Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Special Track on AI and Health

PhysioGMC: Generalizable Multi-modal Coordination for Physiological Signals

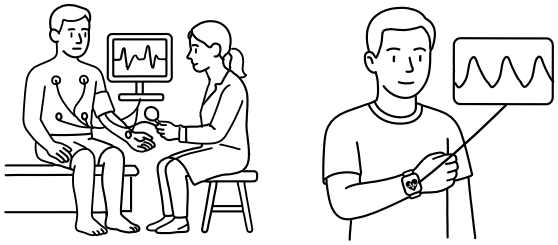
Mudi Zhang, **Anirudh Nakra** and **Min Wu**
University of Maryland, College Park
{mzhang99, anakra, minwu}@umd.edu

Abstract

Physiological signals are widely used for health assessment in clinical and daily-life settings. Established physiological signals collected for inpatient and clinical use are often impractical for patients' daily home use due to their complexity and resource demands. In contrast, wearable signals enable continuous monitoring in everyday life, but many have limited reliability and are not widely understood or accepted in clinical practices. To leverage complementary strengths of clinical and wearable physiological signals, we propose **PhysioGMC**, a **G**eneralizable **M**ulti-modal **C**oordination framework for **Physio**logical signals that explicitly accounts for their strong inter-subject variability. PhysioGMC incorporates both clinical and wearable modalities into the training process to improve cross-subject performance when only a single wearable modality is available at deployment. The framework introduces a cross-modal contrastive learning module comprising two contrastive losses to jointly learn label-relevant, subject-agnostic representations across modalities. The self-supervised contrastive loss aligns latent features across modalities, while the supervised contrastive loss encourages learning label-discriminative features that are invariant to subject identity. Experiments on cardiovascular health monitoring and sleep staging tasks demonstrate that PhysioGMC consistently outperforms existing methods, achieving superior cross-subject performance at test time using only wearable modalities, such as photoplethysmography (PPG).

1 Introduction[1]

Continuous monitoring of physiological signals plays an essential role in assessing human health in both clinical and daily-life settings. In clinical practice, signals such as electrocardiography (ECG), arterial blood pressure (ABP), and polysomnography (PSG) are routinely collected for tasks including arrhythmia detection, hypertension management, and sleep staging. With the widespread adoption of wearable devices such as smartwatches, wearable signals, such as PPG [Hwang *et al.*, 2021], have emerged as scalable, non-invasive alternatives for continuous health monitoring in everyday life. Together, these developments have led to the rapid growth in the amount of physiological data collected across diverse environments.



(a) Clinical Settings (b) Daily-life Settings

Figure 1: Illustration of physiological signal acquisition in (a) clinical settings using ECG and blood pressure measurements, and (b) daily-life settings using continuous PPG monitoring with wearables.

The growing volume of physiological data collected both within and beyond clinical settings has made manual expert review increasingly impractical, motivating deep-learning-based automated analysis systems. Recent advances have achieved strong performance in health monitoring tasks, including atrial fibrillation (AF) detection [Torres-Soto and Ashley, 2020; Das *et al.*, 2022], hypertension assessment [Leitner *et al.*, 2022; BN *et al.*, 2024], sleep apnea detection [Olsen *et al.*, 2024; Sharan *et al.*, 2020], and sleep staging [Kotzen *et al.*, 2023]. Despite these successes, most existing methods rely on a single physiological modality, thereby overlooking the complementary information available across multiple physiological signals [Gil *et al.*, 2010; Mukkamala and Hahn, 2018]. This limitation is particularly important because physiological modalities differ substantially in their reliability, accessibility, and suitability for long-term deployment.

Physiological signals collected in clinical settings, such as ECG, ABP, and PSG, provide more reliable, accurate, and clinically interpretable measurements of human health compared to signals measured by wearable devices. However, measuring these signals requires specialized and cumbersome

[1]Code and appendix: https://github.com/nott1999/PhysioGMC

equipment, making long-term continuous monitoring in non-clinical environments impractical, uncomfortable, and costly, as illustrated in Fig. 1(a). In contrast, wearable devices enable more feasible and user-friendly long-term health monitoring through non-invasive measurements, as shown in Fig. 1(b). This shows a trade-off between the fidelity of clinical-grade signals and the practicality of continuous real-world monitoring. To balance this trade-off, it is desirable to complement the strengths of multimodal physiological modalities by leveraging their inherent relationships to enhance the utility of wearable signals in real-world application scenarios.

Although training with multimodal data has been shown to improve the performance of unimodal models [Zadeh *et al.*, 2020; Ma *et al.*, 2021; Liu *et al.*, 2023; Wu *et al.*, 2024], this paradigm is understudied in the context of physiological sensing. Existing approaches in this domain generally fall into two categories: 1) inferring the absent modalities during testing from the available ones; 2) aligning features across modalities to achieve multimodal coordination. Inference-based methods [Tian *et al.*, 2023; Vo *et al.*, 2024; Ji and Zhou, 2024] propose reconstructing absent physiological modalities from available modalities using generative models and subsequently applying the inferred signals to downstream tasks. However, such approaches are sensitive to reconstruction errors and may introduce additional noise that degrades performance. Recently, some work has focused on achieving multimodal coordination for physiological signals. SiamAF [Guo *et al.*, 2023] employs a Siamese network to learn shared features between ECG and PPG signals, thus improving AF detection accuracy using only PPG at inference. But these existing methods overlook the substantial cross-subject differences inherent in physiological signals, hindering their generalization to unseen subjects.

In this work, we propose PhysioGMC, a multimodal coordination framework designed for physiological signals, which coordinates wearable modalities with clinical-grade modalities during training to improve cross-subject performance of a single wearable modality at test time. During training, PhysioGMC employs a cross-modal contrastive learning module composed of two contrastive losses to align subject-agnostic, label-decisive features across clinical-grade and wearable-grade physiological modalities. The self-supervised cross-modal contrastive loss encourages consistency across features learned from different modalities of the same input data, while the supervised cross-modal contrastive loss learns label-decisive features agnostic to subjects. At test time, PhysioGMC achieves superior performance using only the wearable PPG signal, thereby bridging the gap between the benefits of multimodal learning and the practical constraints of unimodal wearable deployment.

To the best of our knowledge, **we introduce the first generalizable multimodal coordination framework for physiological signals.** Extensive experiments on multiple public datasets demonstrate that our method consistently outperforms existing approaches across several health monitoring tasks, achieving superior cross-subject performance using only a wearable physiological signal, PPG, during inference.

