Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Prototypical clustered federated learning for heart rate prediction*

Yongjie YIN[1], Hui RUAN[1], Yang CHEN[1], Jiong CHEN[2], Ziyue LI[3], Xiang SU[4], Yipeng ZHOU[5], Qingyuan GONG[†‡6]

[1]*School of Computer Science, Fudan University, Shanghai 200438, China*
[2]*NIO, Shanghai 201800, China*
[3]*Information System Department, University of Cologne, Cologne, Germany*
[4]*Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland*
[5]*School of Computing, Faculty of Science and Engineering, Macquarie University, Sydney, Australia*
[6]*Research Institute of Intelligent Complex Systems, Fudan University, Shanghai 200433, China*
[†]E-mail: gongqingyuan@fudan.edu.cn

Received Jan. 27, 2025; Revision accepted Apr. 6, 2025; Crosschecked

Unedited

Abstract: Predicting future heart rate (HR) not only helps in detecting abnormal heart rhythms but also provides timely support for downstream health monitoring services. Existing methods for predicting future HRs still encounter challenges, especially concerning privacy protection and data heterogeneity. To address these challenges, this paper proposes a novel HR prediction framework, PCFedH, which leverages personalized federated learning and prototypical contrastive learning to achieve stable clustering results and better predictions. PCFedH contains two core modules: the Prototypical Contrastive Learning-Based Federated Clustering Module, which characterizes data heterogeneity and enhances HR representation for better clustering, and the Two-Phase Soft Clustered Federated Learning Module, which enables personalized performance improvements for each local model based on stable clustering results. Experimental results from two real-world datasets demonstrate the superiority of our approach over state-of-the-art methods, achieving an average reduction of 3.1% in mean square error. Additionally, we conduct comprehensive experiments to empirically validate the effectiveness of the key components in our proposed method. Among these, the personalization component is identified as the most crucial aspect of our design, indicating its significant impact on the performance of our method. The code for this work is available at https://github.com/pcfedh/pcfedhgithub.com/pcfedh/pcfedh.

Key words: Federated learning; Heart rate prediction; Prototypical contrastive learning
https://doi.org/10.1631/FITEE.2500062 **CLC number:**

1 Introduction

Heart rate (HR) stands as a crucial health metric used for various applications, including heart disease detection (Patidar et al., 2015), fatigue monitoring (Patel et al., 2011), and activity recognition (Mehrang et al., 2017). Future HR prediction is significant because it not only aids in detecting HR anomalies through comparison of actual HRs with predicted values but also supports downstream detection services such as heart disease and fatigue monitoring, enabling them to provide more timely warnings before the onset of a disease. Extensive research efforts have been dedicated to estimating HR to align with the ground truth values closely.

‡ Corresponding author
* Project supported by the National Natural Science Foundation of China (Nos. 62102094 and 62072115), the Shanghai Science and Technology Innovation Action Plan Project (No. 22510713600), the NIO University Programme (NIO UP), China, and the Nordic University Cooperation on Edge Intelligence (No. 168043)
ORCID: Qingyuan GONG, https://orcid.org/0000-0001-7942-8752
© Zhejiang University Press 2025

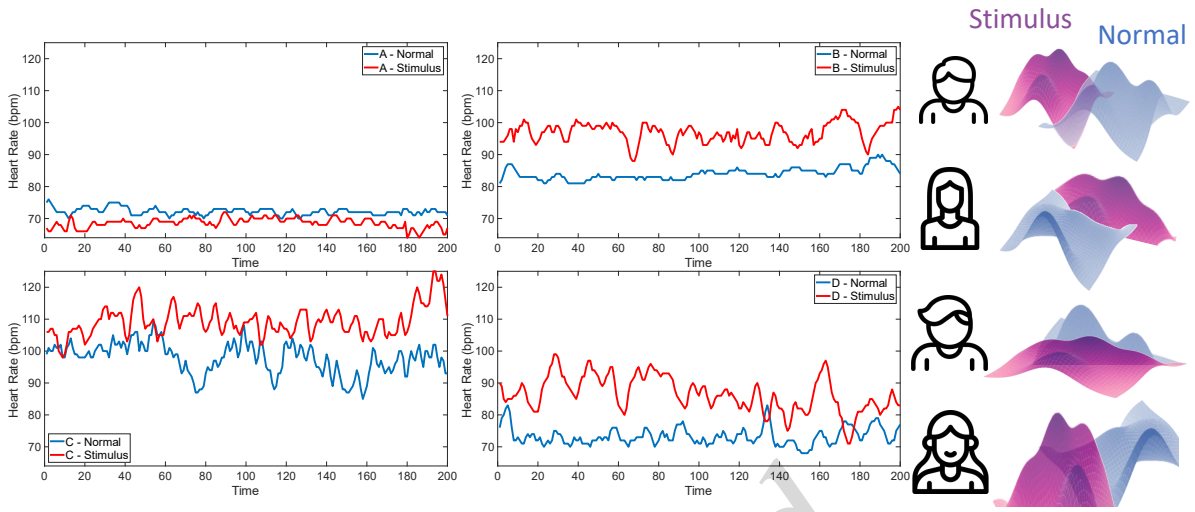


Fig. 1 Different people have different HR (heart rate) patterns and unique pattern differences between normal and stimulus conditions. Different people's two states are in different manifolds.

Unedited

Smart watches (Panwar et al., 2020) and visual methods (Yu et al., 2022) offer more comfortable and non-invasive means of HR estimation compared to alternative health metrics. Consequently, it is feasible and desirable to predict future HR with precise estimation.

Previous studies in this area heavily depend on a traditional training paradigm, which involves aggregating all HR data and processing it on a centralized server. This approach raises significant privacy concerns, particularly due to the sensitive nature of HR data. Protecting such data is crucial, especially in light of the General Data Protection Regulation (GDPR), which mandates stringent safeguards for personal data. Federated learning (FL) represents a promising approach for training machine learning models while safeguarding user privacy. This technique allows models to be trained collaboratively across numerous distributed devices, thereby preserving users' privacy while still achieving accurate predictions. There are several studies that use FL for HR prediction or other cardiovascular health problems (Brisimi et al., 2018; Fang et al., 2020; Yaqoob et al., 2023). They often focus on directly identifying heart diseases, which require a substantial amount of user-specific characteristics.

Non-independent and identically distributed (Non-IID) problem (Zhu et al., 2021) exists in the FL process due to the noticeable differences in data distribution among various clients. It is worth noting the heterogeneity of HR data among individuals. Taking drivers as an example, stressful driving conditions like congested traffic or adverse weather conditions may result in an increase in a driver's HR. This phenomenon is depicted in Figure 1, which shows the HR changes among four drivers during normal driving and stimulated driving. The HR fluctuations under stimulation clearly differ from those during normal driving, and there is also variation in the HR performance among different drivers under the same conditions. Consequently, HR responses can vary markedly among individuals, even in similar activity scenarios, which presents a significant Non-IID problem. This observation underscores the inadequacy of applying a uniform predictive model to all clients.

Personalized FL (Tan et al., 2023) has emerged as the primary method to solve the Non-IID problem of FL by classifying training samples with similar patterns. In this research, we focus on clustered FL, one type of personalized FL that proves more practical for real-world applications. However, existing clustered FL models only regard the clustering process as a tool to improve personalized training effects, ignoring the importance of achieving better clustering outcomes. Accurate clustering is essential because it ensures that each cluster model is tailored to its specific group, thereby improving the overall

performance of cluster prediction models. Furthermore, the clustering process in existing approaches is only beneficial for users who participate in the training process, as it relies on training loss to partition the data, overlooking the users who are temporarily disconnected due to poor network circumstances or high training and transmission costs, which often happen in real scenarios.

To address the above challenges, this paper contributes an HR prediction framework named PCFedH, which integrates the benefits of personalized FL and prototypical contrast learning. PCFedH includes two fundamental modules: a Prototypical Contrastive Learning-Based Federated Clustering Module and a Two-Phase Soft Clustered Federated Learning Module. These modules address data heterogeneity issues in traditional FL from different aspects. The Prototypical Contrastive Learning-Based Federated Clustering Module performs feature extraction and clustering of original HR data to characterize data heterogeneity, which can provide cluster centers for users who did not participate in previous training rounds. The Two-Phase Soft Clustered FL Module overcomes the instability of clustering performance in existing clustered FL methods to further improve the HR prediction performance for each local model. In the first phase, it initializes clustering and conducts precise data classification based on the results of the Prototypical Contrastive Learning-Based Federated Clustering Module. And in the second phase, it conducts personalized training locally through regularization methods and derives multiple global clustered models based on the weight of data types. Our contributions are threefold.

