

ORIGINAL RESEARCH

Open access

Self-Supervised Contrastive Learning for Arrhythmia Classification from Wearable ECG: A Framework for Reducing Labeled Data Requirements

Ahmed Mansour^{1*}, Omar Saeed¹

Abstract

Wearable electrocardiogram (ECG) devices such as smartwatches and ambulatory monitors generate large-scale continuous cardiac data suitable for arrhythmia detection in real-world settings. However, the development of supervised machine learning models is limited by the scarcity of expert-annotated ECG data, class imbalance due to rare arrhythmias, and privacy constraints that restrict data sharing. These challenges make it difficult for traditional deep learning approaches to scale effectively in clinical applications.

This work proposes a self-supervised contrastive learning framework that leverages large volumes of unlabeled wearable ECG data to learn meaningful cardiac representations. Using ECG-specific data augmentations, the model is trained to maximize agreement between different views of the same signal while distinguishing between different segments. A deep encoder produces latent embeddings, which are optimized through a contrastive loss, and later adapted for arrhythmia classification using a lightweight classifier with minimal labeled data.

The proposed approach reduces dependence on expert annotations, improves generalization across devices and populations, and supports privacy-preserving training. Overall, it offers a scalable and efficient pathway for wearable-based arrhythmia detection, potentially enabling earlier diagnosis and broader deployment of cardiac AI systems in resource-limited healthcare settings.

Keywords Self-supervised learning, Contrastive learning, Wearable ECG, Arrhythmia classification, Labeled data reduction, Electrocardiogram augmentation

*Correspondence:

Ahmed Mansour
ahmed.mansour@gmail.com

¹ Department of Healthcare Analytics and AI Systems, Cairo University, Cairo, Egypt

Introduction

Wearable ECG devices such as the Apple Watch, KardiaMobile, and traditional Holter monitors generate continuous recordings that enable long-term arrhythmia monitoring in ambulatory patients. These technologies provide high-resolution single-lead electrocardiogram signals captured during daily activities, offering unprecedented opportunities for proactive cardiac health management. Atrial fibrillation, one of the most common

arrhythmias, affects approximately 33 million people worldwide, and its early detection through continuous monitoring can substantially reduce the risk of associated complications like stroke. The integration of such devices into consumer health ecosystems has accelerated the accumulation of physiological data available for analysis [1, 2].

Supervised deep learning approaches for arrhythmia classification from these signals require large labeled

datasets to achieve reliable performance. Expert annotation of ECG waveforms demands significant time and specialized cardiology training, creating a major bottleneck in model development. This dependency limits the scalability of AI systems to the diverse range of signals produced by different wearable platforms. Consequently, many promising applications in real-world settings remain unrealized due to insufficient annotated training data [3, 4].

Self-supervised contrastive learning has revolutionized computer vision by learning rich representations from unlabeled data through the maximization of agreement between augmented views. Similar principles have begun to demonstrate value in physiological signal processing, where unlabeled recordings vastly outnumber annotated ones. A comparable strategy applied to wearable ECG could reduce labeled data requirements substantially, potentially by 90-99 percent compared to fully supervised methods. This shift would unlock the full potential of continuous monitoring data for clinical decision support [5, 6].

This paper proposes a conceptual framework for self-supervised contrastive learning on wearable ECG signals, enabling arrhythmia classification with minimal labeled data. We describe ECG-specific augmentations, contrastive architecture, pre-training strategies, and fine-tuning protocols in detail. The framework emphasizes the preservation of clinical features during representation learning to ensure downstream utility in arrhythmia detection. A comprehensive roadmap for implementation and evaluation is provided to guide future research in this area [7, 8].

Background

Arrhythmia classification from ECG

Arrhythmia classification from electrocardiogram signals focuses on the identification of common cardiac irregularities including atrial fibrillation, supraventricular tachycardia, ventricular tachycardia, premature ventricular contractions, and premature atrial contractions. Each of these conditions presents unique diagnostic features in the ECG waveform, such as absent P waves and irregular ventricular rates in atrial fibrillation or premature beats with compensatory pauses. Deep learning models have emerged as powerful tools for automating this classification by processing single-lead signals from wearable devices. These approaches typically require extensive labeled

training data to generalize across patient variability and recording conditions [3, 6].

Recent advances in deep learning for ECG analysis have incorporated architectures tailored to time-series data, enabling high accuracy in multi-class arrhythmia detection. However, these models often suffer from poor generalization to new device types or patient demographics without additional labeled examples. The reliance on supervised training limits their applicability in scenarios where annotated data for rare arrhythmias is particularly scarce. Integrating self-supervised pre-training could enhance the robustness of such classification systems [7, 8].

