Information Fusion 117 (2025) 102812

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

Contrastive learning-based multi-view clustering for incomplete multivariate time series

Yurui Li [a], Mingjing Du [a,*], Xiang Jiang [a], Nan Zhang [b]

[a] Jiangsu Key Laboratory of Educational Intelligent Technology, School of Computer Science and Technology, Jiangsu Normal University, Xuzhou, 221116, China
[b] College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou, 325035, China

ARTICLE INFO

Keywords:
Multi-view clustering
Multivariate time series
Missing data
Contrastive learning
Deep learning

ABSTRACT

Incomplete multivariate time series (MTS) clustering is a prevalent research topic in time series analysis, aimed at partitioning MTS containing missing data into distinct clusters. Contrastive learning-based multi-view clustering methods are a promising approach to address this issue. However, existing methods are typically not designed for time series. Specifically, most of these methods struggle to capture the inherent properties of time series, and are susceptible to losing their interdimensional correlations, thereby compromising data integrity. Furthermore, they commonly utilize data augmentation techniques to generate sample pairs for contrastive learning. These existing data augmentation techniques are not suitable for time series, and introduce uncertainty factors, which can diminish the representation learning capacity of contrastive learning. To address the challenges, we propose a contrastive learning-based multi-view clustering method for incomplete multivariate time series (MVCIMTS). In this method, each variable within the MTS is treated as a separate view, enabling a multi-view learning approach. To better leverage the intrinsic information of time series, we utilize a GRU-based model architecture that integrates imputation and clustering within a unified deep learning framework. In this way, missing views can be effectively inferred, and representations suitable for clustering can be learned, thereby enhancing the clustering performance for incomplete time series. Furthermore, we introduce an innovative contrastive learning approach specifically tailored for MTS, which ensures that the exploration of common semantics and clustering consistency across views remains unaffected by uncertainty factors. It assumes that each time series variable within the same sample has similar representations, thereby taking into account the correlation between variables and enhancing the quality of the representations. To the best of our knowledge, this is the first attempt at applying contrastive learning-based multi-view deep clustering to incomplete MTS. We conduct extensive comparative experiments with five multi-view clustering methods and two time series clustering methods on seven benchmark datasets. The results demonstrate that our proposed method is superior to other state-of-the-art methods.

1. Introduction

Multivariate time series (MTS) data has gained increased attention due to sensing technology advancement in real-world applications such as finance [1], biomedicine [2], and climate [3]. Despite the wealth of information contained in time series data, analyzing it remains challenging due to its temporal dependence and interdimensional correlations. Furthermore, in practice, due to the complexity of data collection and transmission, some data points in time series may be absent, resulting in incomplete time series [4,5]. Incomplete MTS clustering [6], as an unsupervised time series analysis technique, has garnered significant attention, resulting in the development of various methods, including those based on information fusion [7], representation learning [8], and unsupervised learning [9]. However, these single-view methods either suffer from high computational costs or ignore the rich information contained in MTS data. Multi-view methods have demonstrated superiority over single-view methods in many application areas by integrating complementary and consensus information across different views [10–15]. Consequently, incomplete multi-view clustering (IMVC) techniques emerge as a promising solution to the above problem [16–20].

Existing IMVC methods can be divided into two categories: traditional [21–26] and deep learning-based methods [27–30]. The majority of current IMVC approaches fall into the traditional category, which generally exhibit limited representation learning capabilities [31–36]. Recently, deep IMVC methods have gained attention due to their powerful representation learning capabilities and scalability [33,37,38].

* Corresponding author.
E-mail addresses: lyr@xs.ustb.edu.cn (Y. Li), dumj@jsnu.edu.cn (M. Du), xjiang@jsnu.edu.cn (X. Jiang), nzhang@wzu.edu.cn (N. Zhang).

https://doi.org/10.1016/j.inffus.2024.102812
Received 23 August 2024; Received in revised form 23 October 2024; Accepted 19 November 2024
Available online 28 November 2024

To further improve the representation learning capability, some deep IMVC methods based on contrastive learning have been proposed [39,40]. However, these methods are not applicable to time series and have the following drawbacks:

(1) Most existing incomplete multi-view clustering methods first learn a suitable representation for the imputation task, followed by applying a clustering algorithm, such as k-means, to group the learned representations [41]. This strategy treats imputation (or representation learning) and clustering as two completely separate processes. Therefore, these methods are deficient in controlling the feedback of clustering performance into the imputation phase, and in ensuring that the optimized representation is more conducive to clustering.

(2) Existing contrastive learning approaches treat a sample and its augmented counterpart as positive pairs, while designating other samples as negative pairs [42,43]. For instance, in the image domain, data augmentation techniques such as rotation, cropping, and adding noise are commonly utilized to generate sample pairs [44,45]. However, given the unique characteristics of time series, these augmentation techniques are not applicable to time series. Furthermore, the design strategies for positive and negative sample pairs based on data augmentation techniques often introduce significant subjectivity and uncertainty, ultimately impacting the data representation capabilities of contrastive learning.

To solve the aforementioned problems, we propose a contrastive learning-based multi-view clustering method for incomplete multivariate time series. Our method integrates missing views recovery and clustering into an end-to-end framework, by jointly optimizing multiple objectives. It facilitates the mutual optimization and enhancement between representation learning and clustering within a unified framework. Firstly, it utilizes a multi-level learning framework to extract features at various levels including low-level, high-level, and semantic features. Specifically, it extracts low-level features using the encoder, and further extracts high-level and semantic features from the low-level features by stacking multilayer perceptrons. Then it performs distinct objectives in different feature spaces in a fusion-free manner, thereby eliminating potential conflicts that may arise between these objectives. Specifically, it recovers missing views and preserves the structural information of the original data through the prediction objective and reconstruction objective on low-level features, respectively. Considering the characteristics of MTS, our method learns common semantics by implementing contrastive learning from a multi-view perspective on high-level features. This approach can fully utilize the information in the raw data and reduce uncertainty. Additionally, our method utilizes cluster information within the high-level features to refine semantic labels, thereby enhancing clustering accuracy. Compared to previous work, the main contributions of this paper are summarized as follows:

- We propose a contrastive learning-based multi-view clustering method for incomplete multivariate time series. It integrates data recovery and clustering within a unified framework, enabling joint training and optimization by integrating multi-view and contrastive learning techniques.
- We design a flexible multi-view learning framework that resolves conflicts among certain objectives, thereby effectively mitigating the loss of interdimensional correlations within multivariate time series. It learns different level features and conducts different objectives in different feature spaces in a fusion-free manner, thus avoiding conflicts between objectives and enabling them to reinforce each other.
- We design a consistent representation learning module for multivariate time series based on contrastive learning. It directly leverages the information from the original data to discover the common semantics across views by employing contrastive learning from a multi-view perspective, thereby avoiding the introduction of uncertainty.

The remainder of this paper is structured as follows: Section 2 provides an overview of contrastive learning and multi-view clustering, as well as related work. In Section 3, we present our proposed method in detail. In Section 4, relevant experiments are conducted to evaluate our method, and the results are analyzed. Finally, Section 5 concludes the paper and outlines future work.

2. Related work

In this section, we briefly review two topics related to this work, i.e., contrastive learning and multi-view clustering.

2.1. Contrastive learning

As one of the most effective unsupervised learning methods, contrastive learning has gained significant attention in representation learning [46–48]. The basic idea of contrastive learning is to seek a latent feature space where the similarity between positive pairs is maximized and the similarity between negative pairs is minimized [49]. The encoder is trained to learn the features of data by maximizing the consistency between positive pairs and minimizing the consistency between negative pairs. Recently, several studies have investigated multi-view learning methods based on contrastive learning. For example, Tian et al. [50] propose a multi-view encoding framework based on contrastive learning to capture the underlying scene semantics. In [51], the authors develop a multi-view representation learning approach to solve the graph classification problem through contrastive learning. In addition, some studies have explored contrastive learning for multi-view clustering [52,53]. Jin et al. [54] propose CPSPAN, which adopts pair-observed data alignment to guide the construction of instance-to-instance correspondence across views. Furthermore, it mines consistent cross-view structural information by maximizing the matching alignment between paired-observed data. Lin et al. [52] propose CCR-Net, which explores the complementarity between views by a designed fusion module based on contrastive learning to learn a shared fusion weight [55,56]. Additionally, it incorporates a consistency representation module to ensure consistency. Wang et al. [57] propose a graph contrastive learning framework to solve incomplete multi-view clustering problems. It mainly consists of two parts: within-view contrastive learning and cross-view consistency learning, to maximize the mutual information of different views in a cluster.

As mentioned above, most contrastive learning studies employ data augmentation techniques to generate different views for constructing positive and negative sample pairs, and then learn consistency from these views [58–60]. This traditional approach assumes that different augmented views within the same instance have similar representations, which introduces a higher degree of uncertainty when applied to time-series data. Different from these existing contrastive learning approaches, our method directly learns consistency from a given multivariate time series dataset. Specifically, our method assumes that MTS originating from the same sample or within the same cluster ought to exhibit similar representations. By considering both instance-level and cluster-level losses, this strategy enhances the exploitation of inter-view informational complementarity in time series multi-view learning tasks.

2.2. Multi-view clustering

Traditional IMVC methods typically employ classical machine learning techniques for representation learning, which can be classified into four major categories: non-negative matrix factorization, kernel techniques, graph learning, and tensor-based. Subsequently, we will briefly discuss the tensor-based and graph learning approaches that are relevant to the baseline algorithm in the later experimental section. The tensor-based IMVC method, TCIMC [19], captures complementary information and spatial structure through a tensor Schatten p-norm-based

completion technique. The graph-based IMVC leverages graph structure information to improve cluster pattern recognition. For instance, PIMVC [23] establishes a graph-regularized projective consensus representation learning model, which learns the consensus representation in a unified low-dimensional subspace.

With the development of deep learning, neural networks' remarkable representation learning capabilities have enabled the extensive application of deep models to incomplete multi-view clustering tasks [57,61–63]. Based on the manner of processing information, the existing deep incomplete multi-view clustering (DIMVC) methods can be categorized into the following four categories: (1) Autoencoder-based methods. By extracting features from the data, deep autoencoders can learn consistent representations to impute missing data and achieve superior clustering performance [64]. Typical methods are [37,65]. Xu et al. [37] propose APADC, which applies adaptive feature projection to map all available data into a common space for feature learning, while considering distribution alignment during this process. The ultimate cluster information is obtained by maximizing mutual information between different views. Xu et al. [65] also propose DIMVC, which performs embedded feature learning on the complete data for each view and employs an EM-like optimization strategy to alternately facilitate feature learning and clustering. (2) GANs-based methods. These approaches directly generate imputation values for missing data using generative adversarial networks (GANs) by exploring mutual representations between views [66]. (3) Contrastive-based methods. A representative work is [67], which explores consistent representations by comparing multiple views and subsequently utilizes these representations to impute missing data. (4) GCN-based methods. It is a method that employs graph embedding techniques to learn node representations from multi-view data. For instance, ICMVC proposed by Chao et al. [68] utilizes GCN to handle missing values in multi-view data, and Xia et al. [69] propose SGCMC, which employs a multi-view shared graph attention encoder to learn graph embedding.

The differences between our method and existing work are as follows. First, almost all existing IMVC approaches [16] treat representation learning and clustering as two separate problems. These methods learn a suitable representation during the imputation phase, subsequently applying a clustering algorithm to group the enhanced representations. In contrast, our approach unifies imputation and clustering into a unified framework and allows them to reinforce each other. The imputation phase yields representations that facilitate effective clustering, while the clustering results inform the imputation to enhance accuracy in inferring missing data and to learn cluster-friendly representations. Second, existing IMVC methods are predominantly focused on the image domain, and there are no dedicated approaches designed specifically for time-series. We propose a multi-view feature learning framework, MVCIMTS, which is specifically designed for time series. It integrates an innovative contrastive learning approach from a multi-view perspective, aiming to learn consistent common semantics and reconstruct view-specific information. Consequently, it effectively harnesses the interdependencies among time-series variables, aiming to exploit the complementarity and consistency between different views during the multi-view learning process.

