IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 25 August 2025, accepted 6 September 2025,
date of publication 11 September 2025, date of current version 17 September 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3608901

RESEARCH ARTICLE

Augmentation-Free Longitudinal Modeling Through Structuring Whitened Embeddings

KAREL FONTEYN, LENNERT BONTINCK, (Graduate Student Member, IEEE), TOM DHAENE, (Senior Member, IEEE), AND DIRK DESCHRIJVER, (Senior Member, IEEE)
IDLab, Ghent University - imec, 9052 Ghent, Belgium

Corresponding author: Karel Fonteyn (karel.fonteyn@ugent.be)

This work was supported in part by the Flemish Government through the AI Research Program, and in part by the Interreg France-Wallonie-Vlaanderen (FWVL) Program and the Province Oost-Vlaanderen as part of the VasculAI Project.

ABSTRACT This paper introduces Structuring Whitened Embeddings, a modality- and encoder-agnostic framework designed to optimize encoders for extracting smooth, progression-aware feature representations by aligning real-world samples. Proportional inter-sample relationships are preserved, enabling the capture of subtle and continuous changes. By establishing relationships directly between samples and thereby avoiding the need for data augmentation, the approach is particularly well suited for transformation-sensitive data, such as medical time series, where even minor sample changes can lead to disproportionate shifts in interpretation. Experimental results on early atrial fibrillation prediction and timestamp imputation, modeling both inter- and intra-patient dynamics, demonstrate significant performance improvements using the optimized features. The framework's augmentation-free design and generalizability across tasks and modalities position it as a practical solution for modeling evolution in complex datasets.

INDEX TERMS Biomedical monitoring, deep learning, feature extraction, patient monitoring, predictive analytics, predictive models, prognostics and health, representation learning, time series analysis, wearable health monitoring.

I. INTRODUCTION

Many modern deep learning (DL) feature extraction approaches rely on augmentation, either to oversample limited datasets or to generate samples for self-supervised learning. However, augmentation necessitates careful design to preserve the true data characteristics, demanding domain-specific knowledge and modality-aware techniques. For transformation-sensitive data, such as medical time series, inadequate augmentation can distort critical patterns, leading to degraded model performance and poor generalization [1]. This issue is particularly acute in domains where subtle, often imperceptible patterns are essential for decision-making. In such settings, even minor signal alterations can render a sample unrepresentative of its class [2]. Furthermore, preserving valid temporal dependencies is crucial, underscoring the need for algorithms that learn directly from real, unaugmented data [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Gerard-Andre Capolino.

This paper introduces the Structuring Whitened Embeddings (SWE) framework, a modality- and encoder-agnostic approach that enhances feature representations by aligning inter-sample relationships. SWE optimizes existing encoders to detect subtle, otherwise overlooked features and learns continuous representations that evolve smoothly across aligned samples. Central to this framework is the concept of Preserving Proportional Relations (PPR), which formalizes how relationships between input samples can be expressed as proportions derived from pairwise distances and mimicked in an auxiliary embedding. By modeling these relationships using only real-world samples, SWE eliminates the need for data augmentation to improve feature extraction. Rather than isolating features that distinguish a specific target, SWE captures how samples evolve with respect to underlying variables through alignment, resulting in smooth, longitudinally-aware representations that reflect subtly progressing features.

The biomedical domain, particularly ECG analysis, exemplifies the challenges associated with transformation-sensitive data. While deep learning has enabled automated

diagnosis from ECGs, a notable gap persists in models that can predict the onset of cardiac conditions such as atrial fibrillation before clinical manifestation [4]. These tasks require models capable of detecting subtle, progressive changes without relying on augmentation, making them ideal testbeds for evaluating augmentation-free, longitudinally aware frameworks such as SWE. Both inter- and intra-group relationships are modeled in the experimental settings by aligning cardiograms based on patient age or relative time of recording. The resulting longitudinal representations capture subtle features indicative of early signs of slow-onset conditions such as atrial fibrillation and improve timestamp imputation from sequences of patient cardiograms. Although evaluated in a biomedical context, the SWE framework is broadly applicable to any domain involving transformation-sensitive data.

The main contributions of this work are three-fold:

- **Conceptual** introduction of PPR for aligning inter-sample relations in latent space.
- **Methodological** implementation of the SWE framework using whitening and structured subbatching to enforce PPR.
- **Empirical** demonstration and analysis of SWE's effectiveness across multiple tasks.

The remainder of this paper is organized as follows: Section II reviews related work; Section III details the theoretical foundations of PPR (III-A) and the implementation of SWE (III-B); Section IV presents the experimental evaluation; Section V discusses the findings; and Section VII concludes the paper.

II. RELATED WORK

A recent shift towards anticipatory modeling in healthcare, where extracted features are used for early risk prediction, highlights the need to detect subtle, progressive patterns [4], [5]. Bontinck et al. [6] introduced ECGencode, a compact network inspired by Filter Bank Common Spatial Patterns, capable of predicting new-onset atrial fibrillation (AF) and detecting AF in normal sinus rhythm ECGs. Although attention-based models are gaining popularity in ECG analysis [7], ResNets remain widely used due to their lower parameter count, faster training, and strong performance in capturing local temporal features critical for arrhythmia detection [8], [9]. Sau et al. [10] demonstrate that their ResNet-based model can uncover latent pathophysiological signals not visible to human interpretation, enabling the prediction of incident hypertension and stratification of risk for major cardiovascular events, such as myocardial infarction and heart failure, even in individuals with normal ECGs and no overt symptoms.

Pham et al. [11] advance this direction by introducing dynamic ECG remodeling as a prognostic marker for sudden cardiac death. Their work demonstrates that monitoring progressive ECG changes significantly enhances risk stratification beyond static ECG features and clinical variables, thereby underscoring the value of longitudinal modeling in healthcare and validating the use of progressively changing physiological signals for early risk prediction. Additionally, longitudinal modeling approaches for irregularly sampled medical time series emphasize the need for deep learning methods capable of capturing both local and global temporal structures across heterogeneous sampling patterns [12].

Extending the notion of progression-aware modeling, relational reasoning provides a complementary approach for capturing structure in complex data. By establishing relationships between connected entities, it enables more nuanced and context-aware feature representations [13]. Unlike traditional feature extraction methods that compare individual samples to target labels, relational reasoning approaches relate a sample to an entire set, thereby facilitating a richer contextual understanding [14], [15]. In self-supervised representation learning, relational reasoning has been proposed as a means of modeling distances between data points, allowing for a continuum of similarities [16]. This contrasts with the binary distinctions captured by contrastive learning, which typically compares anchors with augmented positives and unrelated negatives. While contrastive methods are widely applied to time series data such as ECGs [17], [18], [19], their effectiveness is often constrained by the temporal dependencies inherent in such data and their sensitivity to distortions introduced by augmentations [20]. Relational reasoning offers an alternative by leveraging temporal consistency. While this consistency makes time series data difficult to augment, it enables effective modeling of longitudinal evolution.

A key challenge in self-supervised feature learning without explicit targets is the risk of representation collapse, where the model outputs degenerate to a constant vector [21], [22]. Grill et al. [23] address this in BYOL using an asymmetric architecture and a momentum encoder to maintain representation diversity. Alternatively, feature decorrelation can be achieved through regularization. Ermolov et al. [24] batch samples in the feature space, whereas Zbontar et al. [25] demonstrate in Barlow Twins that aligning feature pairs while minimizing redundancy across vector components can "soft-whiten" representations. Zhang et al. [26] further show that decorrelation can be achieved by introducing base vectors in the latent space and maximizing consistency across views. These approaches offer promising alternatives to contrastive learning, particularly for time series data where augmentation-based methods may be less effective.