2 Related Work

2.1 Inter-subject Variability in Physio. Signals

Physiological signals exhibit substantial inter-subject variability, which limits the generalization of models trained on a fixed set of subjects to unseen individuals. Accordingly, recent studies have explored subject-invariant representation learning for physiological signals. For example, some works [Hwang *et al.*, 2021; Han *et al.*, 2024] adopt adversarial training to disentangle task-related, subject-invariant features from subject-specific features and noises, thereby mitigating inter-subject variability. Another work [Kim *et al.*, 2024] leverages both deep metric learning and contrastive learning strategies to learn subject-invariant features. While these approaches improve inter-subject generalization, they typically consider each physiological signal in isolation and therefore do not exploit complementary information across modalities.

2.2 Multimodal Coordination for Physio. Signals

Previous studies [Zadeh *et al.*, 2020; Rahate *et al.*, 2022] have shown that models trained collaboratively using multimodal data, which is referred to as multimodal co-learning, can outperform those trained solely on unimodal data in downstream unimodal tasks. Motivated by these works, researchers explore multimodal co-learning for physiological signals. MERL [Liu *et al.*, 2024] proposes a cross-modal contrastive learning strategy to align ECG features with text features during the pretraining of an ECG-based encoder. SleepSMC [Ma *et al.*, 2025] improves sleep staging performance when only one modality is present at the testing stage by introducing modality-level contrastive coordination mechanisms. However, existing multimodal co-learning approaches largely overlook the inter-subject variability intrinsic to physiological signals, which limits their effectiveness and motivates our generalizable multimodal coordination framework, PhysioGMC.

2.3 Multimodal Domain Generalization (DG)

Most existing research for DG focuses on learning domain-invariant features for unimodal data. RNA-Net [Planamente *et al.*, 2022] is among the first methods to learn generalized multimodal features across domains. SimMMDG [Dong *et al.*, 2023] aims to achieve domain generalization in multimodal settings by splitting the representations into modality-specific and modality-shared components. CMRF [Fan *et al.*, 2024] proposes learning representations generalizable to unseen domains by constructing modality-consistent flat loss regions via cross-modal knowledge transfer. However, these works primarily focus on other modalities instead of physiological signals. However, these methods are primarily developed for general multimodal data and have not been investigated for physiological signals.

3 Feasibility of Multimodal Coordination for Physiological Signals

The practice of multimodal learning paradigms has demonstrated that models trained with multiple modalities can outperform those trained on a single modality, even if only one

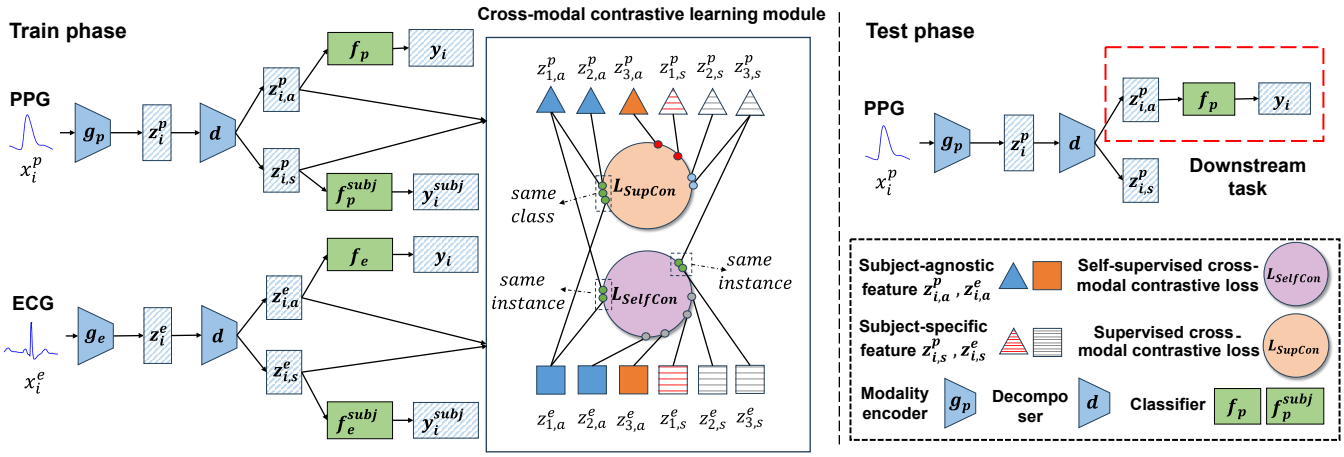


Figure 2: Illustration of the proposed PhysioGMC using PPG and ECG signals as example input modalities. In the cross-modal contrastive learning module, feature colors (blue, orange, etc) indicate categories of downstream tasks, while feature shapes indicate modality types.

modality is available at deployment. Previous studies [Zadeh *et al.*, 2020; Lu, 2023] attribute this advantage to the interrelated and complementary properties of the modalities used during training. Building on this insight, we refine the conditions introduced in [Zadeh *et al.*, 2020] under which multimodal coordination improves upon unimodal learning.

Let $\mathcal{F}_\Theta$ denote a model parameterized by learnable parameters Θ. During training, the model has access to multimodal data $\{x_i \mid i \in M\}$ and the corresponding downstream-task label y, where M denotes the full set of modalities. At test time, only a subset of modalities, $\{x_m \mid m \in S\}$, is available, where $S \subset M$. The set of modalities unavailable at test time is given by the complement $S^c = M \setminus S$.

Theorem 1. *Assume that the modalities in M are mutually informative, i.e., ($I(\{x_m \mid m \in S\}; \{x_n \mid n \in S^c\}) > 0$), and that incorporating additional modalities improves downstream task performance: ($H(y \mid \{x_m \mid m \in M\}) < H(y \mid \{x_m \mid m \in S\}) < H(y)$), where $I(\cdot;\cdot)$ denotes mutual information and $H(\cdot)$ denotes entropy. Under these assumptions, there exist model parameters $\theta \in \Theta$ such that,*

$$\begin{aligned} &H(y \mid \{x_m \mid m \in M\}) < \\ &H(y \mid \{x_m \mid m \in S\}, \theta) < H(y \mid \{x_m \mid m \in S\}). \end{aligned} \tag{1}$$

The right-hand inequality in Eq. (1) holds if the following holds:

$$\begin{aligned} &H(\{x_n \mid n \in S^c\} \mid \{x_m \mid m \in S\}) > \\ &\quad H(\{x_n \mid n \in S^c\} \mid \{x_m \mid m \in S\}, \theta); \end{aligned} \tag{2}$$

$$\begin{aligned} &I(y; \{x_n \mid n \in S^c\} \mid \{x_m \mid m \in S\}) > \\ &\quad I(y; \{x_n \mid n \in S^c\} \mid \{x_m \mid m \in S\}, \theta). \end{aligned} \tag{3}$$

Although Zadeh et al. [Zadeh *et al.*, 2020] identify Eq. (2) as a necessary condition for multimodal co-learning to improve unimodal performance, this condition alone is insufficient. Our extension, summarized in Eq. (3) in Theorem 1, demonstrates that, by capturing connections between modality sets $\{x_m \mid m \in S\}$ and $\{x_n \mid n \in S^c\}$ that preserve label-relevant information, multimodal models can outperform unimodal counterparts, even when only a subset of modalities is available at inference, as stated in Eq. (1).