MFLVC [70] and COMPLETER [41] are two multi-view clustering methods closely related to the proposed MVCIMTS. MVCIMTS differentiates itself from these methods in several aspects: model architecture, learning strategy, and domain applicability. Regarding model architecture, MFLVC and COMPLETER utilize linear layers. In contrast, MVCIMTS is tailored for time series, with its core model structure based on the GRU, a type of recurrent neural network. When dealing with temporal sequences, MVCIMTS' GRU structure is particularly effective at capturing dynamic changes and temporal dependencies within the data. With respect to the learning strategy, COMPLETER is designed to learn suitable representations during the imputation process. The learned representations are applied to the subsequent k-means algorithm for clustering. It is fundamentally a form of representation learning for multiple views. Unlike COMPLETER, our model combines missing view recovery and clustering within a unified framework through integrating multiple objective functions, enabling them to mutually enhance and optimize each other. Its goal is to generate clustering-friendly representations, using feedback from clustering outcomes to enhance imputation accuracy and to further improve representation quality. Furthermore, MVCIMTS integrates multi-view learning with contrastive learning and exploits the inter-variable correlations inherent in time series to more effectively enhance representation learning capability. In multi-view learning, this strategy aims to utilize complementarity and consistency between different views [71]. In terms of domain applicability, MVCIMTS is specifically designed for handling time series with missing values, whereas MFLVC and COMPLETER are not focused on this issue.

3. Method

In this section, we first delve into the motivation behind the method proposed in this paper. Then, we provide brief definitions of the terms, and Table 1 summarizes the key notations used in this paper. Immediately following this, we provide an overview of the overall framework structure of MVCIMTS. Finally, we elaborate on each component of the proposed MVCIMTS in detail.

3.1. Motivation

In this section, we endeavor to construct a multi-view clustering framework for incomplete multivariate time series based on GRU. Multi-view clustering methods are quite common in the fields of image processing, such as [22,23,64]. In recent years, multi-view clustering methods based on contrastive learning have emerged as a new research focus. These methods typically construct positive and negative sample pairs by data augmentation techniques, then employ multi-view methods to learn multiple representations, followed by representation fusion, and finally apply a simple clustering algorithm to group the data. However, these methods fail to account for learning a latent space that is suitable for clustering while simultaneously inferring missing views effectively. Moreover, the uncertainty introduced by data augmentation methods in contrastive learning can lead to instability of the model's clustering performance. Due to the uniqueness of time series, there are few incomplete multi-view clustering methods in the time series domain. Therefore, the key issues lie in: (1) How to balance the relationship between imputation and clustering, and construct a framework that jointly performs time series imputation and clustering. (2) How to construct multi-view data for time series and mitigate the introduction of uncertainty in contrastive learning for time series.

Our goal is to perform clustering on incomplete time series using a multi-view approach, enhancing the model's clustering capabilities while maintaining its stability across varying degrees of missing data. To this end, we propose MVCIMTS, which employs GRU as its primary architecture, applies the concept of multi-view and contrastive learning to multivariate time series. It adopts a one-stage training strategy to simultaneously implement time series imputation and clustering. These enhancements boost the method's clustering performance and expand its scope of application scenarios.

3.2. Overview of model

Notations: Formally, we define a MTS dataset $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N\}$ with N samples. Each $\boldsymbol{X}_i \in \mathbb{R}^{D\times T}$ is the ith sample and $\boldsymbol{x}_i^j$ is the jth time series variable of $\boldsymbol{X}_i$, where D is the number of variables, and T represents the time series length. $\boldsymbol{X}^j \in \mathbb{R}^{N\times T}$ is the jth variable of sample. The dataset is divided into K clusters. Table 1 details important symbols.

Framework architecture: The structure of the proposed MVCIMTS is composed of two parts: imputation and clustering. As depicted in

Table 1
List of key notations used in this paper.

Notations	Descriptions
$\boldsymbol{X}$	MTS dataset
$\boldsymbol{X}_i$	The ith time series sample
$\boldsymbol{X}^j$	The jth time series variable
N	Number of samples
D	Number of variables
T	Length of time series
K	Number of clusters
d, d'	Dimension of low-level and high-level features
$\boldsymbol{Z}$	The low-level features
$\boldsymbol{H}$	The high-level features
$\boldsymbol{Q}$	The semantic features
$f^{(j)}(\cdot), \theta^{(j)}$	Encoder of the jth view, the jth encoder parameter
$g^{(j)}(\cdot), \phi^{(j)}$	Decoder of the jth view, the jth decoder parameter
$\boldsymbol{W}_H, \boldsymbol{W}_Q$	Weight parameters
τ_F, τ_L	Tuning factors
$\lambda_0, \lambda_1, \lambda_2$	Weight coefficients

Fig. 1, the upper part is responsible for inferring missing data, while the lower part implements the multi-view clustering algorithm. Overall, the model learns three levels of features for each view: low-level features $\boldsymbol{Z}^j$, high-level features $\boldsymbol{H}^j$, and semantic features $\boldsymbol{Q}^j$. It then implements different objectives for each feature type and finally unites these multiple learning objectives to achieve data recoverability and clustering. Specifically, in the imputation part, the missing views are recovered by a dual prediction mechanism operating. Meanwhile, view reconstruction is performed on low-level features to learn view-specific representations that maintain the original structural information of the data. In the clustering part, multi-view and contrastive learning is combined to learn cross-view common semantics on high-level features and cluster consistency across all views on semantic features. The imputation and clustering parts mutually reinforce each other. The learned representations in the imputation part should identify clusters well, and the clustering part feeds the clustering results back to the imputation part so that it can infer the missing data more accurately and learn cluster-friendly representations.

3.3. Objective functions

Cross-view dual prediction objective: To infer the missing views, the model utilizes a dual prediction mechanism as shown in Fig. 2. To better illustrate the mechanism principle, we take the example of bi-view data. Specifically, in a latent space parameterized by a neural network, the representation of one view is predicted from another by minimizing the conditional entropy $H(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})$, where $j=1$, $j'=2$ or $j=2$, $j'=1$. $\tilde{\boldsymbol{Z}}^j$ and $\tilde{\boldsymbol{Z}}^{j'}$ are the feature representations with missing values for $\boldsymbol{X}^j$ and $\boldsymbol{X}^{j'}$, respectively. The theoretical explanation of the dual prediction mechanism is shown in Fig. 3.

In Fig. 3, the solid and dashed rectangles represent the information contained in view $\boldsymbol{X}^j$ and view $\boldsymbol{X}^{j'}$, respectively. Mathematically, the mutual information $I(\tilde{\boldsymbol{Z}}^j, \tilde{\boldsymbol{Z}}^{j'})$, denoted as the gray area, quantifies the shared information between $\tilde{\boldsymbol{Z}}^j$ and $\tilde{\boldsymbol{Z}}^{j'}$, where $\tilde{\boldsymbol{Z}}^j$ and $\tilde{\boldsymbol{Z}}^{j'}$ are the representations of $\boldsymbol{X}^j$ and $\boldsymbol{X}^{j'}$, respectively. We learn consistent representations by maximizing $I(\tilde{\boldsymbol{Z}}^j, \tilde{\boldsymbol{Z}}^{j'})$. Furthermore, to promote missing views recovery, the conditional entropy $H(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})$ (the blue area) is minimized, where $j=1, j'=2$ or $j=2, j'=1$. Cross-view consistency learning and missing view recovery exhibit a reciprocal reinforcing relationship. In more detail, on the one hand, maximizing the mutual information $I(\tilde{\boldsymbol{Z}}^j, \tilde{\boldsymbol{Z}}^{j'})$ increases the shared information between the two views, which in turn makes it easier to recover one view from another, and enhances the recoverability of the data. On the other hand, $H(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})$ measures the amount of information in $\tilde{\boldsymbol{Z}}^j$ conditional on $\tilde{\boldsymbol{Z}}^{j'}$. Therefore, minimizing $H(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})$, which corresponds to the process of recovering missing views, promotes the elimination of inconsistent information across views, thereby improving the learning of consistent representations. Considering the above, it is evident that data recovery and cross-view consistency learning can be handled simultaneously and mutually reinforcing each other.

It can be obtained from Fig. 3 that $\tilde{\boldsymbol{Z}}^{j'}$ is completely determined by $\tilde{\boldsymbol{Z}}^{j'}$ if and only if the conditional entropy $H(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'}) = -\mathbb{E}_{\mathcal{P}_{(\tilde{\boldsymbol{Z}}^j,\tilde{\boldsymbol{Z}}^{j'})}}[log\mathcal{P}(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})] = 0$, where $\mathcal{P}(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})$ is a probability distribution. Towards this end, we adopt a general solution that maximizes the lower bound $\mathbb{E}_{\mathcal{P}_{(\tilde{\boldsymbol{Z}}^j,\tilde{\boldsymbol{Z}}^{j'})}}[log\mathcal{Q}(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})]$ of $\mathbb{E}_{\mathcal{P}_{(\tilde{\boldsymbol{Z}}^j,\tilde{\boldsymbol{Z}}^{j'})}}[log\mathcal{P}(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})]$, where $\mathcal{Q}(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})$ is a variational distribution.

In terms of variational distribution, $\mathcal{Q}(\cdot)$ can be of any type, for example a Gaussian distribution or a Laplace distribution. In our experiments, we assume $\mathcal{Q}(\cdot)$ is a Gaussian distribution $\mathcal{N}(\tilde{\boldsymbol{Z}}^j|\kappa^{(j')}(\tilde{\boldsymbol{Z}}^{j'}), \sigma\mathbf{I})$, where $\kappa^{(j')}(\cdot)$ is a parameterized model that maps $\tilde{\boldsymbol{Z}}^{j'}$ to $\tilde{\boldsymbol{Z}}^j$, and $\sigma\mathbf{I}$ is the variance matrix. By ignoring the Gaussian distribution constants, maximizing $\mathbb{E}_{\mathcal{P}_{(\tilde{\boldsymbol{Z}}^j,\tilde{\boldsymbol{Z}}^{j'})}}[log\mathcal{Q}(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})]$ is equivalent to:

$$\min \mathbb{E}_{\mathcal{P}_{(\tilde{\boldsymbol{Z}}^j,\tilde{\boldsymbol{Z}}^{j'})}} \left\| \tilde{\boldsymbol{Z}}^j - \kappa^{(j')}(\tilde{\boldsymbol{Z}}^{j'}) \right\|_2^2 \tag{1}$$

where $\|\cdot\|_2^2$ denotes the squared Euclidean distance. $\mathbb{E}$ represents the mean of the squared Euclidean distance between $\tilde{\boldsymbol{Z}}^j$ and $\kappa^{(j')}(\tilde{\boldsymbol{Z}}^{j'})$ under the probability distribution $\mathcal{P}(\tilde{\boldsymbol{Z}}^j|\tilde{\boldsymbol{Z}}^{j'})$.

With the above dual mapping, the missing representation $\tilde{\boldsymbol{Z}}^j$ can be easily predicted from $\tilde{\boldsymbol{Z}}^{j'}$, which is denoted as:

$$\tilde{\boldsymbol{Z}}^j = \kappa^{(j')}(\tilde{\boldsymbol{Z}}^{j'}) = \kappa^{(j')}(f^{(j')}_{\theta^{(j')}}(\boldsymbol{X}^{j'})) \tag{2}$$

where $f^{(j')}(\cdot)$ is the encoder of the j'th view, and $\theta^{(j')}$ is the j'th encoder parameter. $\tilde{\boldsymbol{Z}}^{j'}$ is the representation of $\boldsymbol{X}^{j'}$.