The proposed SWE framework builds on relational reasoning by encoding the structural relationships between samples onto an auxiliary hyperspherical embedding. While this embedding provides flexibility, direct label supervision is not feasible due to its unstructured nature. Instead, a self-supervised strategy is employed, wherein the relative proportions between samples are established through pairwise distances, and the learning objective is to minimize the discrepancy between the proportions derived from the input and corresponding embedding representations. To prevent a collapsed solution, in which all samples are mapped to a

single point, whitening is applied to encourage dispersion within each batch. By enforcing relational structure through the preservation of inter-sample proportions, while ensuring representation variability through whitening, SWE offers a compelling alternative to augmentation-based contrastive learning approaches, which are often suboptimal for temporally sensitive data. Additionally, by leveraging the time between samples, SWE is well suited for longitudinal ECG analysis and for monitoring individuals at elevated risk before clinical manifestation. This is evidenced by improved performance across multiple ECG tasks, including early prediction of future cardiac conditions and imputation of missing time points.

III. METHODOLOGY

The SWE framework aligns sample representations in an auxiliary embedding to optimize the features extracted by an encoder. By modeling inter-sample relationships in this embedding space through PPR, the model learns how samples are interconnected, yielding insights into how they change in response to underlying variables. Since SWE operates by adding auxiliary modules without modifying the core architecture, it remains modality-agnostic and can be applied to any backbone encoder $E(.)$. Once optimized, the encoder $E(.)$ benefits from this alignment strategy and can be reused for downstream tasks, as its features now reflect meaningful inter-sample relationships. This makes the framework particularly well-suited for applications involving temporal progression, imputation, or the monitoring of subtle patterns.

Rather than mapping samples to absolute targets, SWE aligns abstract sample representations in an unbounded embedding space using proportions derived from metrics that quantify inter-sample relationships. In the input space, such a metric may be defined using differences between continuous target labels, while in the embedding space, it quantifies the distances between the mapped sample representations.

This section introduces the concept of PPR, followed by a description of how SWE implements this in practice. A notation table summarizing the symbols used in this section is provided in Appendix A.

A. PRESERVATION OF PROPORTIONAL RELATIONS

Given a dataset $\mathcal{D}$, consisting of samples $\mathbf{x}$ with associated labels $\mathbf{y}$, where each y_i enables the definition of inter-sample relationships over a subset $\mathcal{S} \subseteq \mathcal{D}$, the goal is to preserve these relationships in the embedding space. For a subset $\mathcal{S}_j = x_1, x_2, \ldots, x_n$ with $n \geq 3$, a relational metric function δ is applied to labels $\mathbf{y}$ to quantify pairwise relationships, for which $\mathcal{S}_j$ can also be referred to as a relational set.

Let $f(.) = g(E(.))$ be a transformation mapping samples $\mathbf{x}$ to an auxiliary embedding and let ϕ be a relational metric function defined in this embedding space. The preservation of proportional relations in $\mathcal{S}_j$ puts the distance between an anchor x_a and a sample x_b of the same relational set in relation to the sum of distances between x_a and any sample in S_j. This can be formulated as: $\forall x_a, x_b \in \mathcal{S}_j \wedge x_a \neq x_b$:

$$\frac{\delta(x_a, x_b)}{\sum_{x_i \in \mathcal{S}_j} \delta(x_a, x_i)} \approx \frac{\phi(f(x_a), f(x_b))}{\sum_{x_i \in \mathcal{S}_j} \phi(f(x_a), f(x_i))} \tag{1}$$

and is also illustrated in Fig. 1.

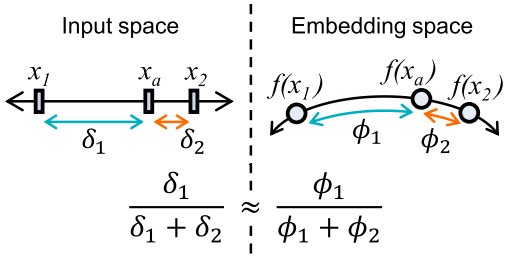

FIGURE 1. Visualization of the PPR concept. With $\delta_1 = \delta(x_a, x_1)$, $\phi_1 = \phi(f(x_a), f(x_1))$ and δ_2, ϕ_2 analogously. The proportions between the anchor x_a and any other sample in the input space are preserved in the embedding after the samples are transformed by $f(.)$.

The loss function for PPR is derived from this equation. To enforce the structure between the input space and the embedding space to match, it minimizes the squared error between the proportions of pairwise distances in the input space (quantified by δ) and their counterparts in the embedding space (quantified by ϕ) for an anchor sample relative to all other samples in the relational set.

Given relational set $\mathcal{S}_j$ and anchor sample $x_a \in \mathcal{S}_j$, the PPR error $\lambda_{\mathcal{S}_j}(x_a)$ and is defined as: $\lambda_{\mathcal{S}_j}(x_a) =$

$$\sum_{x_b \neq x_a} \left| \frac{\delta(x_a, x_b)}{\sum_{x_i \in \mathcal{S}_j} \delta(x_a, x_i)} - \frac{\phi(f(x_a), f(x_b))}{\sum_{x_i \in \mathcal{S}_j} \phi(f(x_a), f(x_i))} \right|^2 \tag{2}$$

The final loss for the relational set for $\mathcal{S}_j$ with anchor x_a is obtained by averaging the PPR error λ over all non-anchor samples:

$$loss_{ppr}^{\mathcal{S}_j, x_a} = \frac{\lambda_{\mathcal{S}_j}(x_a)}{|\mathcal{S}_j| - 1} \tag{3}$$

By minimizing the error between proportions, $E(.)$ is encouraged to extract features that quantify changes between samples, resulting in representations that evolve smoothly across aligned samples. Because the auxiliary embedding is unbounded and only relative positions matter, the model is free to learn a flexible representation. This contrasts with traditional regression models, which constrain the output—and by extension, the network weights—to absolute targets.

However, as the model is trained to predict each set of normalized pairwise distances derived from the input space via δ, using $\phi(f(x))$ in the embedding space, and the loss is defined over a set of targets with predictions produced by a learnable function $f(x)$, PPR can be reduced to a standard regression problem. While a complete mathematical proof of convergence is out of scope for this paper, conceptually, PPR is a regression on relational proportions from input space to embedding space. Therefore, it inherits the convergence properties of supervised regression: the loss is differentiable, bounded below, and—given the smoothness

of δ and ϕ—locally convex in the predicted proportions, allowing gradient-based optimization methods to minimize the error effectively.

B. STRUCTURING WHITENED EMBEDDINGS

SWE implements the concept of PPR by mimicking inter-sample structures in a hyperbolic, whitened projection space. The auxiliary modules first normalize the representations with their L^2-norm, followed by a batch-wise whitening operation that distributes the samples uniformly over the hypersphere. The transformation $f(.)$ is implemented as $w(h(E(.)))$, where $h(.)$ maps the high-dimensional output of the encoder $\mathbf{e} = E(\mathbf{x})$ onto a lower-dimensional hypersphere $\mathbf{h} = h(\mathbf{e})$, and $w(.)$ performs the whitening operation, yielding the whitened embeddings $\mathbf{w} = w(\mathbf{h})$. A graphical overview of SWE is shown in Fig. 2.