- We present PCFedH, an FL method designed specifically for HR prediction, providing enhanced personalized predictions with practical applicability.
- To address the Non-IID nature of people's HR, we synergize prototypical contrastive FL with clustered FL. This approach not only facilitates more stable and improved clustering within FL processes but also leads to superior personalized training outcomes.
- We evaluate the performance of PCFedH on two real-world datasets, with significant improvements over baselines. Compared with state-of-the-art methods, our method achieves an average reduction of **3.1%** in mean square error, highlighting our method's effectiveness in accurately predicting HR.

2 Related work

2.1 Heart rate prediction

In the field of HR prediction, most work adopts a centralized training paradigm for model training. These studies (Alharbi et al., 2021; Oyeleye et al., 2022; Staffini et al., 2022) based on centralized training have demonstrated the effectiveness of recurrent neural networks for HR prediction, but they require the collection of users' raw data, which poses significant privacy and security concerns. Addressing the critical privacy issue, there are also studies leveraging FL to predict HR or other cardiac health issues. Fang et al. (2020) proposed FL-based HR prediction. It relies on real-time exercise speed, which means it is still a detection work instead of a prediction. Brisimi et al. (2018) developed a federated optimization scheme to train a sparse support vector machine (SVM) to predict hospitalizations due to heart diseases based on past patients' electronic health records. Yaqoob et al. (2023) applied a federated SVM to identify heart failure, drawing on the distinct features of patients' heart diseases. These studies require a variety of user-specific features, and moreover, they do not take into account the fact that the HR of each individual exhibits different characteristics of variation due to factors such as age and differences in behavior and activities. Training a single unified model cannot meet the diverse needs of all users.

2.2 Clustered FL

Clustered FL is one of the core strategies of personalized federated learning (Sattler et al., 2021). It not only champions privacy by sharing only model updates rather than raw data but also allows clients to tailor models to closely align with their unique data landscapes. The advantage of clustered FL lies in its pragmatic applicability, notably allowing clients lacking in computational resources or new participants to straightforwardly select an appropriate model for task execution based on their data's cluster categorization. This approach is particularly

Unedited

suitable for HR prediction tasks with significant privacy concerns and notable data heterogeneity across individuals.

IFCA (Ghosh et al., 2022) represents a further evolution in clustered FL, implementing an iterative algorithm that commences with the initialization of K clusters and finally reaches the corresponding global models. Clients are categorized based on the minimal loss function derived from these global models, with each iteration refining the clustering and model personalization process. **FedSoft** (Ruan and Joe-Wong, 2022) further refines this concept of soft clustering, positing that prior clustered FL models imposed overly stringent constraints on clustering. Acknowledging that client data often span multiple categories, **FedSoft** advocates for a nuanced model aggregation approach to minimize training overhead. To address the high computational costs brought by model inference through all training data, Cai et al. (2023) introduced **FedCE**, which is also based on the concept of soft clustering for personalized training. This method uses the loss observed across multiple batches to represent the features of client data distributions, sidestepping the need to specify exact data proportions. However, this approach presumes that each client predominantly belongs to a single class. This assumption may not align with the reality of HR data, where the distribution of data among various categories can be comparably uniform.

In summary, current research on clustered FL merely treats the clustering process as a mechanism to enhance the effects of personalized training, overlooking the significance of achieving accurate clustering outcomes. Moreover, clustering in existing approaches is only useful for users who participate in the training process, whereas non-participating users lack the capability to accurately derive patterns from HR data.

2.3 Prototype FL

The concept of prototypes stems from the domain of prototypical contrastive learning (Li et al., 2021). The contrastive learning approach suggests that each data sample is associated with a specific cluster, with the prototype acting as the central or "average" feature representation of all the samples in that cluster. During the learning process, a sample paired with its cluster's prototype forms a positive pair, enhancing similarities, while pairing a sample with a prototype from another cluster creates a negative pair, emphasizing differences.

In the realm of FL, the integration of prototypes into learning models has been implemented in various ways (Tan et al., 2022a,b; Mu et al., 2023; Qi et al., 2023), primarily within a supervised learning framework. For example, **FedProto** (Tan et al., 2022a) firstly proposed the paradigm where clients share knowledge among them through their prototypes rather than gradients, and **FedPCL** (Tan et al., 2022b) proposed to utilize both local and global prototypes. These approaches, by focusing on supervised tasks, notably classification, inadvertently narrow the scope of prototypes. This limitation confines their utility, underlining a missed opportunity for broader application in unsupervised or semi-supervised learning scenarios, where prototypes could offer substantial benefits in learning from unlabeled or partially labeled data.

In our work, we bridge this critical gap by combining prototype FL and clustered FL. We employ clustering techniques to generate cluster centers from unlabeled HR data and apply prototype FL to enhance the predictors' abilities. These can not only facilitate the training of high-quality global encoders with low additional communication costs but also expand the application of prototype FL to unsupervised tasks.

3 Method

3.1 Problem definition

Leveraging FL in HR prediction typically involves a central server and N clients within a horizontal FL setting, where these clients possess local HR datasets $\{D_1, D_2, ..., D_N\}$. Here, $D_i = \{X_i^h, Y_i^h\}(h = 1, 2, 3, ..., n_i)$, with n_i being the number of data samples. The central server is responsible for model aggregation, whereas the vehicles, serving as clients, participate in FL to process HR data and train models. The input to the model is a univariate time series of HR measured at L consecutive time points $X = \{x_1, x_2, ..., x_L | x_i \in \mathbb{R}^1\}$. The prediction target of the model is the future HR at the next time point $Y = y_{L+1} \in \mathbb{R}^1$ following the time series X. Each client trains its local prediction model (model parameters denoted by w_i) to tackle this regression problem using its local dataset D_i. We will first

Unedited

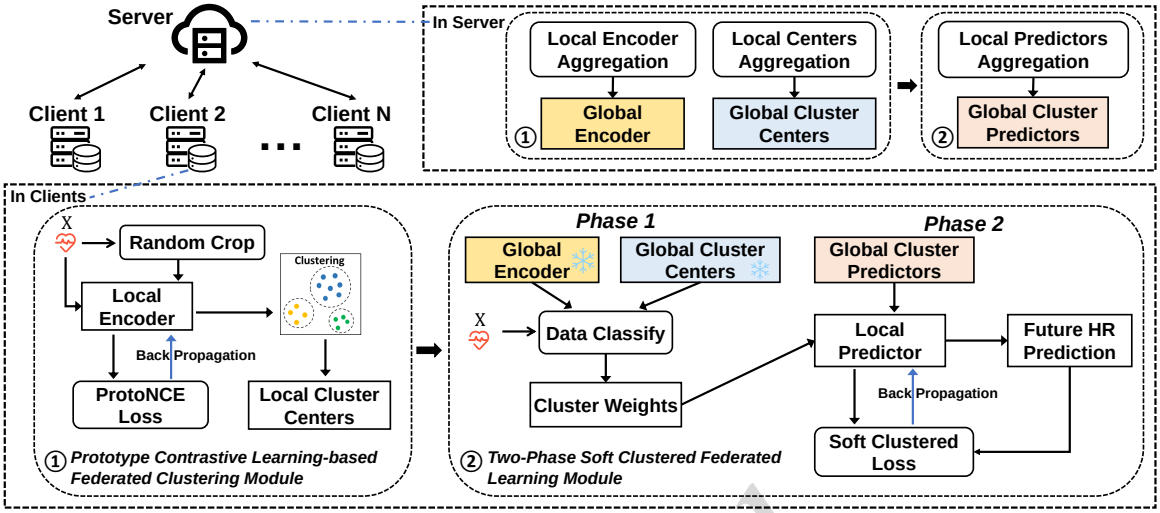


Fig. 2 Architecture of PCFedH. (1) The first module details the process where clients use local data to train the clustering encoder, perform local clustering, and then share the local cluster centers and local model with the server for the aggregation of a global encoder and global cluster centers. (2) The second module describes how clients use the global cluster centers and global encoder to cluster their data, calculate cluster weights in the first phase, and then train their local prediction models with the soft clustered FL approach to improve prediction accuracy on their datasets and contribute to the enhancement of the global cluster prediction models. Note that the second module uses the global encoder and cluster centers pretrained in the first module.