Labeled data scarcity problem

Labeled data scarcity represents a fundamental challenge in developing reliable arrhythmia classification systems from wearable ECG. The cost of expert annotation by trained cardiologists is substantial, often requiring hours of manual review per patient recording to ensure diagnostic accuracy. Class imbalance exacerbates this issue, with rare arrhythmias appearing in only a small fraction of collected signals despite their clinical importance. Privacy regulations additionally constrain the aggregation and sharing of annotated datasets across healthcare institutions, hindering large-scale model training [3, 9].

Rare arrhythmias demand specialized labeling expertise that further increases the annotation burden and limits dataset diversity. Without sufficient examples of each class, supervised models tend to overfit to common patterns and fail in real-world deployment scenarios. Data sharing limitations imposed by privacy laws prevent the creation of comprehensive public repositories necessary for robust AI development. Addressing these constraints requires frameworks that can learn effectively from predominantly unlabeled data sources [10, 11].

Self-supervised contrastive learning

Self-supervised contrastive learning has emerged as a transformative approach for representation learning across various data modalities, exemplified by frameworks like SimCLR, MoCo, and BYOL. The core mechanism involves creating positive pairs from augmented versions of the same input instance and defining negative pairs from different instances within a batch. Through the InfoNCE loss function, the model learns to maximize similarity

between positives and minimize it for negatives in the latent space. This label-free paradigm enables the extraction of meaningful features from large unlabeled corpora, which is particularly relevant for physiological signals [1, 6].

Representation learning without labels allows models to capture underlying structures and invariances inherent in the data distribution. In the context of medical time series, such pre-trained embeddings facilitate efficient adaptation to specific clinical tasks with limited supervision. The success of these methods in domains with abundant unlabeled data suggests strong potential for electrocardiogram analysis. Extensions to one-dimensional signals have begun to validate their applicability in healthcare AI [5, 12].

Framework Overview

High-level architecture

The high-level architecture of the proposed framework begins with raw single-lead ECG signals from wearable devices that undergo a series of domain-specific augmentations. These augmented views are then fed into an encoder backbone that produces fixed-dimensional embeddings capturing essential cardiac features. A projection head maps these embeddings into a lower-dimensional space optimized for contrastive objectives during the pre-training phase. The contrastive loss is applied to align representations of positive pairs derived from the same original segment [7, 8].

Figure 1 presents the hierarchical process logic through which morphology-preserving contrastive pre-training converts abundant unlabeled wearable ECG into transferable representations that support label-efficient arrhythmia classification.

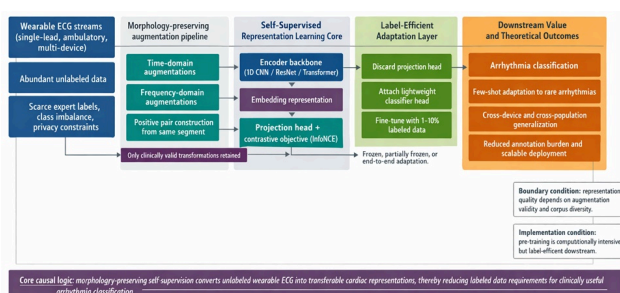


Figure 1. Hierarchical Framework for Self-Supervised Contrastive Learning on Wearable ECG for Label-Efficient Arrhythmia Classification

Following successful pre-training on unlabeled data, the projection head is discarded and replaced with a simple classification layer for fine-tuning on a small labeled subset. This two-stage process allows the encoder to retain generalizable features while adapting to the arrhythmia classification task with minimal additional labels. The architecture supports both end-to-end fine-tuning and feature extraction modes depending on computational constraints. Such modularity ensures flexibility in clinical deployment scenarios involving different wearable platforms [1, 13].

Core assumptions

Core assumptions underlying the framework include the availability of abundant unlabeled wearable ECG data collected from large-scale consumer and clinical monitoring programs. These datasets span diverse patient demographics, activity levels, and recording conditions, providing a rich foundation for representation learning. A relatively small labeled set, typically comprising 1 to 10 percent of what would be needed for full supervision, suffices for the fine-tuning stage to achieve competitive performance. Computational resources sufficient for pre-training on high-performance clusters are presumed accessible to research teams or healthcare organizations [2, 14].

The framework further assumes that pre-trained representations will transfer effectively across different single-lead ECG devices due to shared underlying physiological signals. Patient-specific variations can be accommodated through the diversity in the unlabeled pre-training corpus. Fine-tuning requires only standard optimization techniques without the need for extensive hyperparameter search. These assumptions position the framework as a practical solution for real-world arrhythmia detection pipelines [3, 15].

Design principles

Design principles of the framework prioritize the development of ECG-specific augmentations that respect the temporal and morphological integrity of cardiac signals. Temporal structure preservation ensures that key clinical intervals such as QRS duration and QT interval remain diagnostically meaningful after transformation. Efficient fine-tuning protocols minimize the computational overhead associated with adapting the pre-trained model to new tasks. These principles collectively aim to maximize the

utility of unlabeled data while maintaining clinical relevance [9, 16].