	MIMIC-III [Johnson *et al.*, 2016]			
	ECG		PPG	
	Accuracy(↑)	Macro F1(↑)	Accuracy(↑)	Macro F1(↑)
Intra-subject eval.				
Unimodal	0.980	0.980	0.858	0.857
Multimodal SimCLR	**0.988**	**0.988**	**0.880**	**0.879**
Inter-subject eval.				
Unimodal	0.672	0.671	0.651	0.651
Multimodal SimCLR	**0.680**	**0.679**	**0.683**	**0.683**

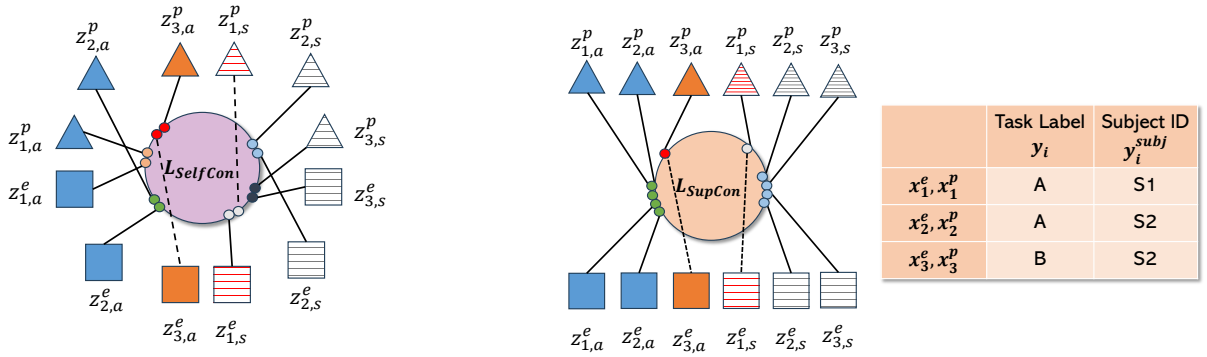
Table 1: Performance comparison for AF detection on a subset of MIMIC-III under intra-subject and inter-subject evaluation settings. In the intra-subject setting, training and testing samples are collected at different times from the same subjects. In the inter-subject setting, evaluation is performed on subjects excluded from training.

To empirically validate this principle in the context of physiological signals, we adopt the multimodal SimCLR [Chen *et al.*, 2020] framework as a representative multimodal co-learning approach and apply it to the AF detection task. It enforces cross-modal alignment by treating temporally synchronized signal samples from different modalities as positive pairs, encouraging the learned label-decisive representations to encode shared information across modalities. Therefore, SimCLR instantiates the conditions for effective multimodal co-learning formalized in Theorem 1, making it an appropriate choice for empirically validating the benefits of multimodal co-learning over unimodal learning.

Despite improved performance over unimodal baselines, the results in Table 1 show that conventional multimodal coordination methods suffer from notable inter-subject variability and limited generalization to unseen subjects. This motivates our proposed framework, PhysioGMC, which collaboratively learns generalizable representations from multimodal physiological signals.

4 Proposed Method

We now provide an overview of the proposed PhysioGMC framework, as illustrated in Fig. 2. PhysioGMC consists of two primary components: 1) a unimodal feature extraction and classification backbone that learns representations from



	Task Label y_i	Subject ID y_i^{subj}
x_1^e, x_1^p	A	S1
x_2^e, x_2^p	A	S2
x_3^e, x_3^p	B	S2

(a) Self-supervised contrastive loss (b) Supervised contrastive loss

Figure 3: Illustration of the proposed cross-modal contrastive objectives: (a) self-supervised contrastive loss and (b) supervised contrastive loss with the corresponding label information. Given an input x_i^m from modality m, we compute $(z_{i,a}^m, z_{i,s}^m) = d(g_m(x_i^m))$, where $z_{i,a}^m$ and $z_{i,s}^m$ denote the label-relevant and subject-specific features, respectively. Along each circle, features pointing to dots of the same color are treated as positive pairs in the contrastive loss. The legends for both subfigures follow those in Fig. 2.

each physiological modality; and 2) a cross-modal contrastive learning module that combines a self-supervised contrastive loss for instance-level alignment across modalities with a supervised contrastive loss for aligning label-relevant features from the same class across modalities while separating them from features of other classes and subject-specific features.

4.1 Unimodal Feature Extraction and Classification

In the proposed framework, we adopt EfficientNet [Tan and Le, 2019] as the backbone encoder for extracting unimodal features. For a given input x_i^m of the modality m, the corresponding unimodal feature z_i^m is obtained as $z_i^m = g_m(x_i^m)$, where $g_m(\cdot)$ is the modality-specific feature encoder and $i \in \{1, \cdots, N\}$ is the input index. The learned feature z_i^m is further decomposed into the label-relevant, subject-agnostic feature $z_{i,a}^m$ and the subject-specific feature $z_{i,s}^m$ using a feature decomposer $d(\cdot)$ These decomposed features are subsequently used for downstream classification.

The modality-specific classification loss is defined as:

$$\mathcal{L}_{cls}^m = \frac{1}{N}\sum_{i=1}^{N}(\mathcal{L}_{CE}(f_m(z_{i,a}^m), y_i) + \mathcal{L}_{CE}(f_m^{subj}(z_{i,s}^m), y_i^{subj})), \tag{4}$$

where $\mathcal{L}_{CE}$ denotes the cross-entropy (CE) loss, f_m and f_m^{subj} are the modality-specific classifiers, y_i is the label of downstream tasks, and y_i^{subj} is the subject ID label. Minimizing the above classification loss encourages the model to learn the label-related features and the identity-related features for each modality separately.

4.2 Cross-modal Contrastive Learning

As discussed in Section 3, multimodal learning paradigms can outperform unimodal counterparts when they capture label-decisive mappings across different modalities. Motivated by this insight, we propose a cross-modal contrastive learning module comprising two objectives: a self-supervised contrastive loss and a supervised contrastive loss. This module learns subject-agnostic, label-relevant representations that are consistently aligned across modalities, thereby implicitly modeling label-discriminative relationships among different physiological signals.