The relational metric ϕ is implemented as the cosine similarity between samples projected onto the whitened hypersphere, enabling the capture of angular relationships that reflect the intrinsic geometry of the embedding space. Unlike Euclidean distance, which measures chord length and introduces distortions at larger angles, cosine similarity more faithfully preserves proportional relationships in the embedding with isotropic variance. Additionally, it avoids the computational overhead and potential numerical instability associated with inverse trigonometric operations required for geodesic distance computation, making it especially suitable for gradient-based optimization in learning systems. The use of cosine similarity also aligns with prior work such as Ermolov et al. [24], and is further supported by recent geometric perspectives in machine learning [27].

Proportional enforcement is most effective when the dimensionality of $\mathbf{h}$ is minimal yet allows sufficient projection freedom. Empirically, both excessively high and low dimensionalities were found to degrade downstream performance, though feature learning remains stable across a broad intermediate range. The dimensionality of $\mathbf{e}$ must match or exceed that of $\mathbf{h}$. Whitening operation $w(.)$ is based on the transform proposed by Siarohin et al. [28], using a fully differentiable Cholesky decomposition [29] to decompose the symmetric covariance matrix $\Sigma_{\mathbf{h}}$ of a batch of embeddings $\mathbf{h}$ into two triangular matrices $\Sigma_{\mathbf{h}} = LL^T$, where L is the lower triangular matrix. The inverse of L is then used to compute the whitening matrix $W_{\mathbf{h}} = L^{-1}$. The whitening operation is then defined as $w(\mathbf{h}) = W_{\mathbf{h}}(\mathbf{h} - \mu_{\mathbf{h}})$, where $\mu_{\mathbf{h}}$ is the mean of the embeddings in $\mathbf{h}$. The resulting whitened embeddings $\mathbf{w} = w(h(\mathbf{e}))$ have identity covariance, with each element having unit variance, thereby decorrelating $\mathbf{w}$ [30].

C. SUB-BATCHING

Due to the high variability of the whitening matrix $W_{\mathbf{h}}$ across batches, care must be taken in how the data are structured. Applying the same whitening matrix to samples that are contrasted with one another can introduce bias and should be avoided [24]. To address this, each sample in a relational set is transformed using a distinct whitening matrix.This is achieved by creating n sub-batches, where each sub-batch is formed by taking every k^{th} element across all $\mathcal{S}_j$, where m is the batch size, and n is the number of samples per relational set. This ensures that elements within the same $\mathcal{S}_j$ are not transformed by the same whitening matrix $W_{\mathbf{h}}^k$. Using distinct matrices for samples within a relational set enables alignment of these samples, preserving relational structure, while avoiding intra-set competition. A graphical representation of this sub-batching strategy is presented in Fig. 3.

IV. EXPERIMENTAL SETTING

Two experimental settings are examined. The first assesses the capabilities of SWE in prognostic tasks, exemplified by the early detection of AF. The second evaluates the continuity of longitudinal features through an imputation task.

The publicly available CODE-15 dataset [31], comprising 345,779 annotated 12-lead ECG recordings from 233,770 patients sampled at 400 Hz and collected across 811 municipalities in Brazil, is used to demonstrate the applicability of SWE. Its large number of recurring patients and prior use in related work make it well suited for benchmarking model performance on longitudinal patterns. The ECGs, originally recorded at lengths varying between 7 and 10 seconds, are padded to 10 seconds. To avoid bias from padded regions, the central 6.4 seconds—consistently containing real signal while avoiding potential inaccuracies at the start and end—are used for testing, while training samples are generated by randomly extracting 6.4-second windows from each ECG.

The experiments employ patient-based cross-validation, dividing the data into 10 folds. For each iteration, a validation fold is left out for early stopping and threshold selection for the downstream task. Apart from baseline removal and normalization, no additional pre-processing is performed, and augmentations are avoided, as even minor alterations could render ECG signals inconsistent with their clinical labels.

During SWE optimization, sub-batch sizes are configured relative to the relational set sizes to match the batch size used during downstream training. With a downstream batch size of 256, SWE uses a batch size of 64 and relational sets of 4 samples, resulting in $64 \times 4 = 256$ ECGs per batch. This ensures that memory requirements remain consistent between optimization and downstream training. Additional training hyperparameters are listed in Appendix B. To assess statistical significance, the normality of model performance distributions is first verified using the Shapiro-Wilk test (summarized in Appendix C), followed by a paired t-test to establish p-values (p). Effect sizes are reported using Cohen's d (d), and 95% confidence intervals (95% CI) are reported for the main metric.

A. UNCOVERING SLOW ONSET FEATURES

Although automated detection of manifested cardiac conditions is well established, early detection remains a

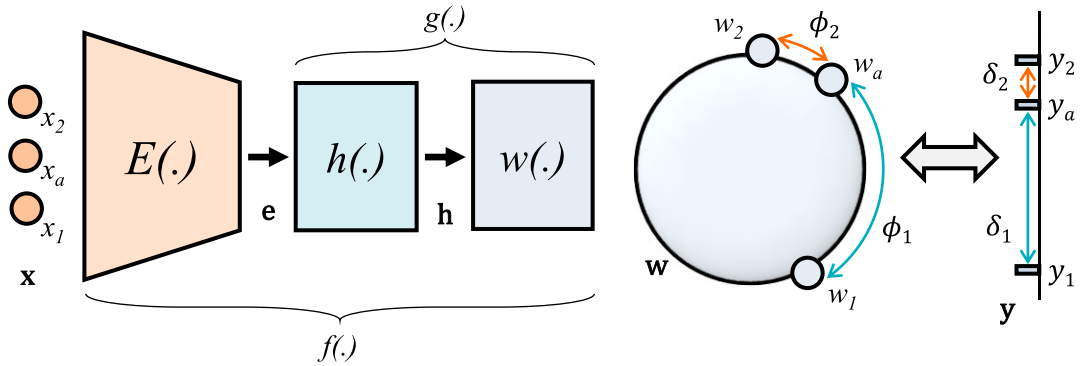

FIGURE 2. Graphical overview of SWE. The samples **x**, for which **y** defines the inter-sample relationships, are mapped onto a whitened hypersphere as representations **w**. Proportional relations are enforced by aligning the embedding space metric distances ϕ with their corresponding input space distances δ.

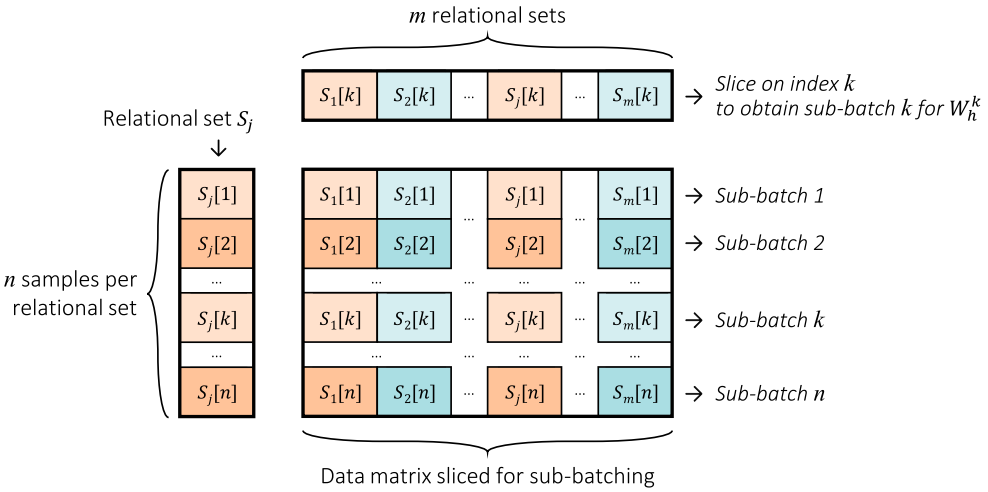

FIGURE 3. Sub-batching for SWE. m relational sets consisting of n elements each, are sliced to ensure that samples contrasted with each other (samples from the same $\mathcal{S}_j$) are transformed using different whitening matrices W_h, resulting in n sub-batches.

challenge. This experiment demonstrates the benefit of SWE in enhancing anticipatory prognoses by extracting subtle precursor features from medical data, thereby outperforming augmentation-based approaches.