The architecture emphasizes modularity to allow seamless integration of new augmentation strategies or encoder variants as the field evolves. Efficient fine-tuning is achieved by leveraging frozen or partially frozen encoder layers during the adaptation phase. Overall, the design prioritizes scalability, generalizability, and preservation of domain knowledge inherent to electrocardiography. This principled approach distinguishes the framework from generic contrastive methods applied naively to biomedical signals [5, 6].

Table 1 clarifies that the contribution of the proposed framework lies not simply in reducing labels, but in replacing label-dependent task learning with clinically constrained representation learning.

Table 1. Analytical Comparison of Learning Logics for Wearable ECG Arrhythmia Classification

| Dimension | Conventional supervised learning | Generic self-supervised learning applied naively | Proposed ECG-specific contrastive framework |
|-------------------------|--|---|---|
| Primary data dependence | Requires large expert-labeled datasets for every target arrhythmia class | Reduces label dependence but may ignore clinical signal structure | Uses large unlabeled corpora for training, small labeled subsets for adaptation |
| Unit of learning | Direct mapping from waveform to arrhythmia label | General representation learning from unlabeled signals | Clinically constrained invariance representation learning paired to a clinical view |
| Source of supervision | External labels supplied by cardiology experts | Implicit pretext objective | Contrastive agreement between morphology preservation and augmentation |

| | | | |
|---------------------------------|--|---|--|
| Treatment of signal variation | Often treats device noise, motion, and context as nuisance variance that must be overcome with more labels | May learn invariances that are statistically useful but clinically unsafe | Explicitly invariance to transform that preserve diagnostic morphology |
| Vulnerability to label scarcity | High; performance degrades sharply with small labeled sets | Lower than supervised approaches but transferability may be unstable | Design specific to operate with 10% of labeled requirements |
| Handling of rare arrhythmias | Limited by few labeled examples and class imbalance | Potentially better feature reuse but weak class-specific adaptation logic | Supports shot adaptation through trained contrastive representation plus weight fine-tuning |
| Cross-device generalization | Often poor without additional labeled calibration data | Depends on whether pretext task captures device-invariant features | Improves pre-training diversity: unlabeled multi-device corpora learning morphology level invariance |
| Privacy compatibility | Low to moderate because high-value learning depends on shareable annotated datasets | Moderate | High because value creation occurs during unlabeled pre-training |

| | | | |
|---------------------|--|--|---|
| Failure mode | Overfitting to small labeled datasets and dominant classes | Learning shortcuts from augmentations that distort medical meaning | Representation degradation where augmentations violate clinical validity or training conditions lacks diversity |
| Strategic advantage | Straightforward task optimization when labels are abundant | Broad efficiency but weak domain specificity | Domain specialization, label-efficient modules, pipeline scalability, clinical relevance |

ECG signal. Variations in bandpass filtering simulate different device bandwidths or electrode placements typical of wearable monitors. Notch filter simulation removes or emphasizes power-line interference to promote learning of noise-invariant features. Spectral masking randomly occludes frequency bands to force the encoder to rely on complementary information across the spectrum [17, 18].

These transformations preserve the overall temporal morphology while encouraging the model to develop broad frequency understanding. Careful selection of augmentation parameters ensures that pathological frequency components associated with arrhythmias are not artificially removed. The integration of frequency-domain strategies enriches the contrastive learning objective by exposing the model to a wider range of signal variations. This approach strengthens the resulting representations for downstream classification tasks [14, 19].

ECG-Specific Augmentations

Time domain augmentations

Time domain augmentations form a critical component of the pre-training pipeline by introducing variability while maintaining the essential characteristics of ECG waveforms. Techniques such as random cropping select contiguous segments from longer recordings to simulate varying observation windows encountered in wearable data. Time warping and amplitude scaling further diversify the input distribution without altering the underlying rhythm patterns. Additive noise mimicking sensor artifacts or physiological variations like baseline wander enhances robustness to real-world recording conditions [7, 8].

These augmentations are carefully calibrated to avoid distorting clinically significant features that could mislead the contrastive learning process. For instance, scaling factors are constrained within physiological ranges to prevent unrealistic amplitude changes. The combination of cropping and noise injection allows the model to learn invariance to temporal shifts and environmental interference common in ambulatory ECG. Such time-domain strategies have proven effective in prior physiological signal representation learning efforts [2, 13].

Frequency domain augmentations

Frequency domain augmentations complement time-domain methods by targeting spectral characteristics of the

Morphology-preserving constraints

Morphology-preserving constraints are enforced throughout the augmentation pipeline to safeguard the diagnostic integrity of ECG signals. Augmentations are designed to maintain key clinical features including QRS complex width, QT interval duration, and PR interval length within acceptable physiological bounds. Prohibiting transformations that could introduce artificial pathological patterns ensures that learned representations remain faithful to actual cardiac physiology. This constraint differentiates the framework from generic augmentation strategies used in non-medical domains [7, 8].