We further define the dual prediction objective as:

$$\mathcal{L}_P = \left\| \kappa^{(j)}(\tilde{\boldsymbol{Z}}^j) - \tilde{\boldsymbol{Z}}^{j'} \right\|_2^2 + \left\| \kappa^{(j')}(\tilde{\boldsymbol{Z}}^{j'}) - \tilde{\boldsymbol{Z}}^j \right\|_2^2 \tag{3}$$

Within-view reconstruction objective: MTS are usually redundant and randomly noisy, so mainstream methods always learn significant representations from the original features. The autoencoder is a widely used unsupervised model for mapping original features into a specific feature space. In our method, we implement the autoencoder structure using the GRU neural network. We process each individual view using an autoencoder to learn its latent representation $\boldsymbol{Z}^j$. This is accomplished by minimizing the reconstruction loss, denoted as $\mathcal{L}_Z$, which serves as the objective function for learning the underlying patterns within the data. We impose a reconstruction objective constraint on low-level features $\boldsymbol{Z}^j$ rather than high-level features $\boldsymbol{H}^j$. This is because, after multiple extractions and transformations, the feature representation tends to contain increasingly complex information, which may deviate from the original data structure. The reconstruction objective for all views is as follows:

$$\mathcal{L}_Z = \sum_{j=1}^{D} \sum_{i=1}^{N} \left\| \boldsymbol{x}_i^j - g^{(j)}_{\phi^{(j)}}(\boldsymbol{z}_i^j) \right\|_2^2 \tag{4}$$

where $\boldsymbol{x}_i^j$ represents the ith sample of $\boldsymbol{X}^j$, $j \in \{1, 2, \ldots, D\}$. $\boldsymbol{z}_i^j \in \mathbb{R}^d$ represents the feature representation of $\boldsymbol{x}_i^j$, and d is the dimension of low-level features. $g^{(j)}(\cdot)$ represent the decoder of the jth view, and $\phi^{(j)}$ is the jth decoder parameter. Thus, the representation of the ith sample in the jth view can be expressed as follows:

$$\boldsymbol{z}_i^j = f^{(j)}_{\theta^{(j)}}(\boldsymbol{x}_i^j) \tag{5}$$

where $f^{(j)}(\cdot)$ is the encoder of the jth view, and $\theta^{(j)}$ is the jth encoder parameter. With Eq. (4), the model retains as much view information as possible.

Cross-view contrastive learning objective: Since the low-level features $\{\boldsymbol{Z}^j\}_{j=1}^D \in \mathbb{R}^{N\times d}$ of each view obtained from Eq. (4) contain both common semantics and view-private information. Some MVC

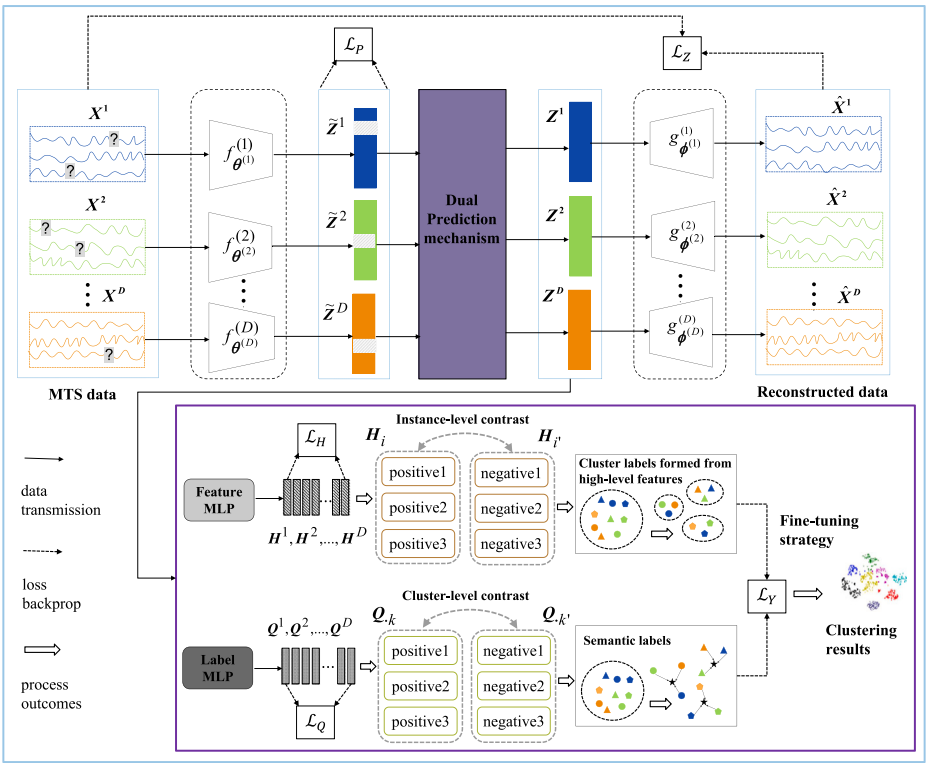


Fig. 1. The MVCIMTS model.

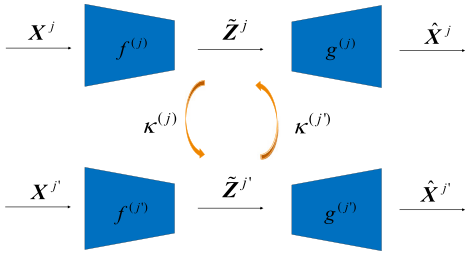


Fig. 2. Dual prediction mechanism.

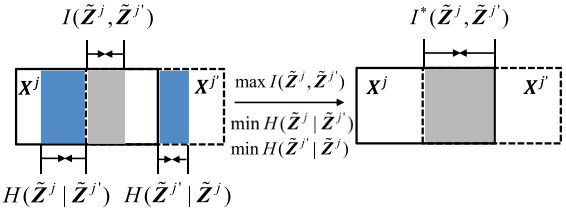


Fig. 3. Theoretical explanation.

methods explore the common semantics and learn consistent multi-view features by enforcing a consistency objective on $\{\boldsymbol{Z}^j\}_{j=1}^D$. In addition, these methods utilize Eq. (4) to impose constraints on $\{\boldsymbol{Z}^j\}_{j=1}^D$ during the reconstruction process to avoid model collapse. However, this strategy forces both the consistency objective and the reconstruction objective to rely on the same set of features, potentially creating a conflict that degrades the quality of $\{\boldsymbol{Z}^j\}_{j=1}^D$. This is because the consistency objective aims to learn the common semantics, whereas the reconstruction objective seeks to preserve the view-private information.

To address the above problem, we consider $\{\boldsymbol{Z}^j\}_{j=1}^D$ as low-level features and proceed to learn an additional layer of features, i.e., high-level features. To this end, we stack a multilayer perceptron (MLP) on $\{\boldsymbol{Z}^j\}_{j=1}^D$, named feature MLP, to obtain the high-level features $\{\boldsymbol{H}^j\}_{j=1}^D$, where each $\boldsymbol{h}_i^j \in \mathbb{R}^{d'}$ and d' is the dimension of high-level features. The feature MLP is a single-layer linear layer, denoted as $F(\{\boldsymbol{Z}^j\}_{j=1}^D; \boldsymbol{W}_H)$, where $\boldsymbol{W}_H$ represents the parameters of this layer. After that, we achieve the consistency objective by employing contrastive learning in the high-level feature space, enabling $\{\boldsymbol{H}^j\}_{j=1}^D$ to focus on learning the common semantics across all views.

Next, we introduce the contrastive learning strategy used in our model in detail. Different from some existing contrastive learning approaches, we regard each variable within MTS as an individual view. Moreover, we treat MTS from the same sample or within the same cluster as positive sample pairs and the remaining MTS as negative sample pairs, thus constructing positive and negative sample pairs of multiple views for contrastive learning. Then, we learn consistent representations by employing contrastive learning from a multi-view perspective. For each high-level feature $\boldsymbol{h}_i^j$, there are $(D \cdot N - 1)$ feature pairs, denoted as $\{\boldsymbol{h}_i^j, \boldsymbol{h}_{i'}^{j'}\}_{i=1,\dots,N}^{j=1,\dots,D}$. Among these, $\{\boldsymbol{h}_i^j, \boldsymbol{h}_i^{j'}\}_{j \neq j'}$ are $(D-1)$ positive feature pairs, representing features from the same sample, while the remaining $D(N-1)$ feature pairs are negative feature pairs, representing features from different sample. In contrastive learning, the similarity of positive pairs should be maximized, and the similarity of negative pairs should be minimized. We utilize the cosine similarity to measure the similarity between two features:

$$d(\boldsymbol{h}_i^j, \boldsymbol{h}_{i'}^{j'}) = \frac{\langle \boldsymbol{h}_i^j, \boldsymbol{h}_{i'}^{j'} \rangle}{\|\boldsymbol{h}_i^j\| \|\boldsymbol{h}_{i'}^{j'}\|} \tag{6}$$

where $\langle \cdot, \cdot \rangle$ is dot product operator. The feature contrastive loss between $\boldsymbol{H}^j$ and $\boldsymbol{H}^{j'}$ is denoted as follows:

$$\ell_{fc}^{(jj')} = -\frac{1}{N} \sum_{i=1}^{N} log \frac{e^{d(\boldsymbol{h}_i^j, \boldsymbol{h}_i^{j'})/\tau_F}}{\sum_{i'=1}^{N} \sum_{v=j,j'} e^{d(\boldsymbol{h}_i^j, \boldsymbol{h}_{i'}^v)/\tau_F} - e^{1/\tau_F}} \tag{7}$$

where τ_F denotes the tuning factor, which governs the similarity in contrastive learning between high-level features, and is set to 1.

We define the feature contrastive loss across all views as follows:

$$\mathcal{L}_H = \frac{1}{2}\sum_{j=1}^{D}\sum_{j'\neq j}\ell_{fc}^{(jj')} \tag{8}$$

Accordingly, the high-level features of each view are denoted as $\boldsymbol{H}^j = \boldsymbol{W}_H\boldsymbol{Z}^j = \boldsymbol{W}_H f^{(j)}(\boldsymbol{X}^j)$. The encoder $f^{(j)}(\cdot)$ serves to filter out random noise from $\boldsymbol{X}^j$ to obtain $\boldsymbol{Z}^j$, and then we perform the reconstruction objective on $\boldsymbol{Z}^j$ to recover the original data. This process not only mitigates the risk of model collapse but also aids in the preservation of both common semantics and view-private information within $\boldsymbol{Z}^j$. $\boldsymbol{W}_H$ facilitates filtering out view-private information in $\{\boldsymbol{Z}^j\}_{j=1}^D$ to obtain $\{\boldsymbol{H}^j\}_{j=1}^D$, and then the consistency objective on $\{\boldsymbol{H}^j\}_{j=1}^D$ allows it to mine common semantics from all views. Thus, the clusters formed based on high-level features are close to true semantic clusters without meaningless noise. Intuitively, high-level features within the same cluster are close to each other, resulting in densely shaped regions.

To obtain semantic features, we operate on low-level features extracted from the raw data. Specifically, the clustering assignment $\{\boldsymbol{Q}^j\}_{j=1}^D \in \mathbb{R}^{N\times K}$ for all views is obtained through a shared MLP stacked on the low-level features $\{\boldsymbol{Z}^j\}_{j=1}^D$, named label MLP, i.e., $L(\{\boldsymbol{Z}^j\}_{j=1}^D; \boldsymbol{W}_Q)$. We learn $\{\boldsymbol{Q}^j\}_{j=1}^D$ from $\{\boldsymbol{Z}^j\}_{j=1}^D$ rather than from $\{\boldsymbol{H}^j\}_{j=1}^D$, because this avoids interaction between $\boldsymbol{W}_H$ and $\boldsymbol{W}_Q$. Also, $\boldsymbol{W}_H$ and $\boldsymbol{W}_Q$ are not influenced by the gradient of $\mathcal{L}_Z$. The last layer of the label MLP is set to a $softmax(\cdot)$ operation that outputs a probability value q_{ik}^j, which denotes the probability that the ith sample in the jth view belongs to the kth cluster. As a result, semantic labels are identified by the maximum element value in the cluster assignment.

