Mask the Redundancy: Evolving Masking Representation Learning for Multivariate Time-Series Clustering

Zexi Tan, Xiaopeng Luo, Yunlin Liu, Yiqun Zhang*

School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China
3123004194@mail2.gdut.edu.cn, luoxiaopeng@mails.gdut.edu.cn, lyunlin2713@gmail.com, yqzhang@gdut.edu.cn

Abstract

Multivariate Time-Series (MTS) clustering discovers intrinsic grouping patterns of temporal data samples. Although time-series provide rich discriminative information, they also contain substantial redundancy, such as steady-state machine operation records and zero-output periods of solar power generation. Such redundancy diminishes the attention given to discriminative timestamps in representation learning, thus leading to performance bottlenecks in MTS clustering. Masking has been widely adopted to enhance the MTS representation, where temporal reconstruction tasks are designed to capture critical information from MTS. However, most existing masking strategies appear to be standalone preprocessing steps, isolated from the learning process, which hinders dynamic adaptation to the importance of clustering-critical timestamps. Accordingly, this paper proposes the Evolving-masked MTS Clustering (EMTC) method, whose model architecture comprises Importance-aware Variate-wise Masking (IVM) and Multi-Endogenous Views (MEV) generation modules. IVM adaptively guides the model in learning more discriminative representations for clustering, while the reconstruction and cluster-guided contrastive learning pathways enhance and connect the representation learning to clustering tasks. Extensive experiments on 15 benchmark datasets demonstrate the superiority of EMTC over eight SOTA methods, where the EMTC achieves an average improvement of 4.85% in F1-Score over the strongest baselines.

Code — https://github.com/yueliangy/EMTC

Introduction

Multivariate Time-Series (MTS) clustering (Li and Liu 2021; Zhang and Sun 2023) is a pivotal unsupervised data analysis task (Zhang et al. 2025; Tan et al. 2025) for discovering intrinsic patterns of temporal data, which are common in domains like activity recognition (Ma et al. 2021), industrial monitoring (Alwan et al. 2022), and medical data analysis (Xie et al. 2025). Although MTS contains rich trend and periodic information (Zhang et al. 2023b), widespread redundant timestamps may obscure the representation of key distribution patterns, thereby compromising the formation of compact and meaningful cluster structures. Conventional approaches implicitly mitigate MTS redundancy through hand-crafted feature engineering, dimensionality reduction (Yang and Shahabi 2004; Li 2019), and elastic time-series alignment (Li and Wei 2020; Cai et al. 2025). Despite their effectiveness in certain scenarios, they often impose strong temporal shape or sample distribution assumptions, and fail to support end-to-end representation learning.

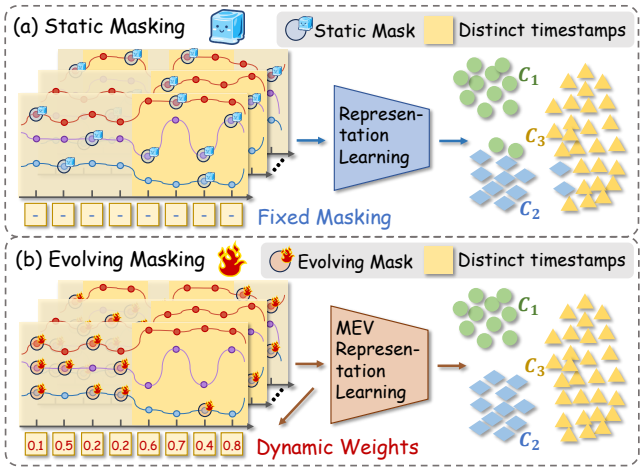

Figure 1: **(a) Static Masking vs. (b) Evolving Masking (ours).** Static masking is a fixed MTS preprocessing strategy to enhance the downstream representation learning. In contrast, our evolving masking dynamically adapts to the representation learning and clustering objective through attention-based timestamp weighting, simultaneously surpassing the MTS redundancy and enhancing the cluster discrimination of representations.

Deep learning paradigms (Lafabregue et al. 2022; Ienco and Interdonato 2023) for MTS clustering usually learn representations and then feed the outputs to algorithms like k-means (MacQueen 1967; Ikotun et al. 2023) to obtain the final clustering results. In this stream, the Autoencoder-based frameworks (Ienco and Interdonato 2020; Li et al. 2023b) are common, which derive latent high-quality embeddings through reconstruction-guided model training. However, their emphasis on reconstruction fidelity may act to retain the MTS redundancy. Moreover, since the clustering objective is usually separated from the representation learn-

*Corresponding Author

ing, the learning process fails to suppress the redundancy from the clustering-friendly perspective. Accordingly, contrastive learning methods (Wu et al. 2024; Li et al. 2025; Wang et al. 2025) have been employed to learn discriminative features by contrasting positive and negative instance pairs. Although it has been proven that the self-supervised contrastive learning can implicitly obtain clustering-friendly representations (Ben-Shaul et al. 2023), their effectiveness critically depends on the construction of instance relationships. That is, when the contrastive strategies do not align with the sought cluster distributions, the learned representations may preserve or even amplify temporal redundancies.