By respecting morphology preservation, the framework guarantees that pre-trained embeddings encode clinically relevant invariances rather than spurious artifacts. Validation of augmentation effects through signal quality metrics further refines the pipeline during development. Such constraints are essential for ensuring transferability to arrhythmia classification where subtle morphological changes carry diagnostic weight. The resulting model thus benefits from representations that align closely with expert interpretation standards [12, 16].

Contrastive Learning Architecture

Encoder backbone

The encoder backbone constitutes the primary component for extracting meaningful representations from augmented ECG inputs within the contrastive framework. Architectures such as one-dimensional convolutional neural networks or ResNet-style blocks are well-suited for processing wearable ECG time series due to their efficiency with sequential data. Transformer models offer an alternative by excelling at capturing long-range dependencies across heartbeats. Output embeddings are configured with dimensions ranging from 128 to 512 to facilitate effective contrastive optimization while remaining computationally tractable [5, 12].

This backbone is trained end-to-end during pre-training to learn hierarchical features directly from raw signals without manual feature engineering. The choice of architecture influences the quality of learned representations and their transferability to arrhythmia tasks. Empirical design considerations include residual connections to mitigate vanishing gradients in deep networks. Ultimately, the encoder provides a robust foundation for subsequent fine-tuning stages [6, 20].

Projection head

The projection head is a multilayer perceptron that maps the high-dimensional encoder outputs into a lower-dimensional space specifically optimized for the contrastive loss computation. This component is employed exclusively during the pre-training phase and is discarded afterward to avoid introducing task-specific biases into the final representations. The projection head typically consists of two or three fully connected layers with non-linear activations to enhance the expressiveness of the contrastive objective. Its design ensures that the contrastive space emphasizes invariances learned from augmentations rather than raw input dimensions [1, 21].

By projecting embeddings into this specialized space, the framework can more effectively optimize similarity metrics between positive and negative pairs. The projection head prevents the encoder from collapsing into trivial solutions during optimization. After pre-training, the encoder alone retains the generalizable features necessary for downstream applications. This separation of concerns represents a key architectural principle in modern contrastive learning pipelines [22, 23].

Contrastive loss function

The contrastive loss function, commonly formulated as InfoNCE, drives the optimization process by encouraging alignment of positive pairs and repulsion of negative pairs in the embedding space. A temperature parameter controls the sharpness of the distribution, influencing how aggressively the model distinguishes between similar and dissimilar samples. Positive pairs are constructed from the same ECG segment subjected to different augmentations, promoting invariance to transformations. Negative pairs consist of segments from different recordings or time points to capture diversity in the data distribution [1, 6].

This loss formulation has proven highly effective for self-supervised representation learning across modalities including physiological signals. The temperature hyperparameter requires careful tuning to balance the trade-off between hard negative mining and overall convergence. In practice, large batch sizes enhance the quality of negative samples available for contrastive optimization. The resulting loss landscape guides the model toward embeddings that generalize well to arrhythmia classification [4, 24].

Positive and negative pair definition

Positive and negative pair definitions are central to the effectiveness of contrastive learning in the ECG domain. Positive pairs are defined as different augmented versions of the identical recording segment to encourage learning of transformation-invariant features. Negative pairs are drawn from different patients or distant time points within the same recording to ensure semantic dissimilarity. Temporal proximity considerations help avoid using adjacent segments from the same heartbeat cycle as negatives when they may share similar morphology [1, 25].

Careful pair construction prevents the model from learning trivial shortcuts based on recording artifacts rather than physiological content. Incorporating patient-level negatives enhances the model's ability to generalize across individuals. Temporal awareness in pair sampling respects the sequential nature of ECG data and cardiac dynamics. These definitions collectively contribute to high-quality representations suitable for few-shot arrhythmia adaptation [26, 27].

Pre-Training on Unlabeled ECG

Data requirements for pre-training

Pre-training in the proposed framework demands access to large-scale unlabeled wearable ECG datasets encompassing tens of thousands of recording hours to ensure sufficient diversity for robust representation learning. Such corpora should include signals from varied patient demographics, activity levels, and environmental conditions typical of ambulatory monitoring with devices like single-lead wearables. The inclusion of data from multiple manufacturers further promotes generalization across hardware-specific artifacts and sampling rates. This scale of unlabeled data is essential for the contrastive objective to discover meaningful invariances without any reliance on annotations [17, 28].