However, in real scenarios, some views of a sample may be assigned incorrect cluster labels due to misleading view-private information. To enhance robustness, it is necessary to achieve clustering consistency, meaning that the semantic features across all views of the same sample should correspond to the same cluster label. In other words, $\{\boldsymbol{Q}_{\cdot k}^j\}_{j=1}^D$ needs to be consistent. To achieve this consistency objective, we adopt contrastive learning, similar to the process we used to learn high-level features. For the jth view, the semantic features $\boldsymbol{Q}_{\cdot k}^j$ have $(D\cdot K-1)$ label pairs, i.e., $\{\boldsymbol{Q}_{\cdot k}^j, \boldsymbol{Q}_{\cdot k'}^{j'}\}_{k=1,\ldots,K}^{j=1,\ldots,D}$, where $\{\boldsymbol{Q}_{\cdot k}^j, \boldsymbol{Q}_{\cdot k}^{j'}\}_{j\neq j'}$ are $(D-1)$ positive label pairs and the remaining $D(K-1)$ label pairs are negative label pairs. Accordingly, we define the label contrastive loss between $\boldsymbol{Q}^j$ and $\boldsymbol{Q}^{j'}$ as follows:

$$\ell_{lc}^{(jj')} = -\frac{1}{K}\sum_{k=1}^{K} log\frac{e^{d(\boldsymbol{Q}_{\cdot k}^j, \boldsymbol{Q}_{\cdot k}^{j'})/\tau_L}}{\sum_{k'=1}^{K}\sum_{v=j,j'} e^{d(\boldsymbol{Q}_{\cdot k}^j, \boldsymbol{Q}_{\cdot k'}^v)/\tau_L} - e^{1/\tau_L}} \tag{9}$$

where τ_L denotes the tuning factor, which governs the similarity in contrastive learning between semantic features, and is set to 1.

We define the clustering-oriented consistency objective as:

$$\mathcal{L}_Q = \frac{1}{2}\sum_{j=1}^{D}\sum_{j'\neq j}\ell_{lc}^{(jj')} + \sum_{j=1}^{D}\sum_{k=1}^{K} s_k^j log s_k^j \tag{10}$$

where $s_k^j = \frac{1}{N}\sum_{i=1}^{N} q_{ik}^j$. The first part of Eq. (10) is designed to learn clustering consistency across all views. The second part serves as a regularization term to prevent all sample from being assigned to a single cluster.

Fine-tuning strategy: The model learns high-level features $\{\boldsymbol{H}^j\}_{j=1}^D$ and semantic features $\{\boldsymbol{Q}^j\}_{j=1}^D$ by the multi-view contrastive learning. We refine the clustering of the semantic features $\{\boldsymbol{Q}^j\}_{j=1}^D$ by matching them with the clusters formed from high-level features $\{\boldsymbol{H}^j\}_{j=1}^D$. This fine-tuning process utilizes cluster information in the high-level features to improve the clustering accuracy of the semantic features. Specifically, we adopt k-means to obtain cluster information. For the jth view, the cluster labels formed from high-level features of the ith sample are as follows:

$$p_i^j = \arg\min_k \left\|\boldsymbol{h}_i^j - \boldsymbol{c}_k^j\right\|_2^2 \tag{11}$$

where $\{\boldsymbol{h}_i^j\}_{i=1}^N$ represent the high-level features of the jth view of all sample, $\{\boldsymbol{c}_k^j\}_{k=1}^K$ denote the K cluster centers of the jth view, and $\boldsymbol{p}^j = \{p_i^j\}_{i=1}^N$ denote the cluster labels formed from high-level features of all sample of the jth view.

We define the semantic labels for the jth view as $\boldsymbol{y}^j = \{y_i^j\}_{i=1}^N \in \mathbb{R}^N$, where each element y_i^j can be calculated as Eq. (12), which is output by the label MLP.

$$y_i^j = \arg\max_k q_{ik}^j \tag{12}$$

Then, we modify $\boldsymbol{y}^j$ by the following matching formula:

$$\min_{\boldsymbol{A}^j} \boldsymbol{M}^j\boldsymbol{A}^j, s.t. \sum_{k=1}^{K} a_{kk'}^j = 1, \sum_{k'=1}^{K} a_{kk'}^j = 1 \tag{13}$$

where $\boldsymbol{A}^j \in \{0,1\}^{K\times K}$ is the boolean matrix and $\boldsymbol{M}^j \in \mathbb{R}^{K\times K}$ denotes the cost matrix. $\boldsymbol{M}^j = \max_{k,k'} \hat{m}_{kk'}^j - \tilde{\boldsymbol{M}}^j$ and $\hat{m}_{kk'}^j = \sum_{i=1}^N \mathbb{I}[y_i^j = k]\mathbb{I}[p_i^j = k']$, where $\mathbb{I}[\cdot]$ represents the indicator function. The modified cluster assignments $\hat{\boldsymbol{p}}_i^j \in \{0,1\}^K$ is a one-hot vector. The kth element of $\hat{\boldsymbol{p}}_i^j$ is 1 when $k = k\mathbb{I}[a_{kk'}^j = 1]\mathbb{I}[p_i^j = k'], k, k' \in \{1,2,\ldots,K\}$. We then fine-tune the model by cross-entropy loss:

$$\mathcal{L}_Y = -\sum_{j=1}^{D}\hat{\boldsymbol{P}}^j log\boldsymbol{Q}^j \tag{14}$$

where $\hat{\boldsymbol{P}}^j = [\hat{\boldsymbol{p}}_1^j, \hat{\boldsymbol{p}}_2^j, \ldots, \hat{\boldsymbol{p}}_N^j] \in \mathbb{R}^{N\times K}$. In this way, we can utilize the cluster information contained in the high-level features to improve clustering results. Finally, the cluster label of the ith sample is:

$$y_i = \arg\max_k\left(\frac{1}{D}\sum_{j=1}^{D} q_{ik}^j\right) \tag{15}$$

3.4. Optimization

Lastly, the overall objective function of our proposed MVCIMTS consists of three components: reconstruction objective, prediction objective, and contrastive learning objective, which are defined as follows:

$$\mathcal{L}_{MVCIMTS} = \mathcal{L}_Z + \lambda_0\mathcal{L}_P + \lambda_1\mathcal{L}_H + \lambda_2\mathcal{L}_Q, \tag{16}$$

where the prediction objective $\mathcal{L}_P$ in Eq. (3) is designed to infer the missing views. $\mathcal{L}_Z$ is the within-view reconstruction objective in Eq. (4), aiming to learn view-specific representations, and maintain the original data structure. The contrastive learning objective includes the cross-view common semantic consistency objective $\mathcal{L}_H$ in Eq. (8) and the clustering consistency objective $\mathcal{L}_Q$ in Eq. (10). Specifically, $\mathcal{L}_H$ is responsible for learning the common semantics of all views, and $\mathcal{L}_Q$ learns the clustering consistency of all views. The coefficients λ_0 is set to a fixed value of 1, and λ_1 and λ_2 are set to 0.1.

The complete optimizing process is described in Algorithm 1.

4. Results

In this section, we evaluate the effectiveness of our proposed MVCIMTS by comparing it with five state-of-the-art IMVC methods and two methods designed for temporal data on seven MTS datasets. First, we present the experimental settings in Section 4.1. Then, we compare our MVCIMTS with state-of-the-art methods in Section 4.2. And, we perform robustness experiments with different missing rates in Section 4.3. Further, we also present the visualization results in Section 4.4. After that, we conduct the ablation studies and parameter analysis in Sections 4.5 and 4.6, respectively. Finally, we investigate convergence analysis in Section 4.7.

4.1. Experimental settings

Baseline algorithms: To evaluate the performance of our proposed method, we compared it with the following five state-of-the-art IMVC methods and two temporal methods.

Algorithm 1: The MVCIMTS Algorithm

Input: Multivariate time series $\{\boldsymbol{X}^j\}_{j=1}^D$; Number of clusters K; The coefficients λ_0, λ_1, and λ_2; Tuning factors τ_F and τ_L; Batch size b; Maximum iteration: $Maxepoch$.

Output: Clustering results $\boldsymbol{y}$.

Initialize $\{\theta^{(j)}, \phi^{(j)}\}_{j=1}^D$ with Eq. (4);
for $epoch = 1$ *to* $Maxepoch$ **do**
  for *batch_size=b in* $\boldsymbol{X}$ **do**
    Perform imputation with Eq. (3), and obtain the low-level features $\{\boldsymbol{Z}^j\}_{j=1}^D$ with Eq. (4);
    Obtain the high-level features $\{\boldsymbol{H}^j\}_{j=1}^D$ with (8), and the semantic features $\{\boldsymbol{Q}^j\}_{j=1}^D$ with (10);
    Compute initial cluster labels formed from high-level features with Eq. (11);
    Compute initial semantic labels with Eq. (12);
    Optimize $\boldsymbol{W}_H$, $\boldsymbol{W}_Q$, $\{\theta^{(j)}, \phi^{(j)}\}_{j=1}^D$ with Eq. (16);
    Modify semantic labels by using cluster labels formed from high-level features with Eq. (13);
    Fine-tune $\boldsymbol{W}_Q$ with Eq. (14);
    Obtain final cluster labels with Eq. (15).
Return Clustering results $\boldsymbol{y}$.

- APADC [37] is a deep IMVC method that considers distribution alignment in feature learning.
- DIMVC [65] presents a deep IMVC framework that maps complete data embedding features into a high-dimensional space to discover linear separability without fusion.
- CPSPAN [54] proposes a cross-view partial sample and prototype alignment deep network. It employs pair-observed data alignment to guide the construction of instance-to-instance correspondence between views.
- TCIMC [19] utilizes the tensor Schatten p-norm-based completion technique to compare the similarity of interview graphs, incorporating complementary information and spatial structure.
- PIMVC [23] proposes a graph-regularized projective consensus representation learning model by learning the consensus representation in a unified low-dimensional subspace.
- CRLI [8] is a deep learning-based method for single-view clustering, specifically designed for incomplete time series.
- VaDER [72] is an end-to-end model, which utilizes a Gaussian mixture variational autoencoder with two LSTMs to cluster multivariate time series with missing data.

For a fair comparison, all baselines are compared using the recommended parameters and network structures.

Network architectures and implementation details: For our proposed MVCIMTS,[1] the following settings are adopted for all datasets. Concretely, the autoencoders of all views are implemented by GRU neural networks with the same structure. The dimensionality of embeddings is set to 10. The activation function is ReLU. We adopt Adam to optimize the model with a learning rate of 0.0001. In the experiments, we set the pre-training epoch of autoencoders to 300. The batchsize is set to 64. The experiment is implemented on PyTorch and Windows 11 with an NVIDIA 3060Ti GPU.

Datasets: We use seven datasets in our experiments,[2] the detailed description of these datasets is provided in Table 2. We construct incomplete multi-view datasets with varying missing rates $(0.1, 0.3, 0.5, 0.7)$.

Evaluation metrics: The clustering effectiveness is evaluated by the rand index (RI), normalized mutual information (NMI), and clustering accuracy (ACC). The higher values indicate better clustering performance.

[1] Code available: https://github.com/Du-Team/MVCIMTS.

[2] Data available: https://www.timeseriesclassification.com/dataset.php.

Table 2
Details of the seven multivariate time series benchmark datasets.