1) SETUP

To mitigate bias towards individual patients, only the first ECG recording per patient is selected, yielding 233,770 ECGs in total, equally divided over 10 folds. The network architecture consists of a ResNet-like feature extractor $E(.)$, based on the model originally proposed for detecting six cardiovascular diseases in ECG signals [9], [31]. A regressor, implemented as a single dense layer with linear activation, processes the extracted features to produce the model output. For this experimental setting, each relational set $\mathcal{S}_j$ for SWE comprises four randomly selected ECGs from different patients, used to pre-optimize the weights of E by mimicking differences in patient age at the time of ECG acquisition.

2) EXPERIMENT: AGE PREDICTION

Features optimized for longitudinal evolution using the SWE framework are particularly advantageous when the downstream task shares the same label used to establish pairwise relations in SWE. Therefore, the encoder optimized via SWE is first assessed for its ability to capture age-related changes and improve age prediction performance.

Two models are compared: a baseline trained from scratch for age prediction and a variant with an SWE-optimized encoder. Following SWE optimization, the entire network undergoes an additional fine-tuning phase to enable accurate age predictions. The results verify that SWE-optimized features significantly improve performance with a large effect ($p < 0.003$, $d > 1.3$), reducing the mean absolute error from 7.17 years (95% CI: ± 0.07) to 7.04 years (95% CI: ± 0.04).

3) EXPERIMENT: FUTURE ATRIAL FIBRILLATION

Although age prediction from ECGs is not clinically actionable, the age label also serves as a proxy for extracting

longitudinal features relevant to predictive tasks. The progression of chronically advancing conditions such as atrial fibrillation (AF) is assumed to correlate with age-related changes [32], making the time-optimized features extracted by the SWE-optimized encoder valuable for detecting its slow onset.

In addition to the unoptimized encoder from Section IV-A2, the SWE-optimized encoder is compared to a BYOL-optimized encoder, equally initialized with weights from the age prediction model. Given BYOL's demonstrated effectiveness on ECG data, it serves as a strong benchmark for augmentation-based self-supervised learning. The augmentations used during BYOL optimization—random resized crop and time out—are relatively mild and were empirically identified as most effective by Mehari and Strodthoff [33]. While large-scale self-supervised pretraining offers notable advantages in low-data regimes, its relative benefit tends to diminish as the volume of labeled data increases [34]. To better isolate the contribution of the optimized encoders, a single-fold training strategy is employed for fine-tuning, with exhaustive evaluation across the left-out folds. Thresholds for all models are determined by maximizing the F1 score on the validation fold.

The encoders are fine-tuned to predict whether a patient currently has or will develop AF. Following the approach in [6], the networks are trained using binary labels indicating whether the patient is known to have AF. Since only the first available ECG per patient is used for training, the label reflects either the presence of AF in the current recording or a future diagnosis. This setup classifies patients as vulnerable to AF (pAF), with approximately 420 such instances per fold. Results summarized in Table 1 show that SWE significantly improves upon the baseline features in terms of F1 score in this prediction setting ($p < 0.01$, $d > 1.0$), and also significantly outperforms BYOL-based feature optimization, with a large effect size for the F1 score improvement ($p < 0.02$, $d > 0.9$).

To further isolate the model's ability to detect precursor features, a new threshold is computed using only patients who develop AF at a later stage and whose training ECG does not yet show AF. This threshold is then applied to predict future AF (fAF), thereby highlighting the model's capacity to detect early indicators of the disease. Each fold contains approximately 128 such cases, with a maximum time to onset of seven years across the dataset. The mean and median times to onset are 625 and 470 days, respectively. Similar to the pAF setting, SWE significantly improves the F1 score over the baseline ($p < 0.002$, $d > 0.3$), and significantly outperforms BYOL for feature optimization ($p < 0.05$, $d > 0.2$). Notably, BYOL's improvement is driven by increased precision, whereas SWE yields a gain in sensitivity. These results verify that the SWE-optimized features enhance the model's ability to identify patients at risk of developing AF. The fAF results are summarized in Table 2.

TABLE 1. pAF — Performance metrics for identifying patients with AF presence and/or future AF recordings. All models are initialized with weights from the age prediction task and further optimized using the technique specified. The F1 scores are reported with 95% confidence intervals.

Opt.	F1 (95% CI)	Sens.	Spec.	Prec.	Auprc
None	0.6323 (±0.0047)	0.6156	0.9931	0.6544	0.6039
BYOL	0.6324 (±0.0044)	0.6099	0.9933	0.6602	0.5959
SWE	**0.6439** (±0.0061)	**0.6206**	**0.9936**	**0.6723**	**0.6159**

TABLE 2. fAF — Performance metrics for identifying patients with future recordings containing AF, despite no current AF presence. All models are initialized with weights from the age prediction task, further optimized using the technique specified. The F1 scores are reported with 95% confidence intervals.

Opt.	F1 (95% CI)	Sens.	Spec.	Prec.	Auprc
None	0.1670 (±0.0061)	0.1957	0.9940	0.1598	0.0872
BYOL	0.1707 (±0.0066)	0.1950	0.9942	0.1617	0.0811
SWE	**0.1790** (±0.0064)	**0.2020**	**0.9943**	**0.1679**	**0.0913**

B. LEVERAGING CONTINUOUS FEATURES FOR TIMESTAMP IMPUTATION

Time series imputation plays a vital role in healthcare, where data quality is critical, labels are sparse, and sampling is irregular [35], [36]. This experiment explores timestamp imputation as an alternative setting to illustrate embedded continuity and progression-awareness of the learned representations. SWE's enhanced ability to capture longitudinal continuity is expected to improve the accuracy of imputing missing values in time series data.

1) SETUP

The human heart signal can be conceptualized as a single continuous time series, with ECG segments representing discrete snapshots. From a sequence of such snapshots, the task is to infer the missing timestamp values. Each sequence consists of a series of records r from a single patient, where each record includes an ECG signal e and a corresponding timestamp t. The timestamps are relative, with the first record in each sequence set to 0. Exactly one timestamp per sequence is missing, excluding the first one. Sequence lengths range from $n_s = 3$ to $n_s = 6$.

Feature extraction is performed using the same encoder architecture E, as described in Section IV-A. For each record, the timestamp is concatenated to the extracted ECG features before being passed to a prediction network, consisting of a bidirectional GRU layer with 32 neurons. The missing timestamp is encoded as -1. The network output consists of the predicted timestamps for all records, including the imputed values. A graphical overview is presented in Fig. 4.

To ensure consistency across all values of n_s, only patients with at least six consecutive ECGs are selected. If multiple records exist for a single day, only the first record is used, resulting in a dataset of 8,521 ECGs from 1,231 patients.