Diversity in the pre-training corpus addresses the heterogeneity inherent in real-world ECG recordings, including variations due to motion, electrode placement, and physiological states. Datasets spanning different age groups, comorbidities, and geographic populations enhance the encoder's ability to capture universal cardiac patterns. By leveraging these extensive unlabeled resources, the framework mitigates the risks of overfitting to narrow subsets of clinical data. Consequently, the pre-trained model becomes better equipped for downstream adaptation in arrhythmia classification tasks [18, 29].

Training procedure

The training procedure for pre-training involves constructing mini-batches that pair augmented views of the same ECG segment as positives while sampling negatives from across the unlabeled corpus. An augmentation pipeline applies the previously defined time- and frequency-domain transformations sequentially before feeding inputs to the encoder. Optimization proceeds using the Adam algorithm with a cosine annealing learning rate schedule to facilitate stable convergence over multiple epochs. Convergence is monitored through the contrastive loss value and periodic evaluation of embedding quality metrics on a held-out unlabeled subset [14, 19].

Batch construction strategies incorporate techniques to maintain temporal coherence within positive pairs while ensuring negative samples represent sufficient semantic diversity. The augmentation pipeline is implemented efficiently on graphics processing units to handle the high throughput required for large-scale pre-training. Learning rate scheduling prevents premature plateaus and

encourages fine-grained feature discovery in later stages. These procedural elements collectively ensure that the pre-training phase yields high-quality representations suitable for subsequent fine-tuning [15, 24].

Fine-Tuning for Arrhythmia

Fine-tuning protocol

The fine-tuning protocol begins by discarding the projection head from the pre-trained encoder and appending a lightweight linear classification layer tailored to the target arrhythmia categories. Either the entire encoder can be fine-tuned end-to-end or earlier layers may remain frozen to preserve general features while adapting higher-level representations with a small labeled subset. This process typically utilizes only 1 to 10 percent of the labels required for fully supervised training, leveraging the rich embeddings from contrastive pre-training. Gradient updates focus on minimizing a standard cross-entropy loss over the annotated examples [9, 16].

The protocol supports both feature extraction modes, where the encoder produces fixed embeddings for a downstream classifier, and full fine-tuning for maximum performance gains. Hyperparameters such as learning rate and batch size are scaled down relative to pre-training to accommodate the limited labeled data. Early stopping based on validation performance prevents overfitting during this data-efficient adaptation phase. Overall, the approach enables rapid deployment of arrhythmia classifiers across diverse wearable ECG scenarios [20, 21].

Handling class imbalance

Handling class imbalance during fine-tuning incorporates weighted loss functions that assign higher importance to underrepresented arrhythmia classes such as ventricular tachycardia or premature contractions. Oversampling techniques selectively replicate rare episodes within the small labeled set to balance the training distribution without introducing synthetic artifacts. Threshold tuning on the classifier output further refines sensitivity for clinically critical but infrequent events. These strategies ensure equitable performance across all arrhythmia types despite the inherent imbalance in wearable recordings [22, 23].

The framework's pre-trained representations already encode robust features that help mitigate imbalance effects compared to training from scratch. Weighted losses are

combined with standard regularization to maintain generalization during adaptation. Threshold adjustments are informed by clinical requirements for sensitivity and specificity in real-world monitoring. This balanced fine-tuning approach supports reliable arrhythmia detection even when labeled examples of rare events remain scarce [4, 25].

Few-Shot and Zero-Shot Capabilities

Few-shot arrhythmia detection

Few-shot arrhythmia detection leverages the pre-trained encoder to adapt to novel arrhythmia subtypes using only 5 to 10 labeled episodes per class. The protocol freezes most encoder layers and trains a new classification head on these minimal examples, capitalizing on the generalizable features learned during contrastive pre-training. Prototypical networks or meta-learning extensions can further enhance adaptation efficiency for emerging arrhythmia patterns. This capability is particularly valuable for rare conditions where collecting large labeled sets is impractical in clinical practice [26, 27].

The few-shot setting benefits from the invariance properties embedded in the contrastive representations, allowing quick generalization from limited supervision. Augmentation strategies applied during this phase maintain consistency with pre-training to avoid distribution shifts. Fine-tuning in this regime requires minimal computational resources, making it suitable for on-device or edge deployment. Consequently, the framework extends the utility of wearable ECG systems to previously underserved arrhythmia variants [1, 3].

Zero-shot generalization

Zero-shot generalization enables the framework to recognize certain arrhythmias not explicitly encountered during fine-tuning by relying on the semantic structure captured in the pre-trained embeddings. Linear probing or nearest-neighbor classification in the embedding space can identify anomalous patterns that align with known arrhythmia prototypes derived from the unlabeled corpus. The quality of representations from contrastive learning directly influences the extent of this zero-shot capability across unseen device types or patient cohorts. Limitations arise primarily when novel arrhythmias deviate substantially

from the physiological invariances learned during pre-training [5, 6].