For the contrastive learning, the extracted unimodal features of the modality m are normalized using L_2 normalization as $\widetilde{z_{i,a}^m} = \frac{z_{i,a}^m}{\|z_{i,a}^m\|}, \widetilde{z_{i,s}^m} = \frac{z_{i,s}^m}{\|z_{i,s}^m\|}$. We assume that each batch of the data used for training comprises N quadruplets, $\{\mathbf{x_i}, y_i, y_i^{subj}\}_{i=1,\cdots,N}$, where $\mathbf{x}_i = [x_i^m; x_i^k]$ denotes temporally synchronized data instances from two physiological modalities m and k, and i is the input index.

Self-supervised Contrastive Loss

As the first component of the proposed module, we design a self-supervised cross-modal contrastive loss, defined in Eq. (5), that enforces alignment between representations of the same input data across modalities.

$$\mathcal{L}_{SelfCon} = -\frac{1}{N}\sum_{i=1}^{N}(log\frac{exp(\widetilde{z_{i,a}^m}\cdot\widetilde{z_{i,a}^k}/\tau)}{\sum_{j=1}^{N}exp(\widetilde{z_{j,a}^m}\cdot\widetilde{z_{j,a}^k}/\tau)} + log\frac{exp(\widetilde{z_{i,s}^m}\cdot\widetilde{z_{i,s}^k}/\tau)}{\sum_{j=1}^{N}exp(\widetilde{z_{j,s}^m}\cdot\widetilde{z_{j,s}^k}/\tau)}), \tag{5}$$

where τ is the temperature scaling parameter, controlling the sharpness of the similarity scores, and $\cdot$ is the inner product of two normalized features to measure their similarity.

The self-supervised contrastive loss encourages representations extracted from temporally paired modalities to be aligned in the embedding space, without relying on label supervision. Consequently, it enables the effective incorporation of large-scale unlabeled multimodal physiological data during training. An illustration of the self-supervised cross-modal contrastive loss is provided in Fig. 3(a).

Supervised Contrastive Loss

Although the self-supervised contrastive loss aligns features across modalities, the learned features may still contain label-irrelevant or subject-specific information. The entanglement

of these two kinds of information can hinder discrimination and generalization of learned representations in downstream tasks, leading to degraded cross-subject performance. Thus, we propose a modified supervised contrastive loss to learn multimodal features that only contain subject-agnostic, label-decisive information. This contrastive loss utilizes both task and subject label information as supervision, drawing features from the same class closer in the embedding space while pushing apart features from different classes, as illustrated in Fig. 3(b). This loss consists of two loss terms: the unimodal contrastive loss term and the cross-modal contrastive loss term.

To construct the supervised contrastive loss, we concatenate the normalized label-relevant subject-agnostic feature $\{\widetilde{z_{i,a}^m}\}_{i=1,\cdots,N}$ and the normalized subject-specific feature $\{\widetilde{z_{i,s}^m}\}_{i=1,\cdots,N}$ into a new feature vector $z_m^{'} = [\{\widetilde{z_{i,a}^m}\}_{i=1,\cdots,N}; \{\widetilde{z_{i,s}^m}\}_{i=1,\cdots,N}]$. The task label y_i is assigned to $\widetilde{z_{i,a}^m}$ while the subject ID label y_i^{subj} is assigned to $\widetilde{z_{i,s}^m}$. Consequently, we define the combined label vector as $y^{'} = [\{y_i\}_{i=1,\cdots,N}; \{y_i^{subj}\}_{i=1,\cdots,N}]$. The modified supervised contrastive loss is denoted as follows:

$$\begin{aligned}\mathcal{L}_{\text{SupCon}} &= \mathcal{L}_{\text{SupCon}}^{\text{uni}} + \beta_1 \mathcal{L}_{\text{SupCon}}^{\text{cross}} \\ &= -\sum_{m\in\mathcal{M}}\sum_{i=1}^{2N}\frac{1}{N^{'}}\sum_{j\in P(i)}\log\frac{\exp(z_{m,i}^{'}\cdot z_{m,j}^{'}/\tau)}{\sum_{r\in A(i)}\exp(z_{m,i}^{'}\cdot z_{m,r}^{'}/\tau)} \\ &\quad -\beta_1\sum_{m\in\mathcal{M}}\sum_{k\in\mathcal{M}^c}\sum_{i=1}^{2N}\frac{1}{N^{'}}\sum_{j\in P(i)}\log\frac{\exp(z_{m,i}^{'}\cdot z_{k,j}^{'}/\tau)}{\sum_{r\in A(i)}\exp(z_{m,i}^{'}\cdot z_{k,r}^{'}/\tau)},\end{aligned} \tag{6}$$

where $i \in I \equiv \{1,\cdots,2N\}$ is the input index. We define $P(i) \equiv \{p : y_p^{'} = y_i^{'}\}$ and $A(i) \equiv I\backslash\{i\}$. $N^{'}$ is the cardinality of $P(i)$. $\mathcal{M}$ is the set of available modalities during the training process. β_1 is the hyperparameter that controls the relative importance of unimodal and cross-modal contrastive learning terms. τ and the similarity operator $\cdot$ have the same definitions as in the self-supervised contrastive loss.

In this supervised contrastive loss, the unimodal contrastive loss term encourages within-class consistency by attracting the same-class features within each modality. Complementarily, the cross-modal contrastive loss term encourages inter-modal consistency by aligning the features of the same class across modalities. Together, these terms guide the model to extract label-relevant information from multimodal data.

A supervised contrastive loss based solely on task labels may still preserve subject-specific information in the learned representations, thereby limiting generalization to unseen subjects. To mitigate this concern, we incorporate subject-specific features $\widetilde{z_{i,s}^m}$ labeled with subject identities into the modified supervised contrastive loss. By treating the label-relevant feature $\widetilde{z_{i,a}^m}$ and the subject-specific feature $\widetilde{z_{i,s}^m}$ extracted from the same input data as belonging to different classes, the proposed supervised contrastive loss effectively disentangles label-relevant information and subject-specific information, yielding label-relevant representations that are more generalizable across subjects.

4.3 Final Loss

The final loss function is defined as the weighted sum of the previously defined losses:

$$\mathcal{L} = \lambda_{cls}\sum_{m\in\mathcal{M}}\mathcal{L}_{cls}^m + \lambda_{self}\mathcal{L}_{SelfCon} + \lambda_{sup}\mathcal{L}_{SupCon}, \tag{7}$$

where λ_{cls}, λ_{self}, and λ_{sup} are the hyperparameters that control the relative importance of different loss terms.