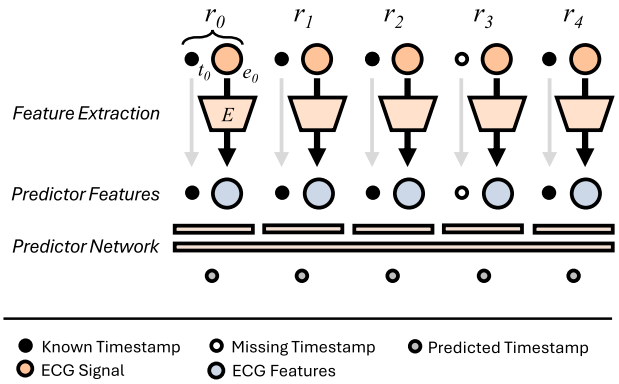

FIGURE 4. Architecture overview of the imputation experiment. The encoder (*E*) extracts features from the ECG signals (*e*), which are concatenated with their corresponding timestamps (*t*). The prediction network integrates features from all records to output a corrected series of timestamps, including the imputed value.

For patients with more than n_s recordings, all valid combinations of n_s records are sampled. For SWE optimization, each relational set $\mathcal{S}_j$ is constructed from ECGs of a single patient, using inter-sample time intervals to define proportional relationships. The root mean squared error is normalized (NRMSE) to account for increased range saturation in longer sequences.

2) EXPERIMENT

The feature extraction network E is initialized with weights from the age prediction model in Section IV-A2 to ensure a fair comparison. While the previous SWE optimization structures the embedding by comparing age across patients, the current experiment sets up SWE to capture longitudinal evolution within individual patients. The relational sets for SWE consist of sequences of recordings from a single patient, enabling the auxiliary embedding to model generalizable patient trajectories.

The prediction networks are trained by minimizing the mean squared error over the entire sequence of time points. Results of this experiment are presented in Table 3. For sequence lengths $n_s = 4$ to $n_s = 6$, the improvement in NRMSE for the missing timestamp is statistically significant with a large effect ($p < 0.005$, $d > 1.1$), with respective relative improvements of 2.33%, 4.42%, and 2.63%. A similar improvement ($p < 0.03$, $d > 0.8$) is observed in the NRMSE over the entire sequence, supporting the conclusion that SWE effectively captures longitudinal patient trajectories in the encoder's feature space.

C. SYNTHETIC INTERPRETABLE EXAMPLE

This final synthetic example provides an interpretable verification of the continuity captured in the feature embeddings by SWE through visualization of the latent spaces. It also demonstrates the modality-agnostic nature of SWE by applying it to image-like inputs, illustrating applicability beyond one-dimensional time series.

TABLE 3. Performance metrics for imputing missing time points in sequences of varying lengths ($n_s = 3$ to $n_s = 6$) across different models. Errors are expressed in days. Metrics include normalized root mean squared error (NRMSE) and mean absolute error (MAE), reported for both the imputed time point and the average error across the full sequence (s.). Results using the age prediction model from Section IV-A2 are included for reference. The relative improvement for the SWE-optimized model over baseline model in terms of NRMSE is 0%, 2.33%, 4.42%, and 2.63%, for sequence lengths $n_s = 3$ to $n_s = 6$, respectively.

n_s	Model	NRMSE (95% CI)	MAE	S. NRMSE	S. MAE
3	Age P.	585.4 (±3.7)	621.4	370.6	254.5
	Base	**213.3** (±4.5)	200.2	107.6	55.4
	SWE	**213.3** (±5.1)	**198.0**	**107.4**	**54.3**
4	Age P.	522.6 (±4.4)	705.9	342.3	307.0
	Base	132.7 (±4.6)	147.3	60.5	36.1
	SWE	**129.6** (±4.1)	**143.8**	**58.9**	**34.6**
5	Age P.	489.1 (±3.3)	771.9	325.9	346.1
	Base	99.4 (±4.9)	118.6	41.2	25.1
	SWE	**95.4** (±4.3)	**115.4**	**39.8**	**24.9**
6	Age P.	467.8 (±4.6)	819.7	315.9	376.0
	Base	76.2 (±6.8)	99.5	29.7	20.3
	SWE	**74.2** (±6.1)	**95.9**	**28.7**	**18.7**

1) SETUP

A setting analogous to reading the time of an analog clock is considered. A synthetic dataset is created in which a 12-hour clock is rendered as a 16 × 16 pixel image with RGB color channels (see Fig. 5). The clock face includes two anchor lines as reference points to support consistent interpretation, as the training set contains multiple rotated instances of the same time point. A simple convolutional neural network (CNN) with three layers is used to extract features, followed by a single dense layer with one neuron for regression.

2) EXPERIMENT

A synthetic dataset of 720 images is generated by applying a random second offset to each minute across the 12-hour range represented by the clock. Twelve time points are selected as test samples and excluded from the training set, along with all time points within 15 minutes before or after each excluded point. The CNN feature extractor is optimized using SWE with relational sets comprising four randomly selected time points.

Dimensionality-reduced visualizations of the embeddings are shown in Fig. 6. After SWE optimization, the auxiliary

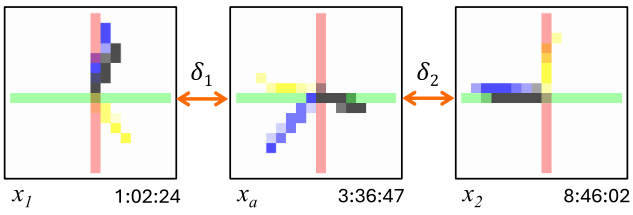

FIGURE 5. Illustrative input for the synthetic example. Two samples, x_1 and x_2, are aligned with the anchor x_a. The metric distance δ quantifies their respective distances to the anchor.