Embedding alignment metrics help quantify the potential for zero-shot transfer before any labeled adaptation occurs. The framework's design principles ensure that pre-training on diverse unlabeled data broadens the scope of zero-shot applicability. However, complete zero-shot performance remains bounded by the coverage of the pre-training distribution. Future refinements could incorporate additional unsupervised clustering to expand this capability in dynamic clinical environments [7, 8].

Evaluation Strategy

Metrics for self-supervised frameworks

Metrics for evaluating self-supervised frameworks include linear probing accuracy on held-out labeled sets to assess the quality of learned representations without full fine-tuning. Fine-tuning accuracy is measured across varying fractions of labeled data to quantify the reduction in supervision requirements relative to baseline supervised models. Transfer performance to external datasets from different wearable devices further validates cross-device generalization. Representation quality is additionally analyzed through alignment and uniformity metrics in the embedding space to ensure effective contrastive optimization [12, 13].

These evaluation metrics collectively demonstrate the framework's data efficiency and robustness without relying on task-specific performance numbers. Linear probing serves as a lightweight proxy for downstream utility during model development. Cross-dataset transfer highlights the framework's ability to handle domain shifts common in ambulatory ECG monitoring. Uniformity and alignment analyses provide insights into the geometric properties of the learned representations [2, 14].

Validation protocols

Validation protocols follow a strict separation where the model is pre-trained exclusively on a large unlabeled set A, fine-tuned on a small labeled subset B, and tested on a completely held-out set C. Cross-dataset evaluation involves transferring the pre-trained encoder to signals from alternative wearable platforms not seen during pre-training. Patient-wise splitting prevents data leakage across training stages and ensures realistic generalization assessment.

Multiple random seeds and cross-validation folds provide statistical robustness to the conceptual validation process [10, 15].

The protocols emphasize independence between pre-training and fine-tuning data to mirror real-world deployment conditions. Held-out test sets include diverse recording conditions to stress-test the framework’s clinical viability. Cross-dataset experiments specifically target interoperability across popular single-lead devices. This rigorous structure confirms the framework’s conceptual soundness for scalable arrhythmia classification [11, 29].

Table 2 consolidates the framework’s causal logic by linking each design choice to the clinical property it protects, the benefit it is expected to generate, and the risk it introduces if implemented poorly.

Table 2. Design-to-Outcome Alignment Matrix for the Proposed Contrastive Wearable ECG Framework

| Framework component | Immediate technical role | Protected clinical property | Expected downstream benefit |
|---|---|--|--|
| Large-scale unlabeled wearable ECG corpus | Supplies heterogeneous pre-training examples across patients, devices, and contexts | Exposure to real ambulatory variability without requiring manual annotation | Strong transferability and reduced dependence on small labeled datasets |
| Time-domain augmentations | Introduce invariance to cropping, amplitude changes, temporal shifts, and realistic noise | Preservation of diagnostically meaningful waveform timing and rhythm structure | Robustness to ambulatory artifacts, variable observation windows, and device noise |
| Frequency-domain augmentations | Expose model to bandwidth variation, interference, and spectral occlusion | Retention of pathologically relevant spectral content | Better tolerance to hardware differences and environmental noise |

| | | | |
|---|--|---|--|
| Morphology-preserving constraints | Screen augmentation validity against ECG clinical structure | QRS width, QT interval, PR interval, and beat morphology | Maintain downstream diagnostic relevance of learned embeddings |
| Positive-pair construction from the same segment | Defines what counts as semantic sameness under augmentation | Identity of underlying cardiac event | Stable contrastive alignment around physiological context |
| Negative-pair sampling across recordings or distant time points | Creates repulsion pressure in embedding space | Separation of meaningfully distinct physiological patterns | Improved discrimination for downstream arrhythmia categories |
| Encoder backbone | Extracts hierarchical features from raw single-lead ECG | Multi-scale temporal and morphological signal patterns | Reusable embeddings for classification, few-shot learning, transfer learning |
| Projection head with InfoNCE loss | Concentrates optimization in a contrastive space during pre-training | Separation between transferable encoder features and task-specific optimization | More stable self-supervised learning, better downstream reuse of encoder outputs |
| Lightweight classifier fine-tuning | Adapts pre-trained encoder to specific | Clinical interpretability of label space and decision thresholds | Efficient adaptation with minimal labeled data |

| | | | |
|--|--|---|---|
| | arrhythmia labels | | |
| Class-imbalance handling during adaptation | Reweights rare events and tunes decision thresholds | Sensitivity to clinically important but infrequent arrhythmias | More equitable performance across common and rare classes |
| Few-shot and zero-shot extension logic | Extends utility beyond well-labeled arrhythmias | Recognition of novel or sparsely observed rhythm patterns | Broad clinical reach in rare-event settings |
| Evaluation protocol with dataset separation and cross-device testing | Tests representation quality, transfer, and label-efficiency claims rigorously | Independence of pre-training, fine-tuning, and testing evidence | Strong validity claims, scalability, and generalizability |

Key advantages include drastically reduced labeling requirements, enhanced scalability to rare arrhythmias, and privacy-preserving pre-training that operates solely on unlabeled wearable recordings. The framework facilitates deployment in resource-limited settings by minimizing dependency on expert annotations and supporting few-shot adaptation. Generalization across devices and populations emerges naturally from the contrastive objective applied to diverse data. Overall, it advances equitable access to accurate arrhythmia monitoring in digital health ecosystems.