5 Experiments and Results

5.1 Experimental Settings

Datasets. We conduct experiments on two subsets derived from the real-world ICU dataset MIMIC-III [Johnson *et al.*, 2016], referred to as MIMIC-AF and MIMIC-Hyper, as well as on a subset of the MESA dataset [Chen *et al.*, 2015], which includes 2237 participants enrolled in a sleep exam. Both MIMIC-AF and MIMIC-Hyper comprise physiological recordings from 300 critically ill subjects. The former subset contains ECG and PPG signals, while the latter contains ABP and PPG signals. All signals are segmented into 10-second windows for both ICU subsets. The MESA subset comprises overnight PSG recordings from 100 subjects, each containing synchronized 30-second ECG and PPG segments.

Downstream Tasks. We use two cardiovascular health monitoring tasks and one sleep monitoring task for evaluation over three datasets: AF detection on the MIMIC-AF, blood pressure classification on the MIMIC-Hyper, and sleep stage classification on the MESA subset. For the MIMIC-AF, each segment is labeled according to the corresponding ICD-9 diagnostic codes. For the MIMIC-Hyper, segments are categorized into four classes: normal, high-normal, Grade 1 hypertension, and Grades 2–3 hypertension. The categorization is based on their systolic blood pressure (SBP) and diastolic blood pressure (DBP) values, following the ESC/ESH guidelines [Mancia *et al.*, 2014]. For the MESA subset, we consider a four-class sleep-staging task comprising wake, light sleep (N1/N2), deep sleep (N3), and Rapid Eye Movement(REM).

Evaluation Settings. To evaluate the generalization performance of PhysioGMC, we perform a five-fold cross-subject validation with no overlap between training and testing subjects. The final results are averaged across all five folds for an overall evaluation. We run each experiment with five random seeds to account for performance variability and report the mean and standard deviation of the five runs' results. Additionally, the Wilcoxon signed-rank test is used to assess the statistical significance of performance improvements achieved by the proposed PhysioGMC over baseline methods.

5.2 Experiment Results

To evaluate the proposed PhysioGMC, we compare it with eight baseline methods. All methods are evaluated under the setting where clinical ECG or ABP signals are available as auxiliary modalities only during training, while PPG serves as the primary modality available during both training and inference.

Method	MIMIC-AF			MIMIC-Hyper		
	Accuracy (↑)	Macro F1 (↑)	AUC-ROC (↑)	Accuracy (↑)	Macro F1 (↑)	AUC-ROC (↑)
Unimodal						
Supervised w/ CE	$0.6667_{\pm 0.003}$	$0.6658_{\pm 0.003}$	$0.7236_{\pm 0.004}$	$0.4146_{\pm 0.006}$	$0.3274_{\pm 0.002}$	$0.6089_{\pm 0.003}$
Supervised w/ CE	$0.667_{\pm 0.003}$	$0.666_{\pm 0.003}$	$0.724_{\pm 0.004}$	$0.415_{\pm 0.006}$	$0.327_{\pm 0.002}$	$0.609_{\pm 0.003}$
Unimodal DG						
BPD [Su *et al.*, 2022]	$0.681_{\pm 0.003}$	$0.679_{\pm 0.003}$	$0.734_{\pm 0.010}$	$0.410_{\pm 0.007}$	$0.334_{\pm 0.003}$	$\underline{0.634}_{\pm 0.008}$
CSGR [Kim *et al.*, 2024]	$0.657_{\pm 0.004}$	$0.655_{\pm 0.004}$	$0.711_{\pm 0.007}$	$0.414_{\pm 0.005}$	$0.324_{\pm 0.002}$	$0.603_{\pm 0.002}$
Multimodal						
Multimodal SimCLR [Chen *et al.*, 2020]	$\underline{0.687}_{\pm 0.004}$	$\underline{0.686}_{\pm 0.004}$	$\underline{0.748}_{\pm 0.007}$	$\underline{0.427}_{\pm 0.005}$	$\underline{0.339}_{\pm 0.003}$	$0.624_{\pm 0.009}$
MERL [Liu *et al.*, 2024]	$0.670_{\pm 0.007}$	$0.668_{\pm 0.007}$	$0.724_{\pm 0.005}$	$0.415_{\pm 0.006}$	$0.327_{\pm 0.003}$	$0.612_{\pm 0.009}$
SiamAF [Guo *et al.*, 2023]	$0.661_{\pm 0.004}$	$0.661_{\pm 0.004}$	$0.715_{\pm 0.006}$	$0.422_{\pm 0.004}$	$0.326_{\pm 0.003}$	$0.608_{\pm 0.007}$
Multimodal DG						
SimMMDG [Dong *et al.*, 2023]	$0.667_{\pm 0.004}$	$0.665_{\pm 0.004}$	$0.720_{\pm 0.006}$	$0.421_{\pm 0.004}$	$0.325_{\pm 0.002}$	$0.603_{\pm 0.005}$
CMRF [Fan *et al.*, 2024]	$0.660_{\pm 0.002}$	$0.659_{\pm 0.001}$	$0.718_{\pm 0.002}$	$0.417_{\pm 0.002}$	$0.323_{\pm 0.003}$	$0.607_{\pm 0.004}$
Proposed PhysioGMC	$\mathbf{0.713}^{*}_{\pm 0.005}$	$\mathbf{0.711}^{*}_{\pm 0.005}$	$\mathbf{0.776}^{*}_{\pm 0.005}$	$\mathbf{0.436}^{*}_{\pm 0.005}$	$\mathbf{0.360}^{*}_{\pm 0.002}$	$\mathbf{0.649}^{*}_{\pm 0.004}$

Table 2: Performance comparison of baseline methods and the proposed PhysioGMC on MIMIC-AF and MIMIC-Hyper datasets for AF detection and blood pressure classification tasks. All methods are trained with multiple modalities as input and evaluated using only PPG signals at inference. Bold and underlined items denote the best and second-best results, respectively. An asterisk $(*)$ indicates a statistically significant improvement over the best baseline method ($p < 0.05$).