Unedited

introduce the definition of the HR prediction problem under the traditional FL scenario, followed by a further analysis and definition of the problem using clustered FL models.

In the traditional FL setting, the model parameters of the local prediction model are identical before local training begins; they are all downloaded from the server to overwrite the local model, and each local training's target function is identical, $\mathcal{L}_P$. Hence, the overall objective function of FL is given by

$$\min \sum_{i=1}^{N} \frac{n_i}{n} \mathcal{L}_i(w_i, D_i), \tag{1}$$

where $\mathcal{L}_i$ is the local target function of the i-th client and n is the total sum of samples across all clients, satisfying $\mathcal{L}_P = \mathcal{L}_1 = \mathcal{L}_2 = ... = \mathcal{L}_N$ and $w_p = w_1 = w_2 = ... = w_N$. Traditional FL methods synchronize the model to all clients by periodically distributing the global model from the server. These approaches to parameter synchronization may bias the model towards samples in greater quantities, potentially reducing performance on local datasets in the face of data heterogeneity among clients.

Next, we will detail the specific problem definition where the heterogeneity in HR is considered. Given the variability of HR, we assume the local dataset comprises data following various distributions, i.e., each data pair X, Y comes from one of K distributions $P_1, P_2, ..., P_K$. The proportion of different distributions in each client's local dataset varies, making the unified synchronization strategy used in traditional FL inapplicable. Therefore, we adopt the concept of clustered FL to define the problem, aiming to enhance each client's predictive performance through personalized prediction. Unlike traditional FL, which relies on a single global model, clustered FL aims to train global model parameters $w_1^p, w_2^p, ..., w_K^p$ for K categories, each corresponding to one of K distributions $P_1, P_2, ..., P_K$, with w_k^p's model performing best on data following P_K. The training of multiple clustered global models also depends on the knowledge of the distribution of the HR data across clients, which is determined by users' activity conditions and personal states and is often not provided in real-world applications. Thus, in our method, we employ clustering methods to effectively address the challenge of data heterogeneity before the clustered FL module.

In summary, we aim to train HR prediction models through personalized FL, intending to preserve user privacy while utilizing historical HR data to predict future HRs. The solution covers two main

aspects:

1. High-quality clustering of each local HR dataset to characterize the heterogeneity of data across clients.
2. Implementation of cluster personalized FL based on clustering results to address the issue of data heterogeneity.

We achieve personalized training through the concept of clustered FL, requiring the server to update K cluster models $w_k^p (k = 1, 2, 3, \ldots, K)$. Each client's local model w_i is trained using the K cluster models combined with the number of data samples from each category n_{ik} in its local data as auxiliary information. The overall objective function for personalized FL is given by

$$\min \sum_{i=1}^{N} \mathcal{L}_i(w_i, w_k^p, n_{ik}), \tag{2}$$

where $\mathcal{L}_i$ is the local target function of the i-th client, with individual differences in local target functions among clients.

3.2 PCFedH architecture

Figure 2 presents the architecture of PCFedH, primarily comprising two core modules: the Prototypical Contrastive Learning-Based Federated Clustering Module and the Two-Phase Soft Clustered Federated Learning Module. In the process of personalized federated training, the effectiveness of local model updates heavily depends on the data quantity of class k from client i, denoted as n_{ik}, which in turn depends on the effectiveness of data clustering. The Prototypical Contrastive Learning-Based Federated Clustering Module is responsible for achieving clustering in each client while preserving privacy. Each client utilizes local data to train a local cluster representation encoder w_i^e and performs local clustering based on the features extracted by the local encoder to generate cluster centers C_i. The server is responsible for aggregating the local cluster centers to obtain a global cluster center C_g and the local encoders into a global cluster representation encoder w_g^e. This process iterates until the clustering results converge. Section 3.3 introduces the specific design of this module algorithm.

The Two-Phase Soft Clustered Federated Learning Module is responsible for training personalized FL prediction models based on the results of clustering. In the first phase of the Two-Phase Soft Clustered Federated Learning Module, data classification is first performed. Each client uses the final global cluster center C_g and global encoder w_g^e obtained from the Prototypical Contrastive Learning-Based Federated Clustering Module to cluster their data, generating different category weights. These weights refer to the proportion of data belonging to various clustering categories relative to the total dataset, denoted as u_{ik}. In the second phase, utilizing these weights, each client uses personalized FL methods to train their local prediction model w_i. This allows them to achieve improved prediction results within their respective datasets, while also enhancing the performance of global cluster predictor models w_k^p, especially improving the performance of models related to categories that are more prevalent in their data types. Section 3.4 introduces the specific design of this module's algorithm.

Algorithm 1 Prototypical Contrastive Learning-based Federated Clustering

Input: Local heart rate dataset D_i from each client, $i = 1, \ldots, N$. $n_i = |D_i|$. $n = \sum_{i=1}^{N} n_i$.

Server executes:

Initialize global encoder w_g^e and global cluster centers $C_g = \{c_1, c_2, ..., c_K\}$.
for each round $T = 1, 2, \ldots$ **do**
  Send w_g^e and C_g to all clients
  for each client i **in parallel do**
    $w_i^e, C_i \leftarrow \text{ClientUpdate}(D_i, w_g^e, C_g)$
  end for
  $w_g^e \leftarrow \sum_{i=1}^{N} \frac{n_i}{n} w_i^e$ *//Aggregate local encoders*
  $C_g \leftarrow \sum_{i=1}^{N} \frac{1}{N} C_i$ *//Aggregate local cluster centers*
end for

ClientUpdate(D_i, w_g^e, C_g):

for each local epoch $E = 1, 2, \ldots$ **do**
  for each batch b in D_i **do**
    Calculate $\mathcal{L}_{ProtoNCE}$ for b.
    Update encoder: $w_i^e \leftarrow w_g^e - \eta \nabla \mathcal{L}_{ProtoNCE}(b, w_g^e)$
  end for
end for
Perform one step of local $K-means$ clustering to update C_i.
return w_i^e and C_i to the server

3.3 Prototypical contrastive learning-based federated clustering

We design the Prototypical Contrastive Learning-based Federated Clustering Module for feature extraction. This module comprises two parts:

Unedited

prototypical contrastive learning, which constructs positive and negative sample pairs through original representations and cluster centers, and contrastive learning through contextual consistency.

Algorithm 1 details the prototypical contrastive learning-based federated clustering. The server initializes a global cluster center C_g and a global cluster representation encoder model w_g^e (line 1 in the main body of Algorithm 1). In each FL round, each client downloads the latest global encoder from the server and updates its local encoder. Subsequently, the local encoder is trained for a contrastive learning task (lines 1-6 in the ClientUpdate function of Algorithm 1). Initially, representation extraction is performed, converting the original HR time series X_h into an encoded feature vector $z_h \in \mathbb{R}^q$ using the local encoder.