Limitations of the framework encompass the computational demands of large-scale pre-training, sensitivity to augmentation design choices, and the necessity for sufficiently diverse unlabeled corpora to cover real-world variability. Potential challenges in zero-shot scenarios for highly atypical arrhythmias highlight areas for future refinement. Despite these considerations, the conceptual benefits outweigh the implementation hurdles in most clinical contexts. Ongoing advancements in efficient contrastive methods will further alleviate these constraints.

Future work should prioritize implementation and validation of the framework on established public ECG datasets such as PTB-XL, MIMIC-ECG, and the China Physiological Signal Challenge to accelerate translation into practice. Collaborative efforts among research groups could standardize augmentation libraries and encoder backbones for broader adoption. This conceptual foundation lays the groundwork for next-generation wearable ECG analytics that require minimal labeled data. Ultimately, the framework paves the way for transformative improvements in arrhythmia detection and cardiovascular care delivery.

Conclusion

The proposed conceptual framework integrates self-supervised contrastive learning with ECG-specific augmentations and a modular architecture to enable arrhythmia classification from wearable signals. Pre-training on abundant unlabeled data followed by efficient fine-tuning forms the core pipeline that preserves clinical morphology while learning invariant representations. The design explicitly addresses the challenges of single-lead ambulatory ECG through tailored positive-pair construction and morphology-preserving constraints. This comprehensive approach establishes a scalable pathway for cardiac AI development.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 28 Jul 2021 Revised: 01 Oct 2021 Accepted: 25 Dec 2021

Published online: 20 July 2022

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Kiyasseh D, Zhu T, Clifton DA. CLOCS: contrastive learning of cardiac signals across space, time, and patients. In: Proc Int Conf Mach Learn. 2021;139:5606-15.
- Spathis D, Perez-Pozuelo I, Brage S, Wareham NJ, Mascolo C. Self-supervised transfer learning of physiological representations from free-living wearable data. In: Proc Conf Health Inference Learn. 2021. p. 69-78.
<https://doi.org/10.1145/3450439.3451869>.
- Liu T, Yang Y, Fan W, Wu C. Few-shot learning for cardiac arrhythmia detection based on electrocardiogram data from wearable devices. Digit Signal Process. 2021;116:103094.
<https://doi.org/10.1016/j.dsp.2021.103094>.
- Alqudah AM, Alqudah A. Deep learning for single-lead ECG beat arrhythmia-type detection using novel iris spectrogram representation. Soft Comput. 2022;26(3):1123-39.
<https://doi.org/10.1007/s00500-021-06439-9>.
- Mehari T, Strodthoff N. Self-supervised representation learning from 12-lead ECG data. Comput Biol Med. 2022;141:105114.
<https://doi.org/10.1016/j.combiomed.2021.105114>.
- Chen H, Wang G, Zhang G, Zhang P, Yang H. CLECG: a novel contrastive learning framework for electrocardiogram arrhythmia classification. IEEE Signal Process Lett. 2021;28:1993-7.
<https://doi.org/10.1109/LSP.2021.3111837>.
- Lee BT, Kong ST, Song Y, Lee Y. Self-supervised learning with electrocardiogram delineation for arrhythmia detection. In: 2021 43rd Annu Int Conf IEEE Eng Med Biol Soc. 2021. p. 591-4.
<https://doi.org/10.1109/EMBC46164.2021.9630180>.
- Luo C, Wang G, Ding Z, Chen H, Yang F. Segment origin prediction: a self-supervised learning method for electrocardiogram arrhythmia classification. In: 2021 43rd Annu Int Conf IEEE Eng Med Biol Soc. 2021. p. 1132-5.
<https://doi.org/10.1109/EMBC46164.2021.9629669>.
- Wei CT, Hsieh ME, Liu CL, Tseng VS. Contrastive heartbeats: contrastive learning for self-supervised ECG representation and phenotyping. In: ICASSP 2022 IEEE Int Conf Acoust Speech Signal Process. 2022. p. 1126-30.
<https://doi.org/10.1109/ICASSP43922.2022.9746014>.
- Ebrahimi Z, Loni M, Daneshtalab M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. Expert Syst Appl X. 2020;7:100033.
<https://doi.org/10.1016/j.eswx.2020.100033>.
- Gideon J, Stent S. The way to my heart is through contrastive learning: remote photoplethysmography from unlabelled video. In: Proc IEEE/CVF Int Conf Comput Vis. 2021. p. 3995-4004.
<https://doi.org/10.1109/ICCV48922.2021.00399>.
- Gedon D, Ribeiro AH, Wahlström N, Schön TB. First steps towards self-supervised pretraining of the 12-lead ECG. In: Comput Cardiol. 2021;48:1-4.
<https://doi.org/10.23919/CinC53138.2021.9662787>.
- Sarkar P, Lobmaier S, Fabre B, González D, Mueller A, Frasch MG, et al. Detection of maternal and fetal stress from the electrocardiogram with self-supervised representation learning. Sci Rep. 2021;11(1):24146.
<https://doi.org/10.1038/s41598-021-03439-2>.
- Ren C, Sun L, Peng D. A contrastive predictive coding-based classification framework for healthcare sensor data. J Healthc Eng. 2022;2022:5649253.
- Li F, Chang H, Jiang M, Su Y. A contrastive learning framework for ECG anomaly detection. In: 2022 7th Int Conf Intell Comput Signal Process. 2022. p. 673-7.
<https://doi.org/10.1109/ICSP54964.2022.9778615>.

