, 1983.

[Platt, 2000] John Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv. Large Margin Classif.*, 10, 06 2000.

[Quan *et al.*, 1998] Stuart Quan, Barbara Howard, Conrad Iber, James Kiley, F. Nieto, George O'Connor, David Rapoport, Susan Redline, John Robbins, Jonathan Samet, and ‡Patricia Wahl. The sleep heart health study: Design, rationale, and methods. *Sleep*, 20:1077–85, 01 1998.

[Tipping and Bishop, 1999] Michael E. Tipping and Christopher M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999.

[Tonekaboni *et al.*, 2021] Sana Tonekaboni, Danny Eytan, and Anna Goldenberg. Unsupervised representation learning for time series with temporal neighborhood coding, 2021.

[Wickstrøm *et al.*, 2022] Kristoffer Wickstrøm, Michael Kampffmeyer, Karl Øyvind Mikalsen, and Robert Jenssen. Mixing up contrastive learning: Self-supervised representation learning for time series. *Pattern Recognition Letters*, 155:54–61, mar 2022.

[Zbontar *et al.*, 2021] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction, 2021.

[Zhang *et al.*, 2018] Guo-Qiang Zhang, Licong Cui, Remo Mueller, Shiqiang Tao, Matthew Kim, Michael Rueschman, Sara Mariani, Daniel Mobley, and Susan Redline. The national sleep research resource: Towards a sleep data commons. *Journal of the American Medical Informatics Association*, pages 572–572, 08 2018.

[Zhang *et al.*, 2022] Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency, 2022.