Prototypical contrastive learning is employed to enhance clustering quality. In this approach, the original sample and its corresponding cluster center form a positive pair, while negative samples are formed with other cluster centers. By contrasting each sample with prototypes of its own and other categories, the goal is to maximize intra-class similarity and minimize inter-class similarity (Li et al., 2021). Features obtained through prototypical contrastive learning demonstrate robustness to noise and are naturally suited for clustering tasks, making clustering more precise and meaningful. The loss function for this learning process is as follows:

$$\mathcal{L}_{pcl} = \sum_{h=1}^{n_i} -\log \frac{\exp(z_h \cdot c_p/\phi_p)}{\sum_{k=1}^{K} \exp(z_h \cdot c_k/\phi_k)}, \quad (3)$$

where z is the encoded feature vector, K is the number of cluster centers, and ϕ is a class density value controlling the gap between features and each cluster center, obtained through the clustering density of cluster centers. c_p is the cluster center belonging to the class of z_h, determined by calculating the cosine similarity:

$$p = \underset{k}{\operatorname{argmax}}\, CosSim(z_j, c_k). \quad (4)$$

To ensure a more stable and efficient convergence of the encoder training process, the training also includes contrastive learning through contextual consistency. This method is particularly suited for time-series data, mitigating the impact of outliers and abrupt changes (Yue et al., 2022). The process of constructing positive and negative samples through contextual consistency includes two parts: random cropping and random masking. The random cropping selects two subsequences, denoted as $[a_1, b_1]$ and $[a_2, b_2]$, from an original input sequence of length L. These subsequences contain repeated segments and are chosen randomly to satisfy the condition $0 < a_1 < a_2 < b_1 < b_2 \leq L$. The second part, random masking, involves applying masking operations to the two previously selected subsequences. This process follows a Bernoulli distribution where each temporal dimension variable has an equal probability of being zeroed out. Specifically, there is a 50% chance for each variable to be masked at each time point. By processing the two subsequences in this manner, there is a high likelihood that they possess similar contextual information, and the random masking enhances robustness against abrupt changes and outliers. The feature vectors f_h and $f_h^{'}$ obtained from encoding these two subsequences are treated as positive sample pairs, while the feature vector pairs of other subsequences within the same training batch serve as negative sample pairs. The loss function for contextual consistency contrastive learning, denoted as $\mathcal{L}_{InfoNCE}$, is formulated as follows:

$$\mathcal{L}_{InfoNCE} = \sum_{h=1}^{n_i} -\log \frac{\exp(f_h \cdot f_h^{'})}{\sum_{j=0}^{r} \exp(f_h \cdot f_j^{'})}, \quad (5)$$

Here, $f_h^{'}$ is the positive sample corresponding to f_h, and r is the number of negative samples, which depends on the size of the training batch.

The goal of each local training session is to minimize the combination of two loss terms:

$$\mathcal{L}_{ProtoNCE} = \mathcal{L}_{pcl} + \alpha \mathcal{L}_{InfoNCE}, \quad (6)$$

where α is a weighting coefficient.

After completing local contrastive learning training, each client uses the local encoder for feature extraction on local data, clustering the features using the K-means algorithm (Jain and Dubes, 1988) concept. In the federated clustering process designed in this paper, the clustering iterations and the entire FL iteration process occur simultaneously. Specifically, after each round of FL, the client uses the global cluster center $C_g = \{c_1, c_2, ..., c_K\}$ provided by the server as the initial cluster center rather than randomly initializing it. For the encoded feature z_j of each client's data point x_j after local training, its Euclidean distance $d(z_j, c_k)$ to each cluster center c_k

Unedited

is calculated. The data point x_j is assigned to the cluster with the nearest distance, thus dividing the local data into K clusters:

$$\text{Cluster}(x_j) = \underset{k}{\text{argmin}}\, d(z_j, c_k), \tag{7}$$

where $\text{Cluster}(x_j)$ indicates the category to which x_j belongs. For each cluster k, the average encoded feature of all data points in the cluster is calculated to serve as the new local cluster center:

$$c'_k = \frac{1}{|S_k|} \sum_{z_j \in S_k} z_j, \tag{8}$$

where S_k is the set of data points assigned to the kth cluster, and $|S_k|$ is the number of data points in that set. The key to the federated clustering process is that each client performs clustering of local data and updates the new cluster centers based on the global cluster center (line 7 in the ClientUpdate function of Algorithm 1), significantly speeding up computation while preventing overfitting of cluster centers due to incomplete data categories at the local level. After generating new local cluster centers, clients upload the local cluster centers and local encoders to the server. The server aggregates these local cluster centers and local encoders to update the global cluster center and global encoder (lines 7–8 in the main body of Algorithm 1). The local training and cluster center update process and the server aggregation process iterate until the global encoder achieves satisfactory performance or the predetermined number of FL rounds is reached.

3.4 Two-phase soft clustered FL

We utilize FedSoft (Ruan and Joe-Wong, 2022), a state-of-the-art clustered FL technique, to train the HR prediction model, addressing the Non-IID challenge. FedSoft's uniqueness lies in its capability to efficiently train both clustered and personalized models. Recognizing that a client might possess diverse data types, FedSoft employs a proximal approach to derive multiple cluster models without multiple training processes. It categorizes data based on training loss using distinct clustering methods and employs an expectation-maximization (EM) algorithm to ensure convergence of both clustering and prediction processes. The EM algorithm also has its limitations. It can be plagued by slow convergence and might settle at local maxima instead of global ones. Furthermore, for new users to utilize the cluster model, they must first classify their data, a step not included in FedSoft. To address FedSoft's limitations, we've eliminated the EM process, opting to cluster data initially.

This module effectively addresses the challenge of Non-IID data in FL through precise data classification and personalized training based on category weights. The accurate classification in the first phase provides a stable foundation for subsequent personalized training, which optimizes the model specifically for the data distribution of each client, enhancing the overall model performance and adaptability. The details of the two phases are shown in Algorithm 2, which will be discussed below.

Phase One: Precise Data Classification. This phase addresses the instability in clustering due to the undertrained initial global model. The

Algorithm 2 Two-Phase Soft Clustered Federated Learning for Heart Rate Prediction

Input: The local heart rate datasets of each client $D_i, i = 1, \ldots, N$; a pretrained global encoder w_g^e; cluster centers $C_g = \{c_1, c_2, \ldots, c_K\}$.

Server executes:

// Phase 1: Data Categorization
Broadcast w_g^e and C_g to all clients;
for each client i **in parallel do**
  for each data point X_i^h **do**
    $j = \underset{k}{\text{argmax}} CosSim(c_k, w_g^e(X_i^h))$;
    $n_{ij} = n_{ij} + 1$;
  end for
  Calculate u_{ik} according to equation 9;
  return n_{ik} to the server;
end for
The server calculates v_{ik} according to equation 10;
// Phase 2: Personalized Federated Learning
Initialize global cluster predictors $w_k^p, k = 1, \ldots, K$
for each round $T = 1, 2, \ldots$ **do**
  The server randomly selects K sets of clients $Sel_k \subset [N]$, where $|Sel_k| = M$, each client is selected with probability v_{ik};
  for each selected client i **in parallel do**
    Download $w_k^p, k = 1, \ldots, K$
    $w_i \leftarrow \text{ClientTrainPredict}(D_i, w_k^p, n_{ik})$;
  end for
  Update $w_k^p \leftarrow \frac{1}{M} \sum_{i \in Sel_k} w_i$;
end for

ClientTrainPredict(D_i, w_k^p, n_{ik}):
for each local epoch **do**
  for each batch b in dataset D_i **do**
    Compute $\mathcal{L}_i(b, w_i, w_k^p, n_{ik})$;
    Update local predictor: $w_i \leftarrow w_i - \eta \nabla \mathcal{L}_i$;
  end for
end for
return w_i to the server;

Uneditable

focus of this phase is to establish a stable and reliable data classification foundation for each client, setting the stage for personalized model training. We utilize the global cluster representation encoder and cluster centers from the Prototypical Contrastive Learning-Based Federated Clustering Module for accurate data classification. By calculating the cosine similarity between each data representation and cluster centers, we compute the number of data points n_{ik} belonging to every cluster k for client i. Subsequently, on each client, the proportion of each type of data is used as the weight for different categories:

$$u_{ik} = \frac{n_{ik}}{n_i}, \tag{9}$$

With the n_{ik} belong to every cluster k, the server computes the aggregation weights v_{ik} for client i as follows:

$$v_{ik} = \frac{n_{ik}}{\sum_i^N n_{ik}} \tag{10}$$

Phase Two: Personalized FL. Building on the classification outcomes from the first phase, the second phase focuses on enhancing the adaptability and prediction accuracy of each client model towards its local data through personalization. To not only simulate real-world scenarios but also reduce the high costs of training and transmission by allowing only a subset of clients to participate in each training round, the server initially selects K groups of clients, $Sel_k \subset \{1, 2, 3, ..., N\}$, based on the probability v_{ik} (line 15 in the main body of Algorithm 2), where $|Sel_k| = M < N$, and M is a predefined hyperparameter denoting the number of clients in each group. This means the M clients in Sel_k are likely to contain more class k data, and these K groups of clients correspond to the aggregation update of K global cluster predictors.