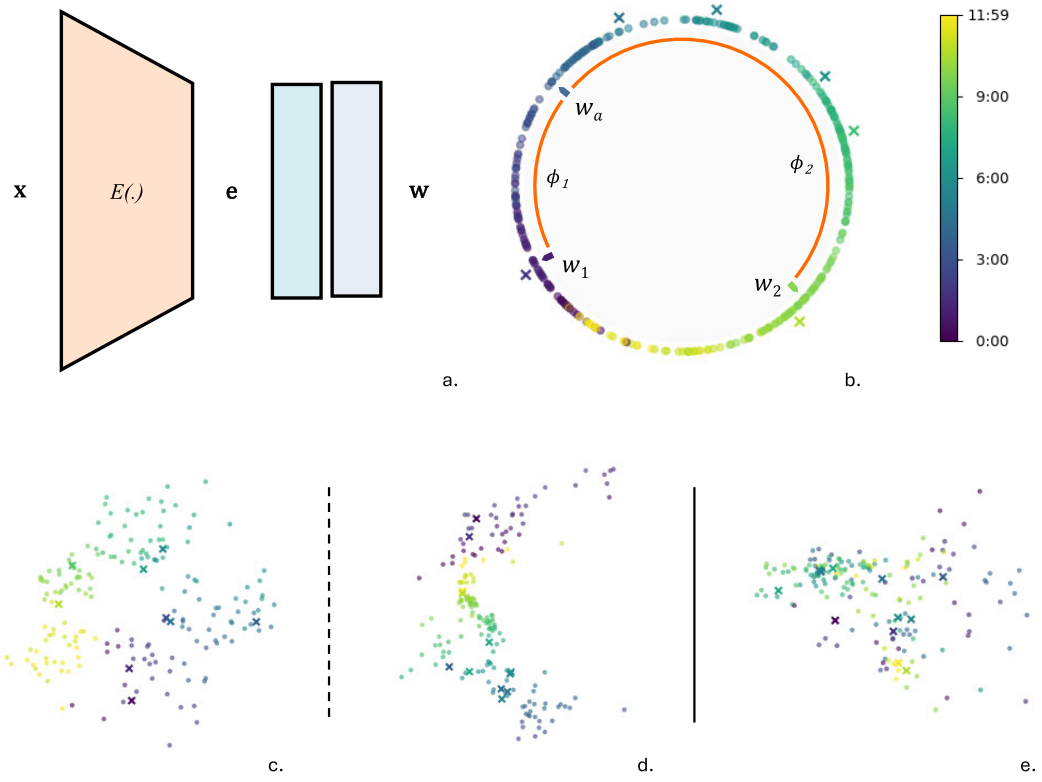

FIGURE 6. Latent space visualizations of the synthetic example. Each dot represents a single training sample; six selected validation samples are highlighted as crosses. (a) provides a reference for the network's variables. (b) shows the whitened embedding **w**, where the inter-sample structure is enforced, with ϕ quantifying the metric distances between embedded samples. (c) visualizes the encoder output **e** immediately after SWE optimization, and (d) shows **e** after fine-tuning for absolute value prediction. (e) serves as a baseline, showing **e** from a model without SWE optimization.

embedding **w**, where SWE enforces the structure, clearly captures the cyclic nature of the clock and preserves its continuity over time. In the bottom row of Fig. 6, the encoder output **e** is visualized, indicating that the cyclic structure of **w** is also captured in the encoder's output. After fine-tuning the SWE-optimized encoder, the latent space is unwrapped to accommodate the dense regressor.

V. DISCUSSION

The experiments demonstrate the applicability, relevance, and benefits of employing SWE to extract augmentation-sensitive features and obtain continuous embeddings. The augmentation-free nature of SWE, combined with its ability to align samples, enables it to excel at the detection of subtle differences, as demonstrated by its improved performance in identifying slow-onset phenomena. In addition, the alignment of samples by SWE promotes continuity in the extracted features, enabling improved performance in imputation tasks.

A. EXTRACTION OF SUBTLE FEATURES

The results in Tables 1 and 2, along with the improvement in age prediction accuracy in Section IV-A2, demonstrate that the SWE framework enables the encoder to extract features that highlight subtle, otherwise difficult-to-detect changes. A key advantage of SWE is its avoidance of augmentations, making it especially suitable for data where even minor modifications can render samples inconsistent with their labels. This is exemplified in Section IV-A3, where SWE is used to detect precursor features for anticipatory prognoses. A statistically significant improvement in F1 score is observed when using SWE-optimized features to identify patients in the early stages of developing AF, outperforming BYOL even when avoiding aggressive augmentation. Additionally, an increase in sensitivity is noted, which is particularly valuable as early interventions can significantly alter patient outcomes. These results underscore the translational potential of SWE in anticipatory diagnostics, where subtle preclinical ECG features indicative of future diseases can be successfully identified.

The results for future AF detection demonstrate substantial improvements over prior work, such as the work by Bontinck et al. [6], which reported an F1 score of 0.0897 ($\pm$0.0270). However, the experimental setups are not directly comparable. Although the same dataset is used, the architecture evaluated in this study processes 6.4-second ECG segments instead of full 10-second recordings, and evaluation is performed using cross-validation rather than single test set bootstrapping commonly used in related work.

To enable comparison with other studies, it is important to highlight that the reported results are obtained by fine-tuning

the encoders on only a single fold (±10% of the dataset for training), using only the first ECG per patient, and starting from weights pre-trained on age prediction. This setup emphasizes the benefit of the optimization. Nonetheless, the employed encoder architecture is state-of-the-art [9], and as such, the performance of the unoptimized model provides a strong and representative baseline.

B. CONTINUITY IN FEATURES

By aligning samples, SWE encourages the auxiliary embedding to mimic the structural relationships present in the input space, thereby capturing features that evolve smoothly between the aligned samples. This is clearly illustrated in Fig. 6, where the cyclic nature of the hour hand on a 12-hour clock face is preserved. In Section IV-B, the encoder is encouraged to capture generalizable longitudinal healthcare patterns by constructing relational sets consisting of individual patient trajectories. This is evaluated through a timestamp imputation experiment, where features are used to infer missing timestamps, confirming the hypothesis that SWE enforces temporal continuity in the feature space. A statistically significant improvement in the NRMSE, as shown in Table 3, supports the conclusion that SWE effectively models longitudinal patient trajectories in the latent space using discrete ECG snapshots, handling irregular time series efficiently. Being encoder-agnostic, SWE can complement existing longitudinal modeling approaches by providing temporally coherent representations that can be integrated into downstream predictive models.

C. DEPLOYMENT

SWE's encoder-agnostic design enables the framework to further optimize models tailored to specific deployment environments without altering inference-time hardware requirements, as the added modules are only present during optimization. Unlike contrastive methods such as BYOL, SWE avoids asymmetric networks that double training parameters [24] during optimization and does not depend on computationally expensive augmentations like resampling. Instead, it leverages two lightweight transformations: a single dense layer for dimensionality reduction, followed by a whitening operation implemented using a fully differentiable Cholesky decomposition—both efficiently supported by GPU-based training frameworks. This makes SWE-optimization feasible on hardware similar to the deployment environment, enabling training in resource-constrained settings where memory is limited, such as edge devices and hospital servers.

VI. LIMITATIONS & FUTURE WORK

While the results demonstrate improved performance through inter-sample alignment, several limitations and opportunities for future research remain. Notably, optimization via SWE can only learn features from samples included in the relational sets, and unaccounted selection bias may inadvertently be introduced during this process.

Although SWE is expected to be robust under highly irregular sampling conditions, comprehensive validation across diverse sampling patterns remains future work. Furthermore, its effectiveness depends on the presence of meaningful relational structures. In settings with highly variable sequences, such structure may be harder to uncover. Future work should assess SWE under such constraints and explore strategies to enhance feature learning. A full theoretical analysis could further strengthen the framework's foundation and broaden its applicability.

Although designed to be modality-agnostic, current evaluation beyond ECG signals is limited to a synthetic example. While this supports conceptual applicability to image-like data, it does not fully validate generalizability to real-world scenarios. Broader evaluation across biomedical and non-biomedical domains, as well as across dataset scales and modalities, remains an important direction.

Finally, while the longitudinal experiment demonstrates embedded evolution, future work should benchmark the added value of SWE when integrated with established longitudinal models. This is particularly relevant for irregularly sampled medical time series, where capturing complex temporal dynamics is critical.

VII. CONCLUSION

This work introduces the Structuring Whitened Embeddings (SWE) framework, an encoder- and modality-agnostic approach that enhances feature representations by aligning inter-sample relationships in an auxiliary embedding space. By preserving proportional relations (PPR), SWE encourages encoders to capture smooth, longitudinal dynamics across samples. The experimental results across diverse tasks demonstrate that SWE-optimized features improve performance in settings where subtle progression patterns and temporal continuity are critical. In particular, it is demonstrated that SWE enhances anticipatory prognosis through the early detection of slow-onset conditions such as atrial fibrillation and improves imputation accuracy by embedding continuity into feature representations, highlighting its flexibility in modeling both inter- and intra-group relationships. The framework's architecture-agnostic and augmentation-free design, along with its robustness to irregular sampling, makes it a versatile tool for predictive tasks involving transformation-sensitive data in healthcare and beyond. Its practical applicability is further supported by the optimization relying on lightweight operations, enabling training in resource-constrained environments. Potential applications include chronic disease monitoring, healthcare trajectory modeling, and treatment response forecasting. Although further validation on external datasets and in prospective studies is needed, SWE offers a scalable and practical solution for modeling evolution in complex, real-world datasets.