Method	MESA subset			
	Accuracy (↑)	Macro F1 (↑)	AUC-ROC (↑)	Kappa (↑)
Unimodal				
Supervised w/ CE	$0.493_{\pm 0.007}$	$0.442_{\pm 0.005}$	$0.720_{\pm 0.003}$	$0.286_{\pm 0.006}$
Unimodal DG				
BPD	$0.488_{\pm 0.008}$	$0.451_{\pm 0.005}$	$0.729_{\pm 0.003}$	$0.292_{\pm 0.007}$
CSGR	$0.484_{\pm 0.008}$	$0.442_{\pm 0.004}$	$0.720_{\pm 0.005}$	$0.286_{\pm 0.005}$
Multimodal				
Multimodal SimCLR	$\underline{0.504}_{\pm 0.010}$	$\underline{0.457}_{\pm 0.006}$	$\mathbf{0.736}_{\pm 0.004}$	$\underline{0.302}_{\pm 0.010}$
MERL	$0.492_{\pm 0.006}$	$0.450_{\pm 0.003}$	$0.726_{\pm 0.005}$	$0.293_{\pm 0.006}$
SiamAF	$0.464_{\pm 0.008}$	$0.408_{\pm 0.009}$	$0.673_{\pm 0.009}$	$0.237_{\pm 0.011}$
Multimodal DG				
SimMMDG	$0.472_{\pm 0.037}$	$0.395_{\pm 0.015}$	$0.717_{\pm 0.011}$	$0.259_{\pm 0.026}$
CMRF	$0.449_{\pm 0.006}$	$0.400_{\pm 0.005}$	$0.734_{\pm 0.002}$	$0.246_{\pm 0.003}$
Proposed PhysioGMC	$\mathbf{0.517}^{*}_{\pm 0.004}$	$\mathbf{0.461}_{\pm 0.004}$	$\underline{0.735}_{\pm 0.005}$	$\mathbf{0.314}^{*}_{\pm 0.004}$

Table 3: Performance comparison of baseline methods and the proposed PhysioGMC on the MESA subset for the sleep stage classification. Bold and underlined items denote the best and second-best results, respectively. An asterisk $(*)$ indicates a statistically significant improvement over the best baseline method ($p < 0.05$).

Method	Macro F1 (↑)		
	0%	40%	70%
Supervised w/ CE	$0.666_{\pm 0.003}$	$0.661_{\pm 0.002}$	$0.656_{\pm 0.002}$
Proposed PhysioGMC	$\mathbf{0.711}_{\pm 0.005}$	$\mathbf{0.708}_{\pm 0.001}$	$\mathbf{0.705}_{\pm 0.004}$

Table 4: Effect of different proportions of missing labels (0%, 40%, 70%) on AF detection performance on the MIMIC-AF dataset.

The Cardiovascular Health Monitoring

The experimental results for AF detection and blood pressure classification are summarized in Table 2. First, we observe that unimodal methods designed to address inter-subject variability in physiological signals, such as BPD [Su *et al.*, 2022], achieve better cross-subject test performance compared to the vanilla supervised learning counterpart. Second, certain multimodal coordination methods, such as SimCLR [Chen *et al.*, 2020], outperform unimodal counterparts, demonstrating the effectiveness of multimodal coordination methods when only unimodal data is available during deployment. In contrast, the two multimodal DG methods [Dong *et al.*, 2023; Fan *et al.*, 2024] show inferior performance, likely due to the strong dominance of ECG/ABP signals in their framework, which may hinder the models' generalization when only PPG signals are available at test time.

Overall, our proposed method consistently outperforms all baseline approaches across all three evaluation metrics on two cardiovascular health monitoring tasks. It achieves about 2.5% absolute improvement in *Accuracy*, *Macro F1*, and *AUC_ROC* scores for the AF detection task compared to the second-best method. For the blood pressure classification task, it improves *Accuracy* by 1%, *Macro F1* by 2%, and *AUC_ROC* scores by 1.6%. Specifically, the proposed method achieves statistically significant improvements over the second-best method across all evaluation metrics.

The Sleep Stage Classification

Table 3 presents experimental results of the sleep stage classification task on the MESA subset. Consistent with the findings in the previous section, multimodal coordination methods, including SimCLR [Chen *et al.*, 2020] and MERL [Liu *et al.*, 2024], outperform unimodal approaches. The proposed framework, PhysioGMC, achieves the best performance among all evaluated methods. These results demonstrate the superiority of learning from multiple physiological signals in a collaborative manner during training, even when only unimodal wearable signals are available at deployment, offering more practical and comfortable options for continuous sleep monitoring.

Extension to Semi-supervised Learning Settings

As discussed in Section 4.2, the proposed framework can extend to semi-supervised learning settings where models are trained using both labeled data and unlabeled data. In this section, we evaluate the performance of PhysioGMC in the semi-supervised scenario by randomly masking labels at ratios of $\{0\%, 40\%, 70\%\}$. We compare PhysioGMC against

ID Method	MIMIC-AF			MIMIC-Hyper		
	Accuracy (↑)	Macro F1 (↑)	AUC-ROC (↑)	Accuracy (↑)	Macro F1 (↑)	AUC-ROC (↑)
Unimodal						
A1 Supervised w/ CE	$0.667_{\pm 0.003}$	$0.666_{\pm 0.003}$	$0.724_{\pm 0.004}$	$0.415_{\pm 0.006}$	$0.327_{\pm 0.002}$	$0.609_{\pm 0.003}$
A2 Supervised contrastive loss	$0.681_{\pm 0.005}$	$0.678_{\pm 0.005}$	$0.728_{\pm 0.007}$	$0.409_{\pm 0.011}$	$0.337_{\pm 0.005}$	$0.626_{\pm 0.007}$
Multimodal						
B1 Proposed w/o self-supervised contrastive loss	$\underline{0.682}_{\pm 0.006}$	$0.681_{\pm 0.006}$	$0.735_{\pm 0.012}$	$0.409_{\pm 0.011}$	$0.334_{\pm 0.006}$	$\underline{0.628}_{\pm 0.007}$
B2 Proposed w/o supervised contrastive loss	$0.682_{\pm 0.002}$	$\underline{0.681}_{\pm 0.002}$	$\underline{0.738}_{\pm 0.002}$	$\underline{0.430}_{\pm 0.005}$	$\underline{0.347}_{\pm 0.003}$	$0.627_{\pm 0.005}$
— Proposed PhysioGMC	$\mathbf{0.713}^{*}_{\pm 0.005}$	$\mathbf{0.711}^{*}_{\pm 0.005}$	$\mathbf{0.776}^{*}_{\pm 0.005}$	$\mathbf{0.436}_{\pm 0.005}$	$\mathbf{0.359}^{*}_{\pm 0.002}$	$\mathbf{0.649}^{*}_{\pm 0.004}$

Table 5: Ablation study evaluating the contribution of each component of the proposed PhysioGMC framework on the MIMIC-AF and MIMIC-Hyper datasets. Bold and underlined items denote the best and second-best results, respectively. An asterisk ($*$) indicates a statistically significant improvement over the best baseline method ($p < 0.05$).

the unimodal supervised learning baseline trained only on the available labeled data. As shown in Table 4, PhysioGMC outperforms the baseline across all label-missing ratios, demonstrating its scalability in semi-supervised settings by leveraging unlabeled data through the self-supervised cross-modal contrastive loss in the training process.