Dataset	Samples	Dimensions	Length	Classes
BM	80	6	100	4
Epilepsy	275	3	207	4
SWJ	27	4	2500	3
LP4	117	6	15	3
LP5	164	6	15	5
PD	10 992	2	8	10
RS	303	6	30	4

4.2. Experimental results and analysis

Table 3 presents experimental results of all methods on seven datasets, where the optimal result is bold and the second best result is underlined.

From Table 3, we can observe that DIMVC performs the worst of all multi-view methods. For instance, DIMVC consistently produces NMI values below 0.1 on the SWJ dataset across four different missing rates, whereas all other methods achieve NMI values exceeding 0.1. The observed performance of DIMVC may be attributed to its reliance on mapping multi-view embedding features to a high-dimensional space to extract complementary information. This strategy potentially necessitates a substantial amount of data for effective training. Consequently, when the information content within the views is insufficient, the method may fail to yield adequate complementary information, resulting in diminished performance. APADC, like DIMVC, is a deep learning approach. In the experiments, APADC achieves superior performance to DIMVC. The main reason for this is that APADC considers distributional alignment between features, which maximizes consistency across different views. Compared to DIMVC, TCIMC and PIMVC are capable of capturing both complementary information and spatial structure information between different views, resulting in better performance. CPSPAN outperforms the other four methods, likely due to its superior ability to learn cross-view information. It takes into account the differences across views while learning cross-view consistency. Although CRLI and VaDER are specifically designed for multivariate time series, they generally perform worse than most multi-view methods. The reason for this could be that the single-view approach they employ does not fully utilize the richness of information contained within the time series nor the complementarity between variables.

For our method, the following conclusions can be drawn: (1) First, the performance of the proposed MVCIMTS is either comparable to or significantly better than that of other methods on all datasets. For instance, our method obtains the highest metric values for all missing

Table 3
Clustering results of all methods on seven datasets..

	Missing rates	0.1			0.3			0.5			0.7		
	Metrics	RI	NMI	ACC	RI	NMI	ACC	RI	NMI	ACC	RI	NMI	ACC
BM	APADC	0.795	0.681	0.65	0.717	0.538	0.565	0.708	0.535	0.525	0.697	0.485	0.515
	DIMVC	0.619	0.087	0.392	0.605	0.084	0.378	0.597	0.082	0.351	0.582	0.075	0.385
	CPSPAN	0.599	0.412	0.45	0.513	0.319	0.425	0.619	0.398	0.45	0.481	0.285	0.4
	TCIMC	0.739	0.483	0.625	0.768	0.507	0.7	0.667	0.336	0.54	0.376	0.091	0.335
	PIMVC	0.753	0.426	0.65	0.712	0.307	0.575	0.641	0.234	0.525	0.555	0.096	0.325
	CRLI	0.551	0.119	0.35	0.653	0.171	0.5	0.606	0.118	0.4	0.604	0.165	0.375
	VaDER	0.412	0.154	0.35	0.231	0.12	0.25	0.455	0.256	0.475	0.258	0.047	0.275
	IMVCTS	**0.858**	**0.728**	**0.825**	**0.794**	**0.595**	**0.725**	**0.786**	**0.588**	**0.7**	**0.804**	**0.601**	**0.75**
Epilepsy	APADC	0.667	0.265	0.517	0.629	0.14	0.393	0.617	0.086	0.366	0.533	0.231	0.458
	DIMVC	0.65	0.104	0.382	0.637	0.093	0.365	0.625	0.035	0.345	0.635	0.074	0.379
	CPSPAN	0.688	0.268	0.453	0.691	0.251	0.482	0.687	0.247	0.474	0.682	0.212	0.474
	TCIMC	0.67	0.272	0.477	0.536	0.157	0.419	0.575	0.082	0.416	0.525	0.075	0.336
	PIMVC	0.622	0.056	0.343	0.646	0.067	0.416	0.623	0.027	0.343	0.554	0.044	0.328
	CRLI	0.56	0.073	0.355	0.551	0.088	0.37	0.575	0.079	0.348	0.543	0.067	0.333
	VaDER	0.632	0.061	0.348	0.629	0.035	0.326	0.622	0.019	0.326	0.618	0.023	0.326
	IMVCTS	**0.836**	**0.608**	**0.774**	**0.765**	**0.477**	**0.696**	**0.744**	**0.41**	**0.667**	**0.739**	**0.389**	**0.652**
SWJ	APADC	0.525	0.306	0.587	0.513	0.286	0.541	0.458	0.245	0.517	0.462	0.259	0.522
	DIMVC	0.522	0.057	0.347	0.519	0.039	0.302	0.496	0.068	0.318	0.476	0.009	0.313
	CPSPAN	0.5	0.124	0.5	0.5	0.295	0.527	0.5	0.109	0.417	0.53	0.189	0.5
	TCIMC	0.626	0.329	0.556	0.561	0.224	0.5	0.496	0.174	0.5	0.452	0.109	0.377
	PIMVC	0.61	0.223	0.574	0.591	0.179	0.546	0.482	0.125	0.501	0.457	0.103	0.434
	CRLI	0.457	0.123	0.4	0.505	0.113	0.467	0.552	0.295	0.533	0.391	0.191	0.467
	VaDER	0.591	0.27	0.533	0.543	0.269	0.47	0.569	0.184	0.458	0.577	0.178	0.428
	IMVCTS	**0.712**	**0.45**	**0.667**	**0.636**	**0.333**	**0.583**	**0.604**	**0.299**	**0.542**	**0.598**	**0.286**	**0.53**
LP4	APADC	0.637	0.37	0.709	0.6	0.312	**0.703**	0.601	0.289	0.684	0.536	0.197	0.667
	DIMVC	0.622	0.12	0.357	0.61	0.094	0.347	0.591	0.059	0.312	0.583	0.071	0.348
	CPSPAN	0.627	0.371	0.682	0.544	0.212	0.641	0.598	0.315	0.65	0.597	0.313	**0.684**
	TCIMC	0.614	0.29	0.638	0.618	0.367	0.609	0.538	0.118	0.479	0.519	0.107	0.479
	PIMVC	0.533	0.187	0.513	0.57	0.196	0.496	0.557	0.108	0.573	0.5	0.022	0.444
	CRLI	0.461	0.022	0.607	0.455	0.035	0.615	0.503	0.011	0.359	0.509	0.027	0.496
	VaDER	0.562	0.223	0.547	0.537	0.048	0.47	0.555	0.1	0.547	0.538	0.032	0.496
	IMVCTS	**0.638**	**0.407**	**0.721**	**0.621**	**0.41**	0.658	**0.623**	**0.406**	**0.692**	**0.622**	**0.398**	0.65
LP5	APADC	0.411	0.156	0.333	0.309	0.121	0.312	0.429	0.181	0.338	0.333	0.128	0.32
	DIMVC	0.468	0.063	0.303	0.48	0.061	0.309	0.443	0.076	0.336	0.431	0.044	0.295
	CPSPAN	0.353	0.214	0.317	0.343	0.198	0.311	0.34	0.195	0.308	0.322	0.166	0.297
	TCIMC	0.78	0.401	0.581	**0.772**	**0.374**	**0.548**	**0.725**	0.258	0.412	0.597	0.098	0.299
	PIMVC	0.736	0.328	0.451	0.727	0.277	0.506	0.67	0.159	0.384	0.576	0.107	0.366
	CRLI	0.484	0.033	0.287	0.523	0.062	0.287	0.634	0.068	0.311	0.638	0.023	0.268
	VaDER	0.666	0.05	0.287	0.674	0.043	0.281	0.673	0.064	0.305	0.673	0.105	0.335
	IMVCTS	**0.803**	**0.452**	**0.592**	0.718	**0.374**	0.488	0.716	**0.367**	**0.482**	**0.717**	**0.356**	**0.476**
PD	APADC	0.876	0.518	0.558	0.872	0.525	0.556	0.88	0.533	0.599	0.888	0.551	0.616
	DIMVC	0.569	0.225	0.437	0.571	0.22	0.449	0.568	0.129	0.405	0.559	0.07	0.394
	CPSPAN	0.898	0.592	0.62	0.902	0.608	**0.666**	0.897	0.598	0.635	0.895	0.594	0.623
	TCIMC	0.818	0.002	0.117	0.813	0.002	0.116	0.815	0.003	0.117	0.804	0.002	0.116
	PIMVC	0.675	0.267	0.303	0.198	0.002	0.109	0.308	0.003	0.111	0.589	0.005	0.113
	CRLI	0.823	0.052	0.188	0.812	0.042	0.167	0.813	0.044	0.173	0.806	0.03	0.154
	VaDER	0.81	0.006	0.128	0.817	0.005	0.123	0.818	0.005	0.122	0.818	0.013	0.139
	IMVCTS	**0.926**	**0.708**	**0.767**	**0.913**	**0.623**	0.652	**0.905**	**0.616**	**0.65**	**0.903**	**0.608**	**0.657**
RS	APADC	0.53	0.106	0.371	0.539	0.116	0.375	0.458	0.233	0.433	0.399	0.153	0.376
	DIMVC	0.645	0.279	0.531	0.605	0.148	0.434	0.625	0.143	0.423	0.592	0.134	0.4
	CPSPAN	0.668	0.272	0.49	0.649	0.239	0.497	0.649	0.251	0.45	0.64	0.194	0.424
	TCIMC	0.662	0.153	0.399	0.667	0.268	0.523	0.648	0.189	0.477	0.593	0.111	0.37
	PIMVC	0.676	0.279	0.55	0.678	0.256	0.51	0.653	0.178	0.483	0.598	0.076	0.364
	CRLI	0.38	0.022	0.296	0.503	0.028	0.296	0.589	0.039	0.329	0.617	0.034	0.336
	VaDER	0.641	0.086	0.441	0.633	0.066	0.375	0.629	0.037	0.362	0.616	0.011	0.296
	IMVCTS	**0.852**	**0.699**	**0.775**	**0.85**	**0.728**	**0.737**	**0.844**	**0.696**	**0.73**	**0.841**	**0.686**	**0.717**

rate cases in the four datasets: BM, Epilepsy, SWJ, and RS. On LP5 with a miss rate of 0.7, MVCIMTS achieves the RI value of 0.717, the NMI value of 0.356, and the ACC value of 0.476, which are 12%, 19%, and 11% higher than the second-best method, respectively. (2) Second, MVCIMTS does not exhibit an overwhelming advantage when confronted with LP4, LP5, and PD datasets, which are characterized by short time steps. The probable reason for this may be the insufficient number of data features learned, which hinders the capture of long-term dependencies within the sequence. However, MVCIMTS demonstrates the greatest improvement on datasets with high missing rates. For instance, on LP4 and RS with a missing rate of 0.7, MVCIMTS outperforms CPSPAN by approximately 12% and 20% in RI, respectively.

The reasons for our method's advanced results can be explained by the following factors: (1) Our method learns different level features by its multi-level feature learning framework. Then it implements the reconstruction objective and consistency objective on different feature spaces in a fusion-free manner to avoid conflicts between the two objectives. This strategy enables mining common semantics across views while maintaining the original information of each view. Also, it protects the representation learning process from the influence of low-quality views, while reducing the loss of association information between views. (2) Our method leverages the characteristics of MTS by employing contrastive learning from a multi-view perspective to extract common semantics across different views. By doing so, it can avoid the introduction of uncertainty and enhance representation quality.