APPENDIX A
NOTATION TABLE

See Table 4.

TABLE 4. Summary of notation used in Section III.

Section	Symbol	Meaning
III-A	$\mathcal{D}$	The dataset
	$\mathcal{S}$	Relational set, consisting of elements within $\mathcal{D}$
	x	Element of $\mathcal{D}$
	y	Target variable associated with x
	$E(.)$	Feature extraction network
	e	Extracted features
	$g(.)$	Projection to an auxiliary embedding
	$f(.)$	$g(E(.))$
	δ	Relational metric function in input space
	ϕ	Relational metric function in embedding space
III-B	$h(.)$	Projection to a lower-dimensional hypersphere
	h	Element in the lower-dimensional hypersphere
	$w(.)$	Whitening transformation
	w	Element in the whitened embedding
	W	Whitening matrix
	$\Sigma_{\mathbf{h}}$	Symmetric covariance matrix of a batch of embeddings $\mathbf{h}$
	L	Lower triangular matrix of $\Sigma_{\mathbf{h}}$ ($\Sigma_{\mathbf{h}} = LL^T$)
	μ	Mean
III-C	m	Number of relational sets in $\mathcal{D}$
	j	Index in the range $(1, m)$, used to indicate a relational set $\mathcal{S}_j$
	n	Number of samples per relational set
	k	Index in the range $(1, n)$, used to slice relational sets and obtain sub-batch k

APPENDIX B
TRAINING HYPERPARAMETERS

All experiments were run on a shared server with an NVIDIA GeForce RTX 2080 Ti GPU and 8 GB of VRAM available, concurrently used by multiple researchers. Networks were trained using the Adam optimizer ($\epsilon = 1 \times 10^{-6}$). The AF experiment (Section IV-A) employs binary focal loss ($\gamma = 2, \alpha = 0.25$), while the imputation task (Section IV-B) uses MSE. To account for the cyclic nature of the clock in Section IV-C, the error is computed using a sine-cosine transformation, ensuring continuity and smooth gradients. Fine-tuning epochs are determined via early stopping on the validation fold. No learning rate scheduling is applied. The ResNet-like encoder network used in Sections IV-A and IV-B contains approximately 438k trainable parameters. Additional training details are summarized in Table 5.

TABLE 5. Training hyperparameters per experiment. While the number of training steps for SWE and BYOL is pre-defined, the number of epochs for fine-tuning is determined via early stopping on the validation fold. Relying on an asymmetric network structure, the batch size for BYOL is halved to be able to run it with the same hardware requirements. Reported values for training steps and wall time represent averages across runs.

Sec.	Training	Learning Rate	Batch Size	Steps	Time
IV-A	BYOL	3×10^{-6}	128	30k	3h
	SWE	3×10^{-5}	256 (64*4)	30k	2h 30m
	Fine-tuning	1×10^{-5}	256	30k	2h 20m
IV-B	SWE	5×10^{-4}	126 (42*3)	3k	0h 45m
	Fine-tuning	1×10^{-3}	128	10k	2h 30m
IV-C	SWE	3×10^{-4}	180 (45*4)	10k	0h 20m
	Fine-tuning	3×10^{-4}	180	200k	4h

APPENDIX C
SHAPIRO-WILK TEST RESULTS

Before assessing statistical significance using the paired t-test, the normality of test results was evaluated using the Shapiro-Wilk test. Test statistics for the early onset AF prediction experiment (Section IV-A) are summarized in Table 6, while results for the imputation experiment (Section IV-B) are listed in Table 7. Results with $p > 0.05$ indicate no significant deviation from normality, supporting the use of parametric statistical methods for evaluating significance in model improvements.

TABLE 6. Shapiro-Wilk test results for normality across models in the early onset AF prediction experiment.

Exp.	Model	W Statistic	p-value
pAF	Base	0.9849	0.3863
	BYOL	0.9927	0.9066
	SWE	0.9882	0.5959
fAF	Base	0.9877	0.5663
	BYOL	0.9765	0.1030
	SWE	0.9883	0.6053

TABLE 7. Shapiro-Wilk test results for normality across models in the imputation experiment. All models with $n_s \geq 4$ yield $p > 0.05$, indicating no significant deviation from normality when sequences of four or more samples are completed.

n_s	Model	W Statistic	p-value
3	Age P.	0.9590	0.7739
	Base	0.9076	0.2648
	SWE	0.8380	0.0418
4	Age P.	0.9687	0.8772
	Base	0.8912	0.1751
	SWE	0.9203	0.3591
5	Age P.	0.9164	0.3276
	Base	0.9356	0.5048
	SWE	0.9604	0.7906
6	Age P.	0.8656	0.0887
	Base	0.9632	0.8217
	SWE	0.9394	0.5469

REFERENCES

[1] G. Iglesias, E. Talavera, Á. González-Prieto, A. Mozo, and S. Gómez-Canaval, "Data augmentation techniques in time series domain: A survey and taxonomy," *Neural Comput. Appl.*, vol. 35, no. 14, pp. 10123–10145, May 2023.

[2] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, and H. Xu, "Time series data augmentation for deep learning: A survey," 2020, *arXiv:2002.12478*.

[3] R. Liu, R. Yin, Y. Liu, and W. Wang, "Unbiased and augmentation-free self-supervised graph representation learning," *Pattern Recognit.*, vol. 149, May 2024, Art. no. 110274.

[4] G. Feng, S. Manimurugan, B. Yi, and Y. Feng, "Towards precision cardiac healthcare: Deep learning and IoT integration for real-time monitoring and personalized diagnosis," *IEEE Internet Things J.*, early access, Jun. 12, 2025, doi: 10.1109/JIOT.2025.3579122.