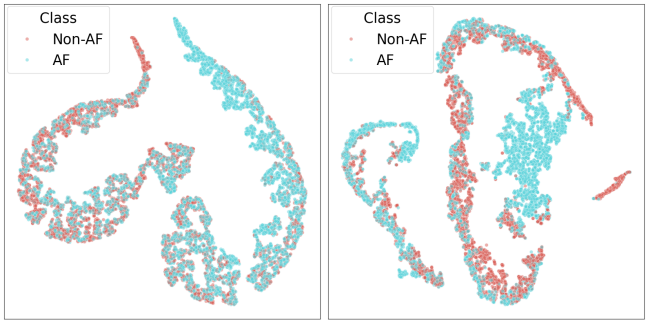
5.3 Ablation Experiments

To further illustrate the contributions of each module in the proposed framework, we conduct ablation experiments evaluating the impact of each component of the proposed PhysioGMC framework on the MIMIC-AF and MIMIC-Hyper datasets. The results of ablation experiments, as shown in Table 5, demonstrate the effectiveness of both self-supervised and supervised cross-modal contrastive losses. The B2 variant, which incorporates the self-supervised contrastive loss only, achieves notable improvements in performance compared to A1 under unimodal testing scenarios, demonstrating the advantage of multimodal coordination paradigms that encourage learning modality-consistent features. The proposed framework, which additionally incorporates the supervised contrastive loss, further enhances cross-subject performance over B2 by explicitly enforcing the disentanglement of subject-agnostic information and subject-specific information in the learned multimodal label-decisive features.

We also evaluate the effectiveness of the proposed supervised contrastive loss when training with a single modality. Compared to the unimodal baseline A1, variant A2 achieves improved cross-subject test performance, suggesting its ability to learn label-relevant representations agnostic to subjects. However, the proposed framework with multimodal contrastive learning modules outperforms the unimodal variant, demonstrating the advantage of learning subject-agnostic, label-relevant features collaboratively across multiple modalities rather than from a single modality.

5.4 Feature Visualization Analysis

In this section, we utilize t-SNE [Maaten and Hinton, 2008] to visualize the PPG features, $z^{p}_{i,a}$, learned by the baseline unimodal model and PhysioGMC on the MIMIC-AF dataset. As shown in Fig. 4, the embeddings learned by PhysioGMC exhibit clearer, more distinguishable classification boundaries. For subjects without AF, the distribution of PPG embeddings is inherently more dispersed due to heterogeneous cardiovascular conditions. Nevertheless, the proposed framework yields a more compact distribution for the non-AF cohort than



(a) Unimodal EfficientNet (b) Proposed PhysioGMC

Figure 4: t-SNE visualization of PPG features ($z^{p}_{i,a}$) on the MIMIC-AF dataset using (a) a unimodal EfficientNet baseline and (b) the proposed PhysioGMC framework.

the unimodal baseline, demonstrating its ability to suppress subject-specific information in the extracted label-relevant features. While there remains some overlap between classes, reflecting the challenges brought by inter-subject variability of PPG signals, these results show improved feature separability for our framework. Further exploration may help achieve more distinct decision boundaries for unseen subjects.

6 Conclusion

We propose PhysioGMC, a generalizable multimodal coordination framework for physiological signals that leverages both clinical-grade and wearable modalities during training. The proposed framework learns label-relevant, subject-agnostic features collaboratively across multiple modalities by integrating proposed self-supervised and supervised cross-modal contrastive losses, leading to improved cross-subject performance for wearable modalities under unimodal deployment scenarios. Experiments on cardiovascular health monitoring and sleep stage classification tasks show that PhysioGMC consistently achieves superior cross-subject performance when only a single wearable physiological signal, PPG, is available during testing for these tasks. This demonstrates the potential of PhysioGMC to support continuous, accessible, and reliable health monitoring using only wearable sensing in real-world applications.

Acknowledgments

This work was supported in part by the U.S. National Science Foundation (NSF) grant 214291.

References

[BN *et al.*, 2024] Suhas BN, Rakshith Sharma Srinivasa, Yashas Malur Saidutta, Jaejin Cho, Ching-Hua Lee, Chouchang Yang, Yilin Shen, and Hongxia Jin. End-To-End Personalized Cuff-Less Blood Pressure Monitoring Using ECG and PPG Signals. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2101–2105, 2024.

[Chen *et al.*, 2015] Xiaoli Chen, Rui Wang, Phyllis Zee, Pamela L Lutsey, Sogol Javaheri, Carmela Alcántara, Chandra L Jackson, Michelle A Williams, and Susan Redline. Racial/Ethnic Differences in Sleep Disturbances: The Multi-Ethnic Study of Atherosclerosis (MESA). *Sleep*, 38(6):877, 2015.

[Chen *et al.*, 2020] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607, 2020.

[Das *et al.*, 2022] Sarkar Snigdha Sarathi Das, Subangkar Karmaker Shanto, Masum Rahman, Md Saiful Islam, Atif Hasan Rahman, Mohammad M. Masud, and Mohammed Eunus Ali. Bayesbeat: Reliable atrial fibrillation detection from noisy photoplethysmography data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(1), 2022.

[Dong *et al.*, 2023] Hao Dong, Ismail Nejjar, Han Sun, Eleni Chatzi, and Olga Fink. SimMMDG: A Simple and Effective Framework for Multi-modal Domain Generalization. In *Advances in Neural Information Processing Systems*, volume 36, pages 78674–78695, 2023.

[Fan *et al.*, 2024] Yunfeng Fan, Wenchao Xu, Haohao Wang, and Song Guo. Cross-modal Representation Flattening for Multi-modal Domain Generalization. In *Advances in Neural Information Processing Systems*, volume 37, pages 66773–66795, 2024.

[Gil *et al.*, 2010] E Gil, M Orini, R Bailón, J M Vergara, L Mainardi, and P Laguna. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiological Measurement*, 31(9):1271, 2010.

[Guo *et al.*, 2023] Zhicheng Guo, Cheng Ding, Duc H. Do, Amit J. Shah, Randall J. Lee, Xiao Hu, and Cynthia Rudin. SiamAF: Learning Shared Information from ECG and PPG Signals for Robust Atrial Fibrillation Detection. *CoRR*, abs/2310.09203, 2023.

[Han *et al.*, 2024] Jinpei Han, Xiao Gu, Guang-Zhong Yang, and Benny Lo. Noise-Factorized Disentangled Representation Learning for Generalizable Motor Imagery EEG Classification. *IEEE Journal of Biomedical and Health Informatics*, 28(2):765–776, 2024.