Each client downloads the global clustering prediction model w_k^p and adjusts the regularization term in the loss function based on the weight of the data category. The proximal update of each client's local prediction model is as follows:

$$\mathcal{L}_P(w_i) = \frac{1}{n_i} \sum_{h=1}^{n_i} (w_i(X_i^h) - Y_i^h)^2 \tag{11}$$

$$\mathcal{L}_i(w_i) = \mathcal{L}_P(w_i) + \frac{\lambda}{2} \sum_{k=1}^{K} u_{ik} ||w_i - w_k^p||^2, \tag{12}$$

where w_i represents the parameters of the local prediction model, and K is the number of cluster centers. $\mathcal{L}_P(w_i)$ in Eq. (12) represents the primary loss function, which in this case is the mean squared error (MSE). The second term $\frac{\lambda}{2} \sum_{k=1}^{K} u_{ik} ||w_i - w_k^p||^2$ is a regularization term, adjusted according to each client's data category distribution u_{ik}, calculating the distance between the model parameters w_i and each global cluster model parameter w_k^p. This mechanism not only retains the personalized features of the local prediction model but also ensures the model is as close as possible to the model corresponding to the dominant category in the local dataset. λ is a regularization parameter that controls the balance between the two terms.

After training, clients upload their local models to the server. The server aggregates each global cluster predictor model for updates. The global cluster model w_k^p is updated through aggregation solely based on the local prediction models of the clients in the group Sel_k, selected by the probability based on the category proportion (line 20 in the main body of Algorithm 2). In this manner, each client's local prediction model is tailored to its specific dataset and encouraged to align closer to the global class prediction model corresponding to the dominant category in the client's data.

3.5 Computational complexity analysis

In this section, we compare our method with other FL methods mentioned in Table 2 in terms of computational complexity, including classical FL methods (e.g., FedAvg (McMahan et al., 2017), FedProx (Li et al., 2020), FedALA (Zhang et al., 2023)), clustered FL methods (e.g., IFCA (Ghosh et al., 2022), FedSoft (Ruan and Joe-Wong, 2022), FedCE (Cai et al., 2023)), and prototype FL methods (e.g., FedProto (Tan et al., 2022a), FedPCL (Tan et al., 2022b)).

Here, we focus on the time complexity of one training epoch on the client side, as the computational resources of the central server are typically sufficient. The results are presented in Table 1.

In this comparison, we consider the computational costs of all operations and omit the specific prediction or classification model employed on the client side. Some basic notations are utilized: d as the dimension of models, B as the sample size, K as cluster count, m as client number, and p as a special hyperparameter of FedALA. From Table 1, it is shown that the time complexity of PCFedH, our proposed FL framework, is similar to and even

Unedited

Methods	FedAvg	FedProx	FedALA	IFCA	FedSoft	FedCE	FedPCL	FedProto	PCFedH(M1)	PCFedH(M2)
Time	$\mathcal{O}$(Bd)	$\mathcal{O}$(Bd)	$\mathcal{O}$(B(d+p))	$\mathcal{O}$(BKd)	$\mathcal{O}$(BKd)	$\mathcal{O}$(BKdlog K)	$\mathcal{O}$(BKdm)	$\mathcal{O}$(BKd)	$\mathcal{O}$(BKd)	$\mathcal{O}$((B+K)d)

Table 1 Computational complexity comparison among our methods and baselines, with notation defined as follows: B (dataset size), d (model dimension), K (cluster count), m (client count), p (a special hyperparameter of FedALA), and M1/M2 (our framework's two core modules)

greater than that of other state-of-the-art (SOTA) cluster and prototype FL methods. This indicates that the computational cost of PCFedH on client sides is acceptable. Meanwhile, PCFedH exhibits higher time complexity than classical FL methods (e.g., FedAvg), which rely on simpler client-side operations. This increase stems from its additional components (e.g., prototype contrast and clustering) designed to enhance highly personalized prediction performance and robustness in Non-IID scenarios. These advantages are empirically validated in Section 5.

4 Convergence analysis

In this section, we provide theoretical guarantees of PCFedH's convergence. As two modules are trained sequentially and independently in our work, we prove their convergence separately in two parts. Specifically, we first proved the convergence of Module 1, where the upper bound of its gradient tends to zero as the model trains. Secondly, we analyzed the convergence of Module 2, where the gap between its global loss and optimal loss tends steadily to 0. For clear illustration and easy writing, we inherit the notation system from Algorithms 1 and 2.

4.1 Convergence of Module 1

We provide insights into the convergence analysis of PCFedH, as demonstrated in Theorem 1. A detailed description and derivations can be found in the Supplementary.

Theorem 1 Combining basic assumptions of smoothness, bounded variance, and dissimilarity, for a sufficiently small learning rate $\eta_1 \leq \sqrt{\frac{N}{TE}}$, it can be proven that parameters of Module 1 are convergent along the training process:

$$\min_{t\in[T]} \mathbb{E}\left[\|\nabla F(\boldsymbol{w}_{t,0}, \boldsymbol{C}_{t,0})\|^2\right] \leq \mathcal{O}\left(\frac{1}{\sqrt{NET}}\right) + \mathcal{O}\left(\frac{EL\sigma^2}{n^2}\sqrt{\frac{NE}{T}}\sum_{i=1}^{m} n_i^2\right) + \mathcal{O}\left(\left(\sigma^2 + \kappa^2 E\right)\frac{NEL^2}{T}\right)$$

This indicates that Module 1 is convergent along the training process, which can be proven in detail in the Supplementary.

4.2 Convergence of Module 2

Module 2 is designed to train efficient personalized local heart-rate predictors based on representative prototypes obtained from Module 1. The framework of Module 2 is similar to FedSoft, whose convergence was completely proven by Ruan and Joe-Wong (2022). Structurally differing from FedSoft in major, Module 2 performs data classification once based on reliable representative prototypes to reduce computational costs. It is clear that these differences cannot change the convergence of the framework.

5 Experiments

5.1 Experiment settings

5.1.1 Data preparation

We evaluate the performance of PCFedH based on two representative HR datasets.

- PPG-DaLiA (Reiss et al., 2019): A real-world dataset containing HR data from the daily life activities of 15 subjects. It has been used in several research studies, including one by Xu et al. (2022). We use the data to generate 15 local HR data.

- Distracted (Taamneh et al., 2017): A real-world dataset from 68 volunteers (63 effective) in a simulated driving environment. It has been used in several research studies, including one by Chen et al. (2021). We use data from both normal and distracted driving conditions to generate 63 local HR data.

Different HR series in the dataset come from distinct individuals, who are configured across different clients in our work. The configuration prevents the sharing of data from the same individual

Unedited

Table 2 Performance comparison of PCFedH and baseline methods, with mean and standard deviation reported

Method \ Dataset	PPG-DaLiA			Distracted		
	MSE	MAE	MAPE	MSE	MAE	MAPE
FedAvg	2.621 ± 2.060	1.121 ± 0.398	1.445 ± 0.583%	20.242 ± 153.889	1.466 ± 4.076	1.486 ± 1.804%
FedProx	2.597 ± 1.784	1.126 ± 0.388	1.441 ± 0.553%	1.665 ± 1.209	0.929 ± 0.393	1.202 ± 0.519%
FedALA	2.615 ± 2.029	1.121 ± 0.396	1.445 ± 0.581%	46.470 ± 13.783	6.610 ± 0.898	8.819 ± 2.525%
IFCA	2.624 ± 2.008	1.125 ± 0.393	1.449 ± 0.575%	19.868 ± 151.646	1.404 ± 4.018	1.402 ± 1.773%
FedSoft	2.737 ± 1.861	1.155 ± 0.394	1.478 ± 0.556%	1.671 ± 1.214	0.930 ± 0.395	1.205 ± 0.522%
FedCE	2.610 ± 1.987	1.121 ± 0.391	1.443 ± 0.573%	18.649 ± 140.433	1.473 ± 3.870	1.532 ± 1.716%
FedProto	23.133 ± 19.097	3.559 ± 1.552	4.445 ± 1.658%	99.373 ± 450.702	4.503 ± 7.050	4.715 ± 3.565%
FedPCL	11.239 ± 24.140	2.410 ± 2.330	3.028 ± 2.996%	38.898 ± 244.490	3.532 ± 5.141	4.144 ± 3.996%
PCFedH	**2.549** ± 1.736	**1.117** ± 0.383	**1.431** ± 0.549%	**1.593** ± 1.175	**0.903** ± 0.398	**1.170** ± 0.526%

MAE, mean absolute error; MAPE, mean absolute percentage error; MSE, mean squared error. It shows PCFedH's superior results in reducing MSE, MAE, and MAPE for the PPG-DaLiA and Distracted datasets. The best result is highlighted in boldface, and the second-best result is underlined

Unedited

across clients. Thus, the heterogeneity of HR data among individuals can be maintained. Additionally, we set the input HR series length to 150, where each continuous subseries is treated as an input, with the subsequent time scale data serving as the prediction target. 80% input-output pairs are used for the training phase, while the remaining 20% pairs are used for the testing phase.