[5] M. Lueken, J. Mettner, N. Spicher, M. Gramlich, N. Marx, S. Leonhardt, and M. D. Zink, "Towards artificial intelligence-based decision support for large-scale screening for atrial fibrillation," *IEEE J. Biomed. Health Informat.*, early access, Jun. 13, 2025, doi: 10.1109/JBHI.2025.3579621.
[6] L. Bontinck, K. Fonteyn, T. Dhaene, and D. Deschrijver, "ECGencode: Compact and computationally efficient deep learning feature encoder for ECG signals," *Expert Syst. Appl.*, vol. 255, Dec. 2024, Art. no. 124775.
[7] F. Zhou and D. Fang, "Multimodal ECG heartbeat classification method based on a convolutional neural network embedded with FCA," *Sci. Rep.*, vol. 14, no. 1, p. 8804, Apr. 2024.
[8] J. A. Wahid, X. Mingliang, M. Ayoub, S. Husssain, L. Li, and L. Shi, "A hybrid ResNet-ViT approach to bridge the global and local features for myocardial infarction detection," *Sci. Rep.*, vol. 14, no. 1, p. 4359, Feb. 2024.
[9] A. H. Ribeiro, M. H. Ribeiro, G. M. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. S. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira, T. B. Schön, and A. L. P. Ribeiro, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nature Commun.*, vol. 11, no. 1, p. 1760, Apr. 2020.
[10] A. Sau, J. Barker, L. Pastika, E. Sieliwonczyk, K. Patlatzoglou, K. A. McGurk, N. S. Peters, D. P. O'Regan, J. S. Ware, D. B. Kramer, J. W. Waks, and F. S. Ng, "Artificial intelligence–enhanced electrocardiography for prediction of incident hypertension," *JAMA Cardiol.*, vol. 10, no. 3, pp. 214–223, 2025.
[11] H. N. Pham, L. Holmstrom, H. Chugh, A. Uy-Evanado, K. Nakamura, Z. Zhang, A. Salvucci, J. Jui, K. Reinier, and S. S. Chugh, "Dynamic electrocardiogram changes are a novel risk marker for sudden cardiac death," *Eur. Heart J.*, vol. 45, no. 10, pp. 809–819, Mar. 2024.
[12] C. Sun, S. Hong, M. Song, and H. Li, "A review of deep learning methods for irregularly sampled medical time series data," 2020, *arXiv:2010.12493*.
[13] T. Albert, B. Eskofier, and D. Zanca, "From patches to objects: Exploiting spatial reasoning for better visual representations," *Discover Appl. Sci.*, vol. 6, no. 5, p. 232, Apr. 2024.
[14] M. Garnelo and M. Shanahan, "Reconciling deep learning with symbolic artificial intelligence: Representing objects and relations," *Current Opinion Behav. Sci.*, vol. 29, pp. 17–23, Oct. 2019.
[15] K. J. Holyoak and H. Lu, "Emergence of relational reasoning," *Current Opinion Behav. Sci.*, vol. 37, pp. 118–124, Feb. 2021.
[16] M. Patacchiola and A. Storkey, "Self-supervised relational reasoning for representation learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 4003–4014.
[17] F. D. Pup and M. Atzori, "Applications of self-supervised learning to biomedical signals: A survey," *IEEE Access*, vol. 11, pp. 144180–144203, 2023.
[18] K. Zhang, Q. Wen, C. Zhang, R. Cai, M. Jin, Y. Liu, J. Y. Zhang, Y. Liang, G. Pang, D. Song, and S. Pan, "Self-supervised learning for time series analysis: Taxonomy, progress, and prospects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 10, pp. 6775–6794, Oct. 2024.
[19] E. Eldele, M. Ragab, Z. Chen, M. Wu, C.-K. Kwoh, X. Li, and C. Guan, "Self-supervised contrastive representation learning for semi-supervised time-series classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 15604–15618, Dec. 2023.
[20] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.
[21] L. Jing, P. Vincent, Y. LeCun, and Y. Tian, "Understanding dimensional collapse in contrastive self-supervised learning," 2021, *arXiv:2110.09348*.
[22] J. Gui, T. Chen, J. Zhang, Q. Cao, Z. Sun, H. Luo, and D. Tao, "A survey on self-supervised learning: Algorithms, applications, and future trends," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 9052–9071, Dec. 2024.
[23] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. Guo, M. G. Azar, and B. Piot, "Bootstrap your own latent—A new approach to self-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 21271–21284.
[24] A. Ermolov, A. Siarohin, E. Sangineto, and N. Sebe, "Whitening for self-supervised representation learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 3015–3024.
[25] J. Žbontar, J. Li, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 12310–12320.
[26] S. Zhang, L. Qiu, F. Zhu, J. Yan, H. Zhang, R. Zhao, H. Li, and X. Yang, "Align representations with base: A new approach to self-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 16579–16588.
[27] M. Papillon, S. Sanborn, J. Mathe, L. Cornelis, A. Bertics, D. Buracas, H. J. Lillemark, C. Shewmake, F. Dinc, X. Pennec, and N. Miolane, "Beyond euclid: An illustrated guide to modern machine learning with geometric, topological, and algebraic structures," *Mach. Learn., Sci. Technol.*, vol. 6, no. 3, Sep. 2025, Art. no. 031002.
[28] A. Siarohin, E. Sangineto, and N. Sebe, "Whitening and coloring batch transform for GANs," 2018, *arXiv:1806.00420*.
[29] D. Dereniowski and M. Kubale, "Cholesky factorization of matrices in parallel and ranking of graphs," in *Proc. 5th Int. Conf. Parallel Process. Appl. Mathematics*, Czestochowa, Poland, 2004, pp. 985–992.
[30] A. C. Koivunen and A. B. Kostinski, "The feasibility of data whitening to improve performance of weather radar," *J. Appl. Meteorol.*, vol. 38, no. 6, pp. 741–749, Jun. 1999.
[31] A. H. Ribeiro, G. M. Paixao, E. M. Lima, M. Horta Ribeiro, M. M. Pinto Filho, P. R. Gomes, D. M. Oliveira, W. Meira Jr., T. B. Schon, and A. L. P. Ribeiro, "CODE-15%: A large scale annotated dataset of 12-lead ECGs," Telehealth Netw. Minas Gerais (TNMG), Minas Gerais, Brazil, Tech. Rep., Jun. 2021, doi: 10.5281/zenodo.4916206.
[32] J. A. Joglar et al., "2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: A report of the American college of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines," *J. Amer. College Cardiology*, vol. 149, no. 1, pp. 109–279, 2023.
[33] T. Mehari and N. Strodthoff, "Self-supervised representation learning from 12-lead ECG data," *Comput. Biol. Med.*, vol. 141, Feb. 2022, Art. no. 105114.
[34] S. Wang, M. Khabsa, and H. Ma, "To pretrain or not to pretrain: Examining the benefits of pretraining on resource rich tasks," 2020, *arXiv:2006.08671*.
[35] M. Kazijevs and M. D. Samad, "Deep imputation of missing values in time series health data: A review with benchmarking," *J. Biomed. Informat.*, vol. 144, Aug. 2023, Art. no. 104440.
[36] C. Sun, H. Li, M. Song, D. Cai, B. Zhang, and S. Hong, "Time pattern reconstruction for classification of irregularly sampled time series," *Pattern Recognit.*, vol. 147, Mar. 2024, Art. no. 110075.

KAREL FONTEYN is currently pursuing the Ph.D. degree with the IDLab group, an affiliation of imec, Faculty of Engineering and Architecture of Ghent University, Belgium. His current research interests include deep learning, self-supervised learning, time series representation learning, and healthcare technology.

LENNERT BONTINCK (Graduate Student Member, IEEE) is currently pursuing the Ph.D. degree with the IDLab group, an affiliation of imec, Faculty of Engineering and Architecture of Ghent University, Belgium. His research interests include data-efficient machine learning, biomedical time series analysis, and patient trajectory prediction.

TOM DHAENE (Senior Member, IEEE) received the Ph.D. degree in electrotechnical engineering from Ghent University, Ghent, Belgium, in 1993. In 1993, he was with the EDA Company Alphabit (currently part of Keysight Technologies). Since 2007, he has been a Full Professor with the IDLab group, an affiliation of imec, Faculty of Engineering and Architecture of Ghent University.

DIRK DESCHRIJVER (Senior Member, IEEE) received the first Ph.D. degree in sciences from the University of Antwerp, Belgium, in 2007, and the second Ph.D. degree in engineering from Ghent University, Belgium, in 2012. Since 2014, he has been a Senior Researcher with iMinds/imec, Ghent. Since 2016, he has been an Associate Professor with Ghent University.