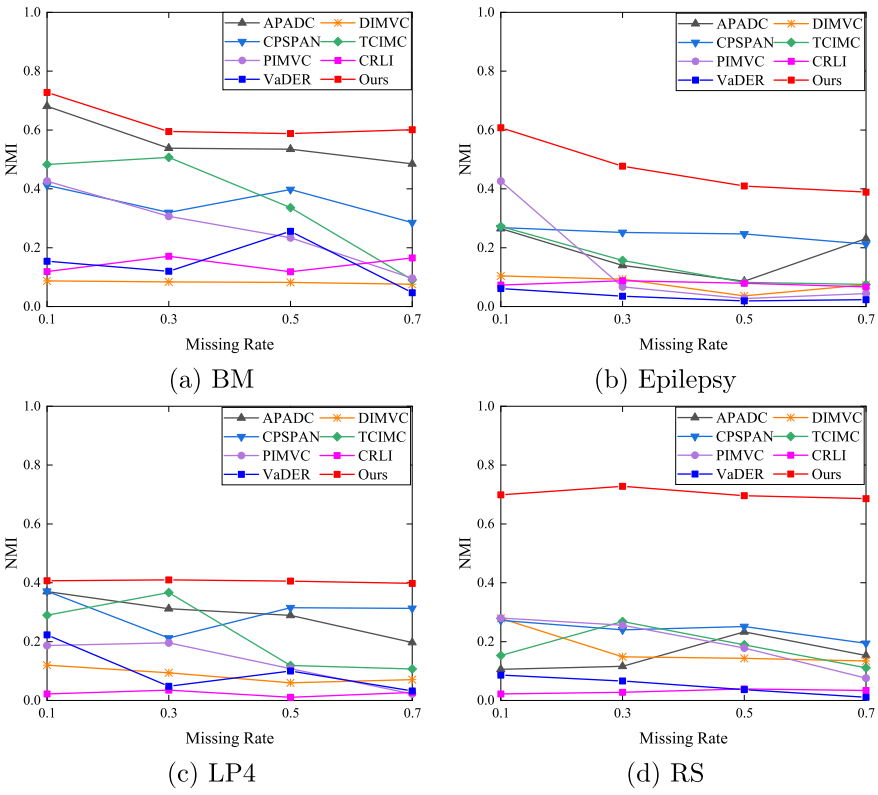


Fig. 4. The clustering NMI of all methods with different missing rates on four datasets.

4.3. Robustness analysis

Fig. 4 depicts the trend of NMI metric for all methods as the missing rate increases from 0.1 to 0.7 across BM, Epilepsy, LP4, and RS datasets. Overall, the NMI values for all methods decrease as the missing rate increases. This is because the more missing data there is, the more challenging it is to learn data distribution. Further, we can observe that with the increase in the missing rate, the proposed MVCIMTS's performance slowly declined in a certain area, while the decline trend of the comparison algorithms is evident, especially on the BM and LP4 datasets. Since our MVCIMTS treats data recovery and consistency learning as a whole, and learns a sufficient representation by jointly optimizing them. This reduces uncertainty in the imputation process, resulting in significant robustness.

4.4. Representation visualization

To complement the evaluation of our method's performance, we further investigate its effectiveness in data recovery. For this purpose, we apply t-SNE visualization to the embedded features and centroids learned by MVCIMTS on RS dataset with four different missing rates, as shown in Fig. 5. In the figure, different clusters are highlighted in different colors, cluster centers are marked with asterisks, sample with complete data are shown as dots, and sample with incomplete data are indicated by crosses. The visualization results indicate that the different clusters are well separated. Consequently, it can be concluded that our method demonstrates excellent robustness under different missing rates.

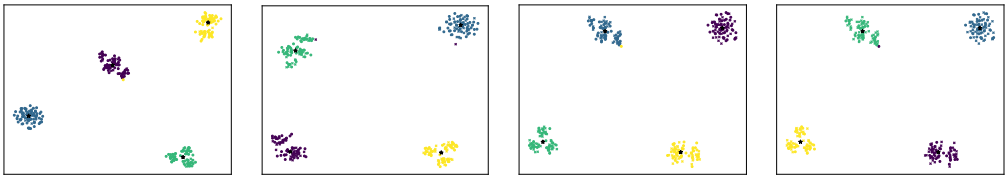
4.5. Ablation studies

We conduct ablation studies on the loss components in Eqs. (14) and (16) to demonstrate the importance of each component of our method. Table 4 presents the detailed results of ablation study, where $\surd$ denotes the adoption of a loss component, and the optimal performance values are emphasized in bold.

According to the experimental results shown in Table 4, we could observe that: (1) The best performance is obtained when full loss terms are used, indicating that all the components play an indispensable role in MVCIMTS. Furthermore, the performance achieved using two loss functions is consistently inferior to the case of using three loss functions, which also validates the significance of each component. (2) (H) perform better than (D), suggesting that reconstruction objective $\mathcal{L}_Z$ plays an essential role. It preserves the characteristics of each view as much as possible to avoid feature space collapse during feature learning. (3) (H) perform better than (E), indicating that the prediction objective $\mathcal{L}_P$ plays a crucial role in recovering missing views. Moreover, by conducting $\mathcal{L}_P$, we enable data recovery and cross-view consistency learning to promote each other. Comparing (F) with (H), it is clear that the learned high-level features by $\mathcal{L}_H$ improve clustering effectiveness. We speculate that the consistent learning implemented by $\mathcal{L}_H$, which maximizes common semantics across views while filtering out irrelevant private information from individual views, thereby improving clustering effectiveness. Further, $\mathcal{L}_H$ provides more performance improvements than $\mathcal{L}_P$. For instance, in RS on RI and NMI, (H) outperforms (E) by about 0.5% and 1%, while (H) outperforms (F) by about 1.46% and 4.48%, respectively. This suggests that it is necessary to maintain the similarity of representation learning across views in the multi-view learning.

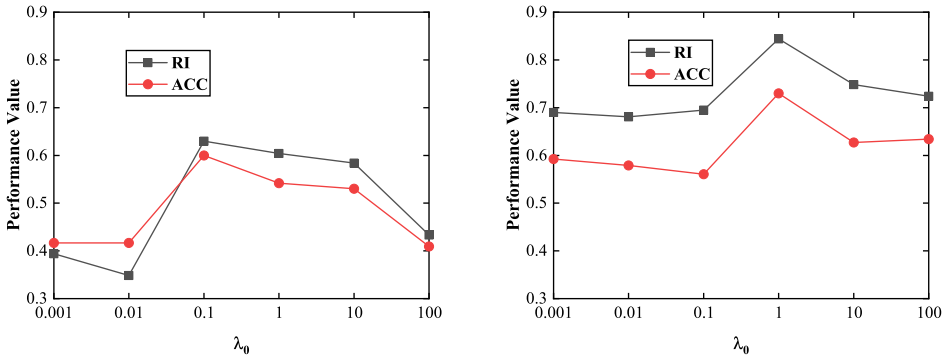
4.6. Parameter analysis

The proposed MVCIMTS consists of five trade-off parameters, i.e., λ_0, λ_1, λ_2, τ_F, and τ_L. In spite of the promising performance of MVCIMTS with these fixed parameters, it is still important to explore the full potential of our method and the influence of these parameters. We conduct experiments on SWJ and RS with a missing rate of 0.5, and utilize the RI and ACC metrics for performance evaluation.



(a) Missing rate = 0.1 (b) Missing rate = 0.3 (c) Missing rate = 0.5 (d) Missing rate = 0.7

Fig. 5. Visualization of the embedding features and centroids on RS with four different missing rates via t-SNE. Dots denote the sample with complete data and crosses represent the sample with incomplete data.



(a) SWJ (b) RS

Fig. 6. Performance of the λ_0 parameter on SWJ and RS with a missing rate of 0.5.

Table 4
Ablation experiments on three datasets with the missing rate of 0.5.

	Loss components				BM		Epilepsy		RS	
	$\mathcal{L}_Z$	$\mathcal{L}_P$	$\mathcal{L}_H$	$\mathcal{L}_Y$	RI	NMI	RI	NMI	RI	NMI
(A)	√	√			0.258	0.047	0.719	0.336	0.622	0.117
(B)	√			√	0.54	0.341	0.622	0.145	0.649	0.3
(C)		√		√	0.231	0	0.725	0.371	0.607	0.202
(D)		√	√	√	0.755	0.521	0.721	0.386	0.837	0.675
(E)	√		√	√	0.75	0.568	0.73	0.39	0.838	0.686
(F)	√	√		√	0.749	0.553	0.718	0.384	0.829	0.651
(G)	√	√	√		0.774	0.57	0.724	0.378	0.829	0.648
(H)	√	√	√	√	**0.786**	**0.588**	**0.744**	**0.41**	**0.844**	**0.696**

Firstly, we assess the impact of λ_0 on our method's performance. Both SWJ and RS datasets are fixed with the following parameters: $\lambda_1 = 0.1, \lambda_2 = 0.1, \tau_F = 1$, and $\tau_F = 1$. Fig. 6 depicts the changes in clustering performance as λ_0 varies for SWJ and RS datasets, as measured by RI and ACC. It is apparent that the clustering performance fluctuates with the variation of λ_0. Accordingly, MVCIMTS requires specific parameter values for λ_0 to maintain clustering accuracy, and clustering performance reaches its optimal value at $\lambda_0 = 1$. Therefore, we recommend setting λ_0 to a fixed value of 1.

Next, we evaluate the performance of the proposed method in relation to λ_1 and λ_2. The remaining parameters are set to fixed values: $\lambda_0 = 1, \tau_F = 1$, and $\tau_L = 1$. Fig. 7 indicates the results of RI and ACC for varying λ_1 and λ_2 values on the SWJ and RS datasets. We change the value of λ_1 and λ_2 in the range of $\{0.001, 0.01, 0.1, 1, 10, 100\}$. From the results, we can observe that MVCIMTS performs relatively smoothly in Fig. 7(c). Additionally, MVCIMTS achieves relatively better performance within a certain range of values for λ_1 and λ_2 in Figs. 7(a), 7(b), and 7(d). The reference value range for λ_1 and λ_2 is $\{0.1, 1\}$.

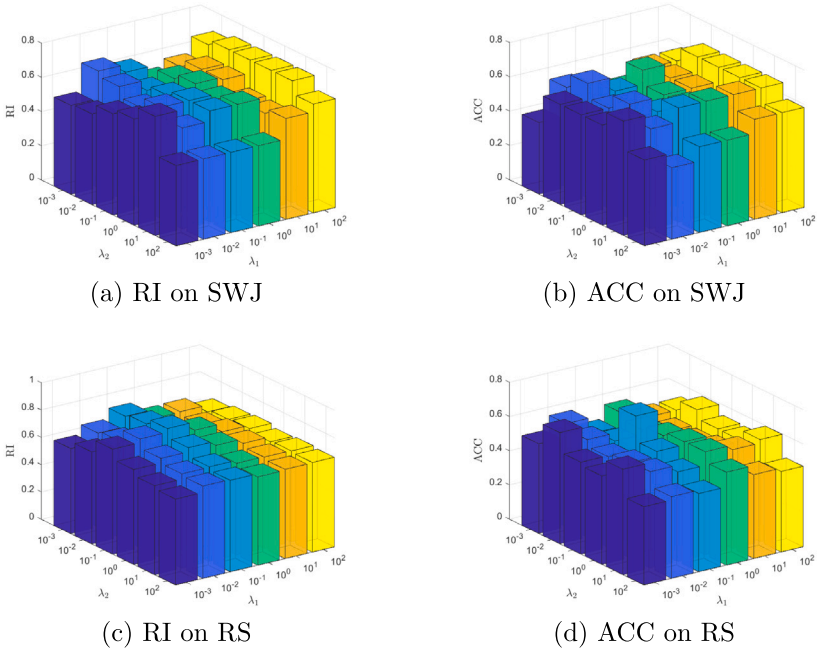
Finally, we investigate how τ_F and τ_L impact clustering performance. As for the remaining parameters, they are fixed: $\lambda_0 = 1, \lambda_1 = 0.1$, and $\lambda_2 = 0.1$. Fig. 8 depicts the results of RI and ACC for varying τ_F and τ_L values on the SWJ and RS datasets. τ_F ranges from $\{0.5, 0.6, 0.7, 0.8, 0.9, 1\}$, and τ_L ranges from $\{0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$. Based on the results shown in Figs. 8(a) and 8(b), we recommend setting τ_F within the range of $\{0.6, 1\}$, and τ_L in the range $\{0.7, 0.8, 1\}$ to achieve high metrics. As demonstrated in Figs. 8(c) and 8(d), we can observe that our proposed method exhibits insensitivity to the parameters τ_F and τ_L on the RS dataset. For simplicity, we set $\tau_F = 1$ and $\tau_L = 1$ in our method.