5.1.2 Implementation details

Following Yue et al. (2022), the encoder in the Prototypical Contrastive Learning-Based Federated Clustering Module is a TS2Vec model, which consists of an input projection layer that maps time series observations to high-dimensional latent vectors, a timestamp masking module for creating augmented context views by randomly masking selected timestamps, and a dilated convolutional neural network module with 10 residual blocks to extract contextual representations at each timestamp. The prediction model in the Two-phase Soft Clustered Federated Learning Module consists of a single-layer long short-term memory module followed by a fully connected layer.

We select Adam (Kingma and Ba, 2015) as the optimizer with a learning rate of 0.01. The local batch size is set to 16. The local update epochs are $E = 5$. The contrastive learning loss function uses a weight coefficient $\alpha = 1$ and regularization parameter $\lambda = 0.001$. Regarding the clustering configuration, this paper sets the number of classes for the PPG-DaLiA dataset to $K = 4$, while for the Distracted dataset, the number of classes is set to $K = 5$. Additionally, the number of selected clients for the PPG-DaLiA dataset is set to $M = 5$, and for the Distracted dataset, the number of selected clients is set to $M = 20$. The round of the Prototypical Contrastive Learning-Based Federated Clustering Module is 10, and the round of the Two-Phase Soft Clustered Federated Learning Module is 500. We implement all the methods using PyTorch and conduct all experiments on two NVIDIA Quadro P5000 GPUs. We use the following three metrics for evaluation: MSE, mean absolute error (MAE), and mean absolute percentage error (MAPE), which are widely used across various fields (Rather et al., 2015; Yao et al., 2019).

5.2 Experimental results

5.2.1 Overall performance

We evaluate the performance of PCFedH on the HR prediction task by comparing it against several FL methods in the main experiment. The traditional FL baseline methods include FedAvg, FedProx, and FedALA. As the foundational framework, FedAvg was the first to create the method of combining model update gradients across different devices. Building upon this, FedProx introduced a special proximal term to tackle the statistical heterogeneity inherent in FL. FedALA further advanced the paradigm through adaptively aggregating the global model and local model towards the local objective on each client. Moreover, we compare our method with several state-of-the-art personalized FL methods, including cluster FL and prototype FL, which

Table 3 Ablation Study: SoftCluster is the most important design

Method	MSE	MAE	MAPE
PCFedH (w/o PCL)	2.553 ± 1.729	1.118 ± 0.381	1.431 ± 0.542%
PCFedH (w/o CRE)	2.565 ± 1.762	1.119 ± 0.381	1.432 ± 0.543%
PCFedH (w/o DCP)	2.737 ± 1.861	1.155 ± 0.394	1.478 ± 0.556%
PCFedH (w/o SC)	2.785 ± 1.957	1.165 ± 0.379	1.490 ± 0.546%
PCFedH	**2.549 ± 1.736**	**1.117 ±0.383**	**1.431 ± 0.549%**

MAE, mean absolute error; MAPE, mean absolute percentage error; MSE, mean squared error.

Table 4 Performance of PCFedH at varying PCL round

Round	MSE	MAE	MAPE
$R_{pcl} = 1$	2.562 ± 1.730	1.122 ± 0.385	1.437 ± 0.552%
$R_{pcl} = 5$	2.548 ± 1.745	1.119 ± 0.389	1.433 ± 0.560%
$R_{pcl} = 10$	2.549 ± 1.736	1.117 ± 0.383	1.431 ± 0.549%
$R_{pcl} = 50$	2.565 ± 1.762	1.118 ± 0.381	1.430 ± 0.542%
$R_{pcl} = 100$	2.549 ± 1.740	1.116 ± 0.381	1.428 ± 0.544%

MAE, mean absolute error; MAPE, mean absolute percentage error; MSE, mean squared error.

are introduced in Section 2.

Table 2 presents the experiment results. The proposed method PCFedH outperforms all the baselines with an average 3.1% MSE reduction, 1.6% MAE, and 1.7% MAPE reduction, demonstrating its superior ability to predict HR while preserving privacy. Additionally, through comparisons of various methods, we have reached the following two conclusions: (1) **The robustness of regularized local updates**. Considering that FedSoft, FedProx, and PCFedH all employ regularization strategies for local updates, we can observe that methods based on regularized updates perform well in both datasets. This indicates that for HR time series data, regularized updates, as opposed to overwriting updates, have better robustness. (2) **The importance of high-quality clustering for personalized clustered FL**. Clustered FL aims to enhance the model's personalization capabilities by grouping similar clients. The effectiveness of clustering directly affects the quality of personalization. Poor clustering results may lead to the model's inability to correctly identify and adapt to the specificity of client data, thus limiting the model's personalized performance. As shown by the performance of FedSoft and IFCA, merely applying clustering algorithms does not guarantee performance improvement. Similarly, approaches like FedProto and FedPCL, which merely consider prototypical sharing without a clustering algorithm, fail to yield accurate predictive outcomes. In contrast, the excellent performance of PCFedH on two datasets emphasizes the importance of an efficient clustering mechanism for improving the personalization capabilities of FL systems. This indicates that PCFedH can more effectively recognize and utilize differences among clients, enhancing the model's personalized adaptability through precise clustering and successfully expanding the prototype technology into unsupervised time-series pattern classification.

5.2.2 Ablation study

In this section, we present the results of ablation experiments to test the effectiveness of the four key components: Prototypical Contrastive Learning, Cluster Representation Encoder, the Data Classification Phase, and Soft Clustered FL. The ablation results are shown in Table 3.

Prototypical Contrastive Learning (PCL): To underscore the significance of prototypical contrastive learning, we consider PCFedH (w/o PCL), a variant of PCFedH that only uses the sub-series contrastive learning to train the representation encoder. The comparison highlights the essential role of prototypical contrastive learning, which significantly improves the prediction ability of the model. The enhancement stems from the ability of PCL to enhance the alignment of data with its cluster center, benefiting clustering outcomes and consequently improving prediction performance.

Cluster Representation Encoder (CRE): The importance of the cluster representation encoder is demonstrated by comparing it to PCFedH (w/o CRE), a variant of PCFedH that employs only original time-series data for clustering. This comparison validates the necessity of the cluster representation encoder, as it can learn deeper representations that yield improved clustering results.

Data Classification Phase (DCP): To illustrate the significance of the data classification phase, we examine PCFedH (w/o DCP), a variant of PCFedH that is just the same as FedSoft. The noticeable decline in performance with this variant indicates that relying solely on the EM process for clustering limits the efficacy of both clustering and prediction outcomes. This finding validates the importance of the data classification phase, which provides superior clustering performance and enhances the results of clustered FL.