[Hwang *et al.*, 2021] Sunhee Hwang, Sungho Park, Dohyung Kim, Jewook Lee, and Hyeran Byun. Mitigating Inter-Subject Brain Signal Variability For EEG-Based Driver Fatigue State Classification. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 990–994, 2021.

[Ji and Zhou, 2024] Hui Ji and Pengfei Zhou. Advancing PPG-Based Continuous Blood Pressure Monitoring from a Generative Perspective. In *Proceedings of the 22nd ACM Conference on Embedded Networked Sensor Systems*, SenSys '24, page 661–674, 2024.

[Johnson *et al.*, 2016] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[Kim *et al.*, 2024] Hyeonji Kim, Jaehoon Kim, and Seoung Bum Kim. Cross-subject generalizable representation learning with class-subject dual labels for biosignals. *Knowledge-Based Systems*, 295:111855, 2024.

[Kotzen *et al.*, 2023] Kevin Kotzen, Peter H. Charlton, Sharon Salabi, Lea Amar, Amir Landesberg, and Joachim A. Behar. SleepPPG-Net: A Deep Learning Algorithm for Robust Sleep Staging From Continuous Photoplethysmography. *IEEE Journal of Biomedical and Health Informatics*, 27(2):924–932, 2023.

[Leitner *et al.*, 2022] Jared Leitner, Po-Han Chiang, and Sujit Dey. Personalized Blood Pressure Estimation Using Photoplethysmography: A Transfer Learning Approach. *IEEE Journal of Biomedical and Health Informatics*, 26(1):218–228, 2022.

[Liu *et al.*, 2023] Hong Liu, Dong Wei, Donghuan Lu, Jinghan Sun, Liansheng Wang, and Yefeng Zheng. M3AE: Multimodal Representation Learning for Brain Tumor Segmentation with Missing Modalities. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(2):1657–1665, 2023.

[Liu *et al.*, 2024] Che Liu, Zhongwei Wan, Cheng Ouyang, Anand Shah, Wenjia Bai, and Rossella Arcucci. Zero-Shot ECG Classification with Multimodal Learning and Test-time Clinical Knowledge Enhancement. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 31949–31963, 2024.

[Lu, 2023] Zhou Lu. A Theory of Multimodal Learning. In *Advances in Neural Information Processing Systems*, volume 36, pages 57244–57255, 2023.

[Ma *et al.*, 2021] Mengmeng Ma, Jian Ren, Long Zhao, Sergey Tulyakov, Cathy Wu, and Xi Peng. Smil: Multimodal learning with severely missing modality. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(3):2302–2310, 2021.

[Ma *et al.*, 2025] Shuo Ma, Yingwei Zhang, Yiqiang Chen, Hualei Wang, Yuan Jin, Wei Zhang, and Ziyu Jia.

SleepSMC: Ubiquitous Sleep Staging via Supervised Multimodal Coordination. In *International Conference on Representation Learning*, volume 2025, pages 24170–24191, 2025.

[Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[Mancia *et al.*, 2014] Giuseppe Mancia, Robert Fagard, Krzysztof Narkiewicz, Josep Redon, Alberto Zanchetti, Michael Böhm, Thierry Christiaens, Renata Cifkova, Guy De Backer, Anna Dominiczak, et al. 2013 ESH/ESC practice guidelines for the management of arterial hypertension: ESH-ESC the task force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Blood pressure*, 23(1):3–16, 2014.

[Mukkamala and Hahn, 2018] Ramakrishna Mukkamala and Jin-Oh Hahn. Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Predictions on Maximum Calibration Period and Acceptable Error Limits. *IEEE Transactions on Biomedical Engineering*, 65(6):1410–1420, 2018.

[Olsen *et al.*, 2024] Mads Olsen, Jamie M. Zeitzer, Risa Nakase-Richardson, Valerie H Musgrave, Helge B. D. Sørensen, Emmanuel Mignot, and Poul J. Jennum. A Deep Transfer Learning Approach for Sleep Stage Classification and Sleep Apnea Detection Using Wrist-Worn Consumer Sleep Technologies. *IEEE Transactions on Biomedical Engineering*, 71(8):2506–2517, 2024.

[Planamente *et al.*, 2022] Mirco Planamente, Chiara Plizzari, Emanuele Alberti, and Barbara Caputo. Domain Generalization Through Audio-Visual Relative Norm Alignment in First Person Action Recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1807–1818, 2022.

[Rahate *et al.*, 2022] Anil Rahate, Rahee Walambe, Sheela Ramanna, and Ketan Kotecha. Multimodal Co-learning: Challenges, applications with datasets, recent advances and future directions. *Information Fusion*, 81:203–239, 2022.

[Sharan *et al.*, 2020] Roneel V. Sharan, Shlomo Berkovsky, Hao Xiong, and Enrico Coiera. ECG-Derived Heart Rate Variability Interpolation and 1-D Convolutional Neural Networks for Detecting Sleep Apnea. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 637–640, 2020.

[Su *et al.*, 2022] Jie Su, Zhenyu Wen, Tao Lin, and Yu Guan. Learning Disentangled Behaviour Patterns for Wearable-based Human Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(1), 2022.

[Tan and Le, 2019] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114, 2019.

[Tian *et al.*, 2023] Xin Tian, Qiang Zhu, Yuenan Li, and Min Wu. Cross-Domain Joint Dictionary Learning for ECG Inference From PPG. *IEEE Internet of Things Journal*, 10(9):8140–8154, 2023.

[Torres-Soto and Ashley, 2020] Jessica Torres-Soto and Euan A Ashley. Multi-task deep learning for cardiac rhythm detection in wearable devices. *NPJ digital medicine*, 3(1):116, 2020.

[Vo *et al.*, 2024] Khuong Vo, Mostafa El-Khamy, and Yoojin Choi. PPG-to-ECG Signal Translation for Continuous Atrial Fibrillation Detection via Attention-based Deep State-Space Modeling. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–7, 2024.

[Wu *et al.*, 2024] Zhenbang Wu, Anant Dadu, Nicholas J. Tustison, Brian B. Avants, Mike A. Nalls, Jimeng Sun, and Faraz Faghri. Multimodal Patient Representation Learning with Missing Modalities and Labels. In *The Twelfth International Conference on Learning Representations*, 2024.

[Zadeh *et al.*, 2020] Amir Zadeh, Paul Pu Liang, and Louis-Philippe Morency. Foundations of Multimodal Co-learning. *Information Fusion*, 64:188–193, 2020.