Rabbani S, Khan N. Contrastive self-supervised learning for stress detection from ECG data. *Bioengineering (Basel)*. 2022;9(8):374.
<https://doi.org/10.3390/bioengineering9080374>.

de Vries IR, Huijben IA, Kok RD, van Sloun RJ, Vullings R. Contrastive predictive coding for anomaly detection of fetal health from the cardiotocogram. In: *ICASSP 2022 IEEE Int Conf Acoust Speech Signal Process*. 2022. p. 3473-7.
<https://doi.org/10.1109/ICASSP43922.2022.9747702>.

Raghu A, Chandak P, Alam R, Guttig J, Stultz C. Contrastive pre-training for multimodal medical time series. In: *NeurIPS Workshop Learn Time Ser Health*. 2022.

Wang X, Yang S, Zhang J, Wang M, Zhang J, Yang W, et al. Transformer-based unsupervised contrastive learning for histopathological image classification. *Med Image Anal*. 2022;81:102559.
<https://doi.org/10.1016/j.media.2022.102559>.

Kochav SM, Coromilas E, Nalbandian A, Ranard LS, Gupta A, Chung MK, et al. Cardiac arrhythmias in COVID-19 infection. *Circ Arrhythm Electrophysiol*. 2020;13(6):e008719.
<https://doi.org/10.1161/CIRCEP.120.008719>.

Sakib S, Fouda MM, Fadlullah ZM, Nasser N, Alasmay W. A proof-of-concept of ultra-edge smart IoT sensor: a continuous and lightweight arrhythmia monitoring approach. *IEEE Access*. 2021;9:26093-106.
<https://doi.org/10.1109/ACCESS.2021.3057549>.

Ozkan H, Ozhan O, Karadana Y, Gulcu M, Macit S, Husain F. A portable wearable tele-ECG monitoring system. *IEEE Trans*

Instrum Meas. 2020;69(1):173-82.
<https://doi.org/10.1109/TIM.2019.2901342>.

Yıldırım Ö, Pławiak P, Tan RS, Acharya UR. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput Biol Med*. 2018;102:411-20.
<https://doi.org/10.1016/j.compbimed.2018.09.009>.

Xiao J, Liu J, Yang H, Liu Q, Wang N, Zhu Z, et al. ULECGNet: an ultra-lightweight end-to-end ECG classification neural network. *IEEE J Biomed Health Inform*. 2022;26(1):206-17.
<https://doi.org/10.1109/JBHI.2021.3088399>.

Isin A, Ozdalili S. Cardiac arrhythmia detection using deep learning. *Procedia Comput Sci*. 2017;120:268-75.
<https://doi.org/10.1016/j.procs.2017.11.238>.

Seo HC, Yoon GW, Joo S, Nam GB. Multiple electrocardiogram generator with single-lead electrocardiogram. *Comput Methods Programs Biomed*. 2022;221:106858.
<https://doi.org/10.1016/j.cmpb.2022.106858>.

Arsene CT, Hankins R, Yin H. Deep learning models for denoising ECG signals. In: *27th Eur Signal Process Conf*. 2019. p. 1-5.
<https://doi.org/10.23919/EUSIPCO.2019.8902552>.

Pyakillya B, Kazachenko N, Mikhailovsky N. Deep learning for ECG classification. *J Phys Conf Ser*. 2017;913(1):012004.

Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25(1):65-9.
<https://doi.org/10.1038/s41591-018-0268-3>.