Soft Clustered FL (SC): The value of soft

Unedited

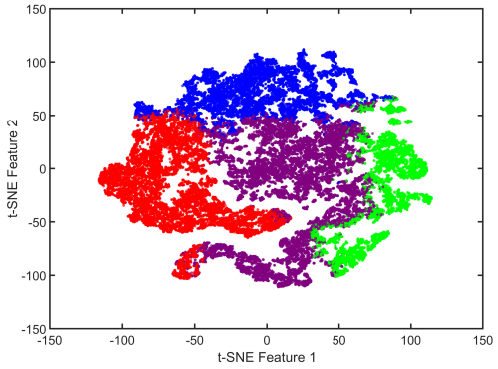
clustered FL is demonstrated through a comparison with PCFedH (w/o SC), a variant of PCFedH that omits the SoftCluster component. In this variant, data are initially segmented and categorized using the fast prototypical federated contrastive learning module. Unique global models are independently trained for each data type following this categorization. The marked decline in this variant's performance suggests that different data types can still mutually benefit each other. This highlights the necessity of soft clustered FL, as it allows clients to learn personalized local models without the constraints of hard clustering.

5.2.3 Hyperparameter sensitivity analysis

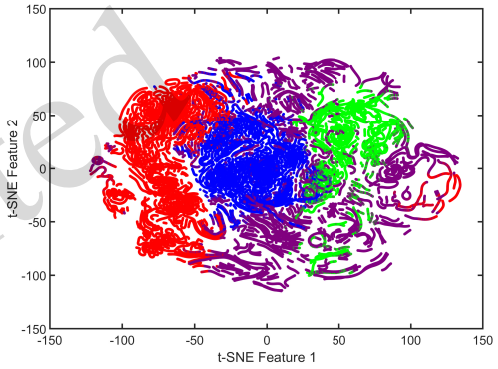
We investigated the sensitivity of the learning round R_{pcl} within the Two-Phase Soft Clustered Federated Learning Module, with our findings detailed in Table 4. By selecting an optimal value for R_{pcl}, we can achieve a balance between performance and computational cost. The results indicate that setting R_{pcl} to 10 yields satisfactory performance. Considering the negligible gains observed when increasing R_{pcl} from 10 to 100, we opted to set $R_{pcl} = 10$ for PCFedH to maintain efficiency without sacrificing accuracy. **Takeaway:** Setting $R_{pcl} = 10$ in the Two-Phase Soft Clustered Federated Learning Module achieves satisfactory performance, maintaining efficiency without significant sacrifice in accuracy for PCFedH.

5.2.4 Effectiveness of cluster representation

To illustrate the effectiveness of our cluster representation, we employed t-SNE (Maaten and Hinton, 2008), a technique primarily used for feature dimensionality reduction to simplify cluster identification. This method was applied to visualize the representation of PPG-DaLiA dataset, as demonstrated in Figure 3. We show the visualizations of the representations obtained by PCFedH's enhanced clustering and the original data. The visualization indicates that the application of prototypical cluster learning significantly enhances the clarity of clustering. **Takeaway:** Employing prototypical cluster learning enhances clustering clarity and stability, surpassing traditional methods such as FedSoft clustering.



(a)



(b)

Fig. 3 Visualizations of PPG-DaLiA dataset representations: (a) demonstrating the enhanced representation by PCFedH and (b) showing the original representation. Points of different colors represent different classes. The visualizations highlight how prototypical cluster learning remarkably improves clustering clarity.

5.2.5 Effectiveness of personalization improvement

In Figure 4, as a case, we show the predictive outcomes for client 3, highlighting the differential performance of PCFedH and FedProx. Client 3's dataset encompasses four sequential driving segments, characterized by HRs that surpass the average of the general population. PCFedH consistently delivers predictions that more accurately reflect the ground truth, attributable to its ability to generate higher predictive values that accommodate the clients' elevated HR levels. Conversely, FedProx's predictions fall short, gravitating towards the mean levels of the broader population due to its lack of personalized training, thereby yielding less precise predictive values. Moreover, differential performance between PCFedH and FedProx among all clients is

Unedited

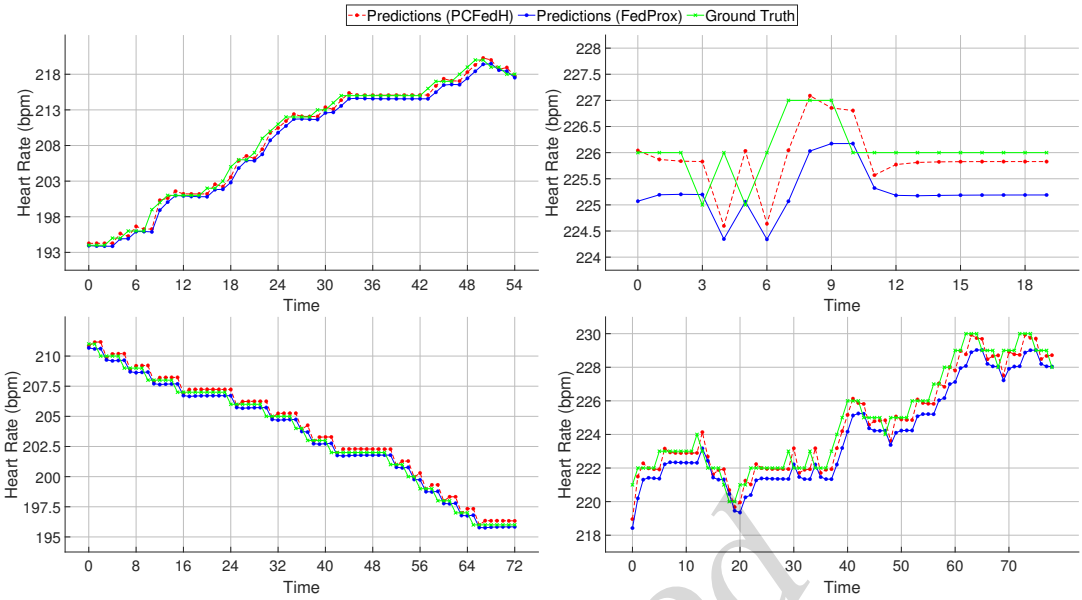


Fig. 4 Comparative analysis of HR predictions during consecutive Driving Segments. This set of four subgraphs depicts the predictive performance of two methods, PCFedH and FedProx, against the ground truth for client 3's HR over time. The PCFedH predictions are consistently closer to the actual HR values, particularly in higher ranges, suggesting a better personalized fit for this client's data compared to the more generalized predictions of FedProx. HR, heart rate.

Unedited

shown in the Supplementary. **Takeaway:** PCFedH consistently improves performance across clients in the Distracted dataset compared to FedProx. Visualizations of predictive outcomes for client 3 highlight PCFedH's superior accuracy in reflecting individual characteristics, especially in scenarios with elevated HRs.

6 Conclusion and future work

In our work, we contribute a novel federated learning framework, PCFedH, specifically designed for predicting heart rate. Addressing critical privacy concerns related to the collection of HR data, PCFedH efficiently harnesses heterogeneity in HRs to provide tailored predictions. Different from classical clustering FL, by integrating prototypical contrastive learning with clustered FL and applying the contextual consistency data augmentation technology, our method accurately classifies user data into distinct categories. This enables more precise and personalized HR predictions while ensuring data privacy, even in Non-IID scenarios with statistical heterogeneity among users. At the same time, prototype learning is used to aggregate prototypes among users to ease local data bias. Through rigorous experimentation with two real-world datasets, our approach has demonstrated a significant advancement over existing state-of-the-art methods, achieving an average reduction of **3.1%** in mean square error. Furthermore, our comprehensive experiments have empirically validated the effectiveness of the key components of PCFedH.

Our work has two main limitations. Firstly, the Prototypical Contrastive Learning-Based Federated Clustering Module requires predefined clustering types K, which may need to be dynamically adjusted as the diversity of activity conditions expands in the FL process. This could potentially affect the adaptability of our model. Secondly, our current experiments are limited to single-step predictions. Improving the ability to predict future HR trends could provide a more comprehensive safeguard for the clients' health by anticipating future HR trends more effectively.

For future work, we plan to enhance the client selection mechanism in FL, aiming to involve fewer clients in each training round to reduce costs effectively. Additionally, we aim to develop methods for dynamically adjusting the pre-defined clustering types K in real-time to maintain model adaptability during Federated Clustering. Furthermore, we

intend to extend our prediction capabilities beyond single steps to anticipate future HR trends more effectively, thereby enhancing health monitoring in FL settings.