4.7. Convergence analysis

Fig. 9 shows the loss value trend of MVCIMTS on PD and RS datasets with a missing rate of 0.5 as the number of iterations increases. As shown in the figures, we can observe fluctuations in the loss curve, which may be attributed to the alternating optimization strategy employed during the training process. Furthermore, it can be seen that the loss value exhibits an overall downward trend and decreases rapidly during the first few steps, indicating that MVCIMTS possesses good convergence property.

5. Conclusions

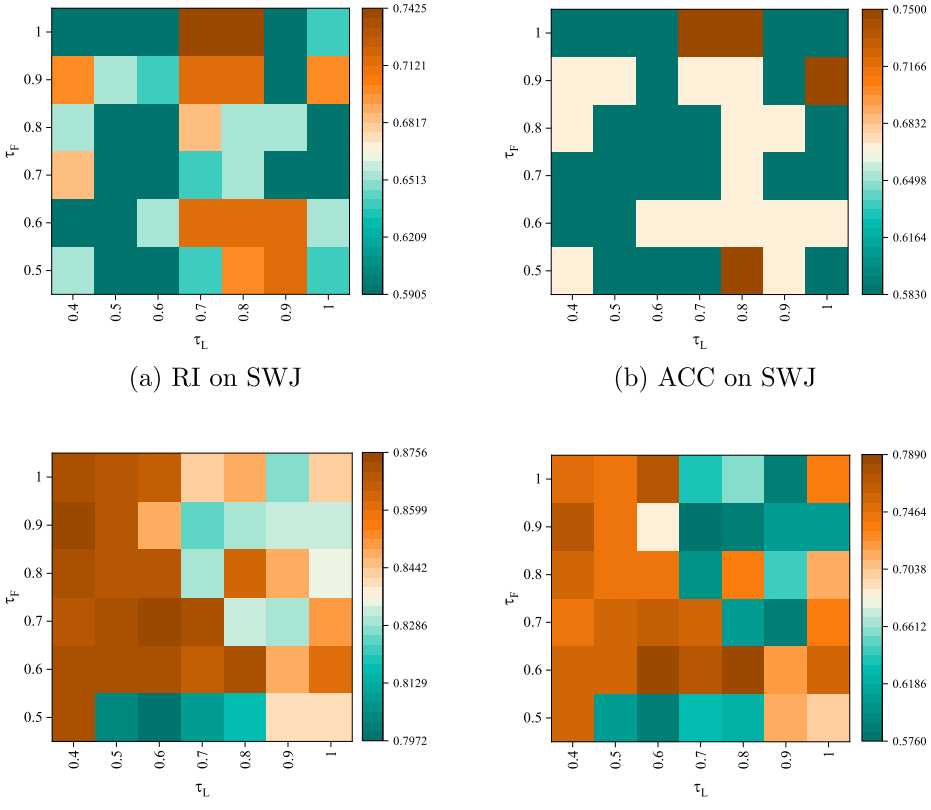
In this paper, we propose a novel contrastive learning-based multi-view clustering method for incomplete multivariate time series. This method employs a multi-level feature learning framework to learn features at different levels of the time series, enabling the model to better understand and exploit structural information within the time series. Firstly, we utilize view-specific encoders to learn view-specific features at different levels, including low-level, high-level, and semantic features. Subsequently, we implement distinct objectives in different feature spaces to mitigate potential conflicts among them and reduce the loss of association information between views. In addition, by leveraging the characteristics of multivariate time series, our method employs contrastive learning from a multi-view perspective to achieve both representation and clustering consistency, thereby enhancing clustering performance. Comparative experimental results on various datasets show that our proposed MVCIMTS is superior to other state-of-the-art methods in multi-view clustering tasks for incomplete multivariate time series.



(a) RI on SWJ (b) ACC on SWJ

(c) RI on RS (d) ACC on RS

Fig. 7. Performance of the λ_1 and λ_2 parameters on SWJ and RS databases with a missing rate of 0.5.



(a) RI on SWJ (b) ACC on SWJ

(c) RI on RS (d) ACC on RS

Fig. 8. Performance of the τ_F and τ_L parameters on SWJ and RS databases with a missing rate of 0.5.

MVCIMTS uses the existing views within a sample to infer the missing views, inherently assuming that there is a correlation between the views. However, when there is not a significant correlation between the views, MVCIMTS may lose its applicability. In the future, we intend to infer information about missing views from the instances within the same cluster. We will also focus on the impact of feature distribution differences between complete and incomplete data on representation learning. Additionally, we plan to introduce causal reasoning to evaluate how variables interact with each other, and to further design generative and clustering methods.

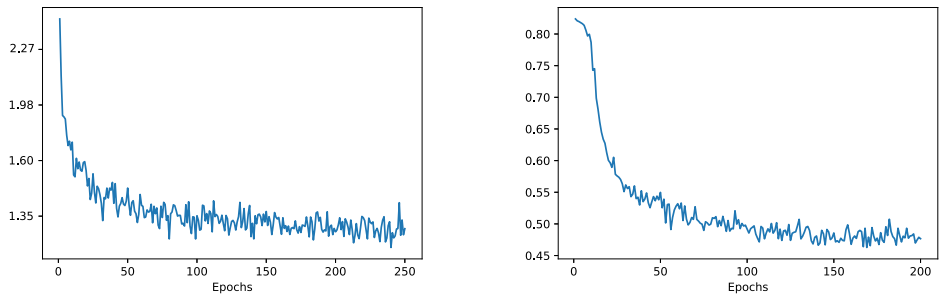


Fig. 9. Convergence performance on (a) PD and (b) RS databases with a missing rate of 0.5.

CRediT authorship contribution statement

Yurui Li: Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mingjing Du:** Writing – original draft, Validation, Supervision, Project administration, Funding acquisition. **Xiang Jiang:** Writing – review & editing, Validation. **Nan Zhang:** Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Nos. 62006104, 62006076) and the Key Science and Technology Innovation Project of Wenzhou (No. ZF2024002).

Data availability

Data will be made available on request.

References

[1] S. Majumdar, A.K. Laha, Corrigendum to "Clustering and classification of time series using topological data analysis with applications to finance", Expert Syst. Appl. 166 (2021) 114140.
[2] S. Li, P. Zhang, W. Chen, L. Ye, K.W. Brannan, N.-T. Le, J.-i. Abe, J.P. Cooke, G. Wang, A relay velocity model infers cell-dependent RNA velocity, Nature Biotechnol. 42 (1) (2024) 99–108.
[3] Z. Zhou, W. Tang, M. Li, W. Cao, Z. Yuan, A novel hybrid intelligent SOPDEL model with comprehensive data preprocessing for long-time-series climate prediction, Remote Sens. 15 (7) (2023) 1951.
[4] Y. Zhu, B. Jiang, H. Jin, M. Zhang, F. Gao, J. Huang, T. Lin, X. Wang, Networked time-series prediction with incomplete data via generative adversarial network, ACM Trans. Knowl. Discov. Data 18 (5) (2024) 115:1–115:25.
[5] Q. Ma, S. Li, G.W. Cottrell, Adversarial joint-learning recurrent neural network for incomplete time series classification, IEEE Trans. Pattern Anal. Mach. Intell. 44 (4) (2022) 1765–1776.
[6] Y. Li, M. Du, W. Zhang, X. Jiang, Y. Dong, Feature weighting-based deep fuzzy C-means for clustering incomplete time series, IEEE Trans. Fuzzy Syst. (2024).
[7] W. Alahamade, I. Lake, C.E. Reeves, B. de la Iglesia, A multi-variate time series clustering approach based on intermediate fusion: A case study in air pollution data imputation, Neurocomputing 490 (2022) 229–245.
[8] Q. Ma, C. Chen, S. Li, G.W. Cottrell, Learning representations for incomplete time series clustering, in: Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI Press, 2021, pp. 8837–8846.
[9] J. Enes, R.R. Expósito, J.D. Fuentes, J.L. Cacheiro, J. Touriño, A pipeline architecture for feature-based unsupervised clustering using multivariate time series from HPC jobs, Inf. Fusion 93 (2023) 1–20.
[10] J. Wu, O. Wyman, Y. Tang, D. Pasini, W. Wang, Multi-view 3D reconstruction based on deep learning: A survey and comparison of methods, Neurocomputing 582 (2024) 127553.
[11] A. Kumar, J. Yadav, A review of feature set partitioning methods for multi-view ensemble learning, Inf. Fusion 100 (2023) 101959.
[12] G. He, H. Wang, S. Liu, B. Zhang, CSMVC: A multiview method for multivariate time-series clustering, IEEE Trans. Cybern. 52 (12) (2022) 13425–13437.
[13] N. Zhang, S. Sun, Multiview unsupervised shapelet learning for multivariate time series clustering, IEEE Trans. Pattern Anal. Mach. Intell. 45 (4) (2023) 4981–4996.
[14] J. Xu, Y. Ren, H. Tang, Z. Yang, L. Pan, Y. Yang, X. Pu, P.S. Yu, L. He, Self-supervised discriminative feature learning for deep multi-view clustering, IEEE Trans. Knowl. Data Eng. 35 (7) (2023) 7470–7482.
[15] C. Cui, Y. Ren, J. Pu, J. Li, X. Pu, T. Wu, Y. Shi, L. He, A novel approach for effective multi-view clustering with information-theoretic perspective, in: Proceedings of the 37th Conference on Neural Information Processing Systems, Vol. 36, 2023.
[16] X. Liu, M. Li, C. Tang, J. Xia, J. Xiong, L. Liu, M. Kloft, E. Zhu, Efficient and effective regularized incomplete multi-view clustering, IEEE Trans. Pattern Anal. Mach. Intell. 43 (8) (2021) 2634–2646.
[17] X. Han, F. Zhou, Z. Ren, X. Wang, X. You, View-specific anchors coupled tensorial bipartite graph learning for incomplete multi-view clustering, Inform. Sci. 664 (2024) 120335.
[18] Z. Wang, L. Li, X. Ning, W. Tan, Y. Liu, H. Song, Incomplete multi-view clustering via structure exploration and missing-view inference, Inf. Fusion 103 (2024) 102123.
[19] W. Xia, Q. Gao, Q. Wang, X. Gao, Tensor completion-based incomplete multiview clustering, IEEE Trans. Cybern. 52 (12) (2022) 13635–13644.
[20] J. Yin, S. Sun, Incomplete multi-view clustering with reconstructed views, IEEE Trans. Knowl. Data Eng. 35 (3) (2023) 2671–2682.
[21] X. Liu, X. Zhu, M. Li, L. Wang, E. Zhu, T. Liu, M. Kloft, D. Shen, J. Yin, W. Gao, Multiple kernel k-means with incomplete kernels, IEEE Trans. Pattern Anal. Mach. Intell. 42 (5) (2020) 1191–1204.
[22] J. Yin, S. Sun, Incomplete multi-view clustering with cosine similarity, Pattern Recognit. 123 (2022) 108371.
[23] S. Deng, J. Wen, C. Liu, K. Yan, G. Xu, Y. Xu, Projective incomplete multi-view clustering, IEEE Trans. Neural Netw. Learn. Syst. (2023).
[24] J. Wen, Z. Zhang, Z. Zhang, L. Zhu, L. Fei, B. Zhang, Y. Xu, Unified tensor framework for incomplete multi-view clustering and missing-view inferring, in: Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI Press, 2021, pp. 10273–10281.
[25] Z. Lv, Q. Gao, X. Zhang, Q. Li, M. Yang, View-consistency learning for incomplete multiview clustering, IEEE Trans. Image Process. 31 (2022) 4790–4802.
[26] Z. Li, C. Tang, X. Zheng, X. Liu, W. Zhang, E. Zhu, High-order correlation preserved incomplete multi-view subspace clustering, IEEE Trans. Image Process. 31 (2022) 2067–2080.
[27] C. Liu, J. Wen, Z. Wu, X. Luo, C. Huang, Y. Xu, Information recovery-driven deep incomplete multiview clustering network, IEEE Trans. Neural Netw. Learn. Syst. (2023).
[28] H. Cai, W. Huang, S. Yang, S. Ding, Y. Zhang, B. Hu, F. Zhang, Y. Cheung, Realize generative yet complete latent representation for incomplete multi-view learning, IEEE Trans. Pattern Anal. Mach. Intell. 46 (5) (2024) 3637–3652.
[29] J. Pu, C. Cui, X. Chen, Y. Ren, X. Pu, Z. Hao, P.S. Yu, L. He, Adaptive feature imputation with latent graph for deep incomplete multi-view clustering, in: M.J. Wooldridge, J.G. Dy, S. Natarajan (Eds.), Proceedings of the 38th AAAI Conference on Artificial Intelligence, AAAI Press, 2024, pp. 14633–14641.
[30] G. Xu, J. Wen, C. Liu, B. Hu, Y. Liu, L. Fei, W. Wang, Deep variational incomplete multi-view clustering: Exploring shared clustering structures, in: M.J. Wooldridge, J.G. Dy, S. Natarajan (Eds.), Proceedings of the 38th AAAI Conference on Artificial Intelligence, AAAI Press, 2024, pp. 16147–16155.
[31] K. Maninis, S. Popov, M. Nießner, V. Ferrari, Vid2CAD: CAD model alignment using multi-view constraints from videos, IEEE Trans. Pattern Anal. Mach. Intell. 45 (1) (2023) 1320–1327.