Contributors

All the authors contributed to the heart rate prediction study. Yongjie YIN, Hui RUAN, Yang CHEN, Jiong CHEN, and Qingyuan GONG proposed the motivation of the study. Ziyue LI directed the design of PCFedH framework. Yongjie YIN and Hui RUAN performed the experiments and drafted the paper. Yipeng ZHOU and Xiang SU guided the convergence proof of PCFedH. All the authors commented on previous versions of the paper. All the authors read and approved the final version of the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

Unedited

References

Alharbi A, Alosaimi W, Sahal R, et al., 2021. Real-Time System Prediction for Heart Rate Using Deep Learning and Stream Processing Platforms. *Complexity*, 2021:5535734. https://doi.org/10.1155/2021/5535734

Brisimi TS, Chen RD, Mela T, et al., 2018. Federated learning of predictive models from federated electronic health records. *Int J Med Inform*, 112:59-67. https://doi.org/10.1016/j.ijmedinf.2018.01.007

Cai LX, Chen NY, Cao YZH, et al., 2023. FedCE: Personalized federated learning method based on clustering ensembles. Proc 31^{st} ACM Int Conf on Multimedia, p.1625-1633, https://doi.org/10.1145/3581783.3612217

Chen J, Jiang YN, Huang ZX, et al., 2021. Fine-grained detection of driver distraction based on neural architecture search. *IEEE Trans Intell Transport Syst*, 22(9):5783-5801. https://doi.org/10.1109/TITS.2021.3055545

Fang L, Liu XL, Su X, et al., 2020. Bayesian inference federated learning for heart rate prediction. Proc 9^{th} EAI Int Conf on Wireless Mobile Communication and Healthcare, p.116-130, https://doi.org/10.1007/978-3-030-70569-5_8

Ghosh A, Chung J, Yin D, et al., 2022. An Efficient Framework for Clustered Federated Learning. *IEEE Trans Inform Theory*, 68(12):8076-8091. https://doi.org/10.1109/TIT.2022.3192506

Jain AK, Dubes RC, 1988. Algorithms for clustering data. Prentice-Hall, Inc., Saddle River, USA.

Kingma DP, Ba LJ, 2015. Adam: A method for stochastic optimization. Proceedings of the 3^{rd} International Conference on Learning Representations, p.1-15.

Li JN, Zhou P, Xiong CM, et al., 2021. Prototypical Contrastive Learning of Unsupervised Representations, https://arxiv.org/abs/2005.04966

Li T, Sahu AK, Zaheer M, et al., 2020. Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems*, 2:429-450.

van der Maaten L, Hinton G, 2008. Visualizing Data using t-SNE. *J Mach Learn Res*, 9(86):2579-2605.

McMahan B, Moore E, Ramage D, et al., 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, p.1273-1282.

Mehrang S, Pietila J, Tolonen J, et al., 2017. Human Activity Recognition Using A Single Optical Heart Rate Monitoring Wristband Equipped with Triaxial Accelerometer. European Medical and Biological Engineering Conf, p.587-590, https://doi.org/10.1007/978-981-10-5122-7_147

Mu XT, Shen YL, Cheng K, et al., 2023. FedProc: Prototypical contrastive federated learning on non-iid data. *Future Gener Comp Syst*, 143:93-104. https://doi.org/10.1016/j.future.2023.01.019

Oyeleye M, Chen TH, Titarenko S, et al., 2022. A predictive analysis of heart rates using machine learning techniques. *Int J Environ Res Public Health*, 19(4):2417. https://doi.org/10.3390/ijerph19042417

Panwar M, Gautam A, Biswas D, et al., 2020. Pp-net: A deep learning framework for ppg-based blood pressure and heart rate estimation. *IEEE Sensors J*, 20(17):10000-10011. https://doi.org/10.1109/JSEN.2020.2990864

Patel M, Lal SKL, Kavanagh D, et al., 2011. Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Syst Appl*, 38(6):7235-7242. https://doi.org/10.1016/j.eswa.2010.12.028

Patidar S, Pachori RB, Acharya UR, 2015. Automated diagnosis of coronary artery disease using tunable-q wavelet transform applied on heart rate signals. *Knowl-Based Syst*, 82:1-10. https://doi.org/10.1016/j.knosys.2015.02.011

Qi Z, Meng L, Chen ZT, et al., 2023. Cross-silo prototypical calibration for federated learning with non-iid data. Proc 31^{st} ACM Int Conf on Multimedia, p.3099-3107, https://doi.org/10.1145/3581783.3612481

Rather AM, Agarwal A, Sastry VN, 2015. Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Syst Appl*, 42(6):3234-3241. https://doi.org/10.1016/j.eswa.2014.12.003

Reiss A, Indlekofer I, Schmidt P, et al., 2019. Deep ppg: Large-scale heart rate estimation with convolutional neural networks. *Sensors*, 19(14):3079. https://doi.org/10.3390/s19143079

Ruan YC, Joe-Wong C, 2022. FedSoft: Soft Clustered Federated Learning with Proximal Local Updating. Proc 36^{th} AAAI Conf on Artificial Intelligence, p.8124-8131, https://doi.org/10.1609/aaai.v36i7.20785

Sattler F, Müller KR, Samek W, 2021. Clustered Federated Learning: Model-Agnostic Distributed Multitask Optimization Under Privacy Constraints. *IEEE Trans Neural Netw Learning Syst*, 32(8):3710-3722. https://doi.org/10.1109/TNNLS.2020.3015958

Staffini A, Svensson T, Chung UI, et al., 2022. Heart rate modeling and prediction using autoregressive models and deep learning. *Sensors*, 22(1):34. https://doi.org/10.3390/s22010034

Taamneh S, Tsiamyrtzis P, Dcosta M, et al., 2017. A multimodal dataset for various forms of distracted driving. *Sci Data*, 4(1):170110. https://doi.org/10.1038/sdata.2017.110

Tan AZ, Yu H, Cui LZ, et al., 2023. Towards Personalized Federated Learning. *IEEE Trans Neural Netw Learning Syst*, 34(12):9587-9603. https://doi.org/10.1109/TNNLS.2022.3160699

Tan Y, Long GD, Liu L, et al., 2022a. Fedproto: Federated prototype learning across heterogeneous clients. Proc 36^{th} AAAI Conf on Artificial Intelligence, p.8432-8440, https://doi.org/10.1609/aaai.v36i8.20819

Tan Y, Long GD, Ma J, et al., 2022b. Federated learning from pre-trained models: A contrastive learning approach. Proc 36^{th} Int Conf on Neural Information Processing System, p.19332-19344.

Xu MA, Moreno A, Nagesh S, et al., 2022. Pulseimpute: A novel benchmark task for pulsative physiological signal imputation. Proc 36^{th} Conf on Neural Information Processing Systems, p.26874-26888.

Yao HX, Tang XF, Wei H, et al., 2019. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. Proc 33^{th} AAAI Conf on Artificial Intelligence, p.5668-5675, https://doi.org/10.1609/aaai.v33i01.33015668

Yaqoob MM, Nazir M, Khan MA, et al., 2023. Hybrid Classifier-Based Federated Learning in Health Service Providers for Cardiovascular Disease Prediction. *Appl Sci*, 13(3):1911. https://doi.org/10.3390/app13031911

Yu ZT, Shen YM, Shi JG, et al., 2022. Physformer: Facial video-based physiological measurement with temporal difference transformer. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.4176-4186, https://doi.org/10.1109/CVPR52688.2022.00415

Yue ZH, Wang YJ, Duan JY, et al., 2022. Ts2vec: Towards universal representation of time series. Proc 36^{th} AAAI Conf on Artificial Intelligence, p.8980-8987, https://doi.org/10.1609/aaai.v36i8.20881

Zhang J, Hua Y, Wang H, et al., 2023. FedALA: Adaptive Local Aggregation for Personalized Federated Learning. Proceedings of the AAAI Conference on Artificial Intelligence, p.11237-11244.

Zhu HY, Xu JJ, Liu SQ, et al., 2021. Federated learning on non-iid data: A survey. *Neurocomputing*, 465:371-390. https://doi.org/10.1016/j.neucom.2021.07.098

Unedited