[32] L. Peng, Y. Yang, Z. Wang, Z. Huang, H.T. Shen, MRA-Net: Improving VQA via multi-modal relation attention network, IEEE Trans. Pattern Anal. Mach. Intell. 44 (1) (2022) 318–329.
[33] J. Xu, Y. Ren, G. Li, L. Pan, C. Zhu, Z. Xu, Deep embedded multi-view clustering with collaborative training, Inform. Sci. 573 (2021) 279–290.
[34] Q. Wang, Z. Ding, Z. Tao, Q. Gao, Y. Fu, Generative partial multi-view clustering with adaptive fusion and cycle consistency, IEEE Trans. Image Process. 30 (2021) 1771–1783.
[35] N. Liang, Z. Yang, Z. Li, W. Han, Incomplete multi-view clustering with incomplete graph-regularized orthogonal non-negative matrix factorization, Appl. Intell. 52 (13) (2022) 14607–14623.
[36] X. Liu, P. Song, Incomplete multi-view clustering via virtual-label guided matrix factorization, Expert Syst. Appl. 210 (2022) 118408.
[37] J. Xu, C. Li, L. Peng, Y. Ren, X. Shi, H.T. Shen, X. Zhu, Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering, IEEE Trans. Image Process. 32 (2023) 1354–1366.
[38] M. Shang, C. Liang, J. Luo, H. Zhang, Incomplete multi-view clustering by simultaneously learning robust representations and optimal graph structures, Inform. Sci. 640 (2023) 119038.
[39] T. Brüsch, M.N. Schmidt, T.S. Alstrøm, Multi-view self-supervised learning for multivariate variable-channel time series, in: D. Comminiello, M. Scarpiniti (Eds.), Proceedings of the 33rd IEEE International Workshop on Machine Learning for Signal Processing, IEEE, 2023, pp. 1–6.
[40] Y. Ren, J. Pu, Z. Yang, J. Xu, G. Li, X. Pu, S.Y. Philip, L. He, Deep clustering: A comprehensive survey, IEEE Trans. Neural Netw. Learn. Syst. (2024).
[41] Y. Lin, Y. Gou, Z. Liu, B. Li, J. Lv, X. Peng, COMPLETER: incomplete multi-view clustering via contrastive prediction, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 11174–11183.
[42] W. Peng, Z. Zhang, W. Dai, X. Fu, L. Liu, L. Liu, N. Yu, A multi-view comparative learning method for spatial transcriptomics data clustering, in: Proceedings of the 10th IEEE International Conference on Bioinformatics and Biomedicine, 2023, pp. 287–292.
[43] Z. Huang, H. Chen, Z. Wen, C. Zhang, H. Li, B. Wang, C. Chen, Model-aware contrastive learning: Towards escaping the dilemmas, in: Proceedings of the 40th International Conference on Machine Learning, Vol. 202, 2023, pp. 13774–13790.
[44] C. Niu, H. Shan, G. Wang, SPICE: semantic pseudo-labeling for image clustering, IEEE Trans. Image Process. 31 (2022) 7264–7278.
[45] W.V. Gansbeke, S. Vandenhende, S. Georgoulis, M. Proesmans, L.V. Gool, SCAN: learning to classify images without labels, in: Proceedings of the 16th European Conference, in: Lecture Notes in Computer Science, vol. 12355, 2020, pp. 268–285.
[46] Y. Li, P. Hu, J.Z. Liu, D. Peng, J.T. Zhou, X. Peng, Contrastive clustering, in: Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI Press, 2021, pp. 8547–8555.
[47] J.M. Giorgi, O. Nitski, B. Wang, G.D. Bader, DeCLUTR: Deep contrastive learning for unsupervised textual representations, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), Proceedings of the 11th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, 2021, pp. 879–895.
[48] S. Wu, Y. Zheng, Y. Ren, J. He, X. Pu, S. Huang, Z. Hao, L. He, Self-weighted contrastive fusion for deep multi-view clustering, IEEE Trans. Multimed. 26 (2024) 9150–9162.
[49] H. Chen, Y. Wang, B. Lagadec, A. Dantcheva, F. Brémond, Joint generative and contrastive learning for unsupervised person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 2004–2013.
[50] Y. Tian, D. Krishnan, P. Isola, Contrastive multiview coding, in: A. Vedaldi, H. Bischof, T. Brox, J. Frahm (Eds.), Proceedings of the 16th European Conference on Computer Vision, Vol. 12356, Springer, 2020, pp. 776–794.
[51] K. Hassani, A.H.K. Ahmadi, Contrastive multi-view representation learning on graphs, in: Proceedings of the 37th International Conference on Machine Learning, Vol. 119, PMLR, 2020, pp. 4116–4126.
[52] R. Lin, Y. Lin, Z. Lin, S. Du, S. Wang, CCR-Net: Consistent contrastive representation network for multi-view clustering, Inform. Sci. 637 (2023) 118937.
[53] Y. Lu, Q. Li, X. Zhang, Q. Gao, Deep contrastive representation learning for multi-modal clustering, Neurocomputing 581 (2024) 127523.
[54] J. Jin, S. Wang, Z. Dong, X. Liu, E. Zhu, Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment, in: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2023, pp. 11600–11609.
[55] Y. Ren, X. Chen, J. Xu, J. Pu, Y. Huang, X. Pu, C. Zhu, X. Zhu, Z. Hao, L. He, A novel federated multi-view clustering method for unaligned and incomplete data fusion, Inf. Fusion 108 (2024) 102357.
[56] Y. Ren, J. Pu, C. Cui, Y. Zheng, X. Chen, X. Pu, L. He, Dynamic weighted graph fusion for deep multi-view clustering, in: Proceedings of the 33rd International Joint Conference on Artificial Intelligence, ijcai.org, 2024, pp. 4842–4850.
[57] Y. Wang, D. Chang, Z. Fu, J. Wen, Y. Zhao, Graph contrastive partial multi-view clustering, IEEE Trans. Multimed. 25 (2023) 6551–6562.
[58] Z. Shu, B. Li, C. Mao, S. Gao, Z. Yu, Structure-guided feature and cluster contrastive learning for multi-view clustering, Neurocomputing 582 (2024) 127555.
[59] B. Peng, G. Lin, J. Lei, T. Qin, X. Cao, N. Ling, Contrastive multi-view learning for 3D shape clustering, IEEE Trans. Multimed. 26 (2024) 6262–6272.
[60] Y.H. Tsai, Y. Wu, R. Salakhutdinov, L. Morency, Self-supervised learning from a multi-view perspective, in: Proceedings of the 9th International Conference on Learning Representations, OpenReview.net, 2021.
[61] H. Wang, Q. Wang, Q. Miao, X. Ma, Joint learning of data recovering and graph contrastive denoising for incomplete multi-view clustering, Inf. Fusion 104 (2024) 102155.
[62] W. Lv, C. Zhang, H. Li, X. Jia, C. Chen, Joint projection learning and tensor decomposition based incomplete multi-view clustering, 2023, arXiv preprint arXiv:2310.04038.
[63] C. Zhang, H. Li, C. Chen, X. Jia, C. Chen, Low-rank tensor regularized views recovery for incomplete multiview clustering, IEEE Trans. Neural Netw. Learn. Syst. 35 (7) (2024) 9312–9324.
[64] J. Wen, Z. Wu, Z. Zhang, L. Fei, B. Zhang, Y. Xu, Structural deep incomplete multi-view clustering network, in: G. Demartini, G. Zuccon, J.S. Culpepper, Z. Huang, H. Tong (Eds.), Proceedings of the 30th ACM International Conference on Information and Knowledge Management, ACM, 2021, pp. 3538–3542.
[65] J. Xu, C. Li, Y. Ren, L. Peng, Y. Mo, X. Shi, X. Zhu, Deep incomplete multi-view clustering via mining cluster complementarity, in: Proceedings of the 36t AAAI Conference on Artificial Intelligence, AAAI Press, 2022, pp. 8761–8769.
[66] J. Gui, Z. Sun, Y. Wen, D. Tao, J. Ye, A review on generative adversarial networks: Algorithms, theory, and applications, IEEE Trans. Knowl. Data Eng. 35 (4) (2023) 3313–3332.
[67] Y. Lin, Y. Gou, X. Liu, J. Bai, J. Lv, X. Peng, Dual contrastive prediction for incomplete multi-view representation learning, IEEE Trans. Pattern Anal. Mach. Intell. 45 (4) (2023) 4447–4461.
[68] G. Chao, Y. Jiang, D. Chu, Incomplete contrastive multi-view clustering with high-confidence guiding, in: Proceedings of the 38th AAAI Conference on Artificial Intelligence, Vol. 38, 2024, pp. 11221–11229.
[69] W. Xia, Q. Wang, Q. Gao, X. Zhang, X. Gao, Self-supervised graph convolutional network for multi-view clustering, IEEE Trans. Multimed. 24 (2021) 3182–3192.
[70] J. Xu, H. Tang, Y. Ren, L. Peng, X. Zhu, L. He, Multi-level feature learning for contrastive multi-view clustering, in: Proceedings of the 32nd Conference on Computer Vision and Pattern Recognition, 2022, pp. 16030–16039.
[71] Y. Zhang, Q. Huang, B. Zhang, S. He, T. Dan, H. Peng, H. Cai, Deep multiview clustering via iteratively self-supervised universal and specific space learning, IEEE Trans. Cybern. 52 (11) (2022) 11734–11746.
[72] J. de Jong, M.A. Emon, P. Wu, R. Karki, M. Sood, P. Godard, A. Ahmad, H. Vrooman, M. Hofmann-Apitius, H. Fröhlich, Deep learning for clustering of multivariate clinical patient trajectories with missing values, GigaScience 8 (11) (2019) giz134.