To explicitly address the temporal redundancy, attention (Ienco and Interdonato 2020) has been adopted to perform soft redundancy filtering through dynamic timestamp weighting (Ding, Sun, and Zhao 2023). Although this type of method considerably relieves the impact of redundancy and enhances the representation of timestamps, it also inherently preserves the full input structure containing the redundant information. More importantly, the learned attention weights may be misled by highly activated yet non-informative patterns, thus failing in prioritizing sparse-but-critical features. In contrast, masking mechanisms offer a more structured approach to redundancy suppression in MTS representation learning (Ashok et al. 2024; Eldele et al. 2024; Zhang et al. 2024). Conventional static masking (Fu and Xue 2022; Zhang et al. 2023a) compels models to learn from partial observations through reconstruction. Although masking has been well validated in acquiring general representations, the fixed schemes are incompetent in task adaptation as demonstrated in Figure 1 (a). Most recently, dynamic masking (Li et al. 2023a; Shi et al. 2025) has been developed to learn the masking. Nevertheless, learning masks for MTS redundancy suppression and structural discriminative representation remains underexplored.

This paper, therefore, proposes Evolving-masked MTS Clustering (EMTC) to explicitly address the critical redundancy issue in MTS clustering. As illustrated in Figure 1 (b), EMTC adopts attention-based Importance-aware Variate-wise Masking (IVM) to dynamically determine the redundant timestamps to be excluded in the current learning epoch, and introduces Multi-Endogenous Views (MEV)-based representation learning to facilitate a more informative and generalized learning process. More specifically, the IVM module assigns masks to timestamps with low importance to explicitly prevent them from participating in the representation learning, where the attention-based importance is learned for each time series. Then, the MEV is generated based on the masked MTS, and a dual-path architecture is designed to learn representations based on the MEV: 1) Consistency and Reconstruction Learning (CRL) performs intra- and inter-view reconstruction to extract invariant features that are robust to masked redundancy and view-specific information, respectively, and 2) Clustering-guided MEV Contrastive learning (CMC) adopts clusters as a basis for data augmentation, which serves to enhance the cluster separation in the embedding space and also connects the clustering objective into representation learning. The contributions are as follows:

- **A novel learnable redundancy masking mechanism.** This paper designs a content-aware attention mechanism for timestamp scoring, which dynamically guides the evolving masking of redundant timestamps. This directly and adaptively suppresses temporal redundancy, enhancing representation discriminability for MTS clustering.
- **Synergistic design of IVM-MEV complementary.** A complementary redundancy masking and multi-view learning mechanism has been established, where MEV facilitates sufficient interaction of the MTS to avoid the dominance of the crisp IVM masking, while IVM serves to eliminate the redundancy amplified by the MEV.
- **Representation and clustering connected paradigm.** EMTC is an early exploration that integrates contrastive learning into MTS clustering. By leveraging dynamically updated clustering results to guide data augmentation, it connects both the representation learning and the redundancy masking learning to the MTS clustering objective.

Related Work

Redundancy Suppression in MTS Clustering. Traditional MTS clustering tackles redundancy via feature engineering or linear dimensionality reduction (e.g., PCA variants) (Yang and Shahabi 2004; Barragan, Fontes, and Embiruçu 2016; Li 2019; Han and Woo 2022), projecting data into lower-dimensional spaces but constrained by linear assumptions. Dynamic Time Warping (DTW) methods (Li and Wei 2020; Shen and Chi 2023; Cai et al. 2025) offer adaptive alignment but lack dynamic redundancy suppression. Sample partition algorithms like k-means (MacQueen 1967; Sinaga and Yang 2020; Ikotun et al. 2023) and hierarchical clustering (Hamerly and Elkan 2003; Wang, Wirth, and Wang 2007; Meng et al. 2023) rely on predefined metrics and are sensitive to initialization. Deep learning paradigms learn discriminative representations end-to-end. Autoencoder frameworks like DeTSEC (Ienco and Interdonato 2020) and RDDC (Trosten et al. 2019) derive compact embeddings, while contrastive methods (TimesURL (Liu and Chen 2024), FCACC (Wang et al. 2025), MVCIMTS (Li et al. 2025)) enhance feature distinctiveness through augmented views, though efficacy depends on augmentation design. Other innovations include CDCC (Peng et al. 2024a) for cross-domain contrastive learning, TimeDRL (Chang et al. 2024) for disentangled representation learning, and k-Graph (Boniol et al. 2025) for interpretability. Although competent in capturing complex temporal structures, most lack explicit dynamic mechanisms to suppress evolving temporal redundancy.

Masking Strategies for MTS Analysis. The principle of masked learning demonstrates significant benefits across domains like computer vision and NLP, enhancing efficiency and precision. For instance, DynaMask (Li et al. 2023a) dynamically selects optimal mask resolution in instance segmentation, while Trainable Dynamic Mask Sparse Attention (Shi et al. 2025) achieves adaptive sparsity for long-context modeling. Inspired by these advances, MTS representation learning has adopted masking to mitigate redundancy. Methods like Ti-MAE (Li et al. 2023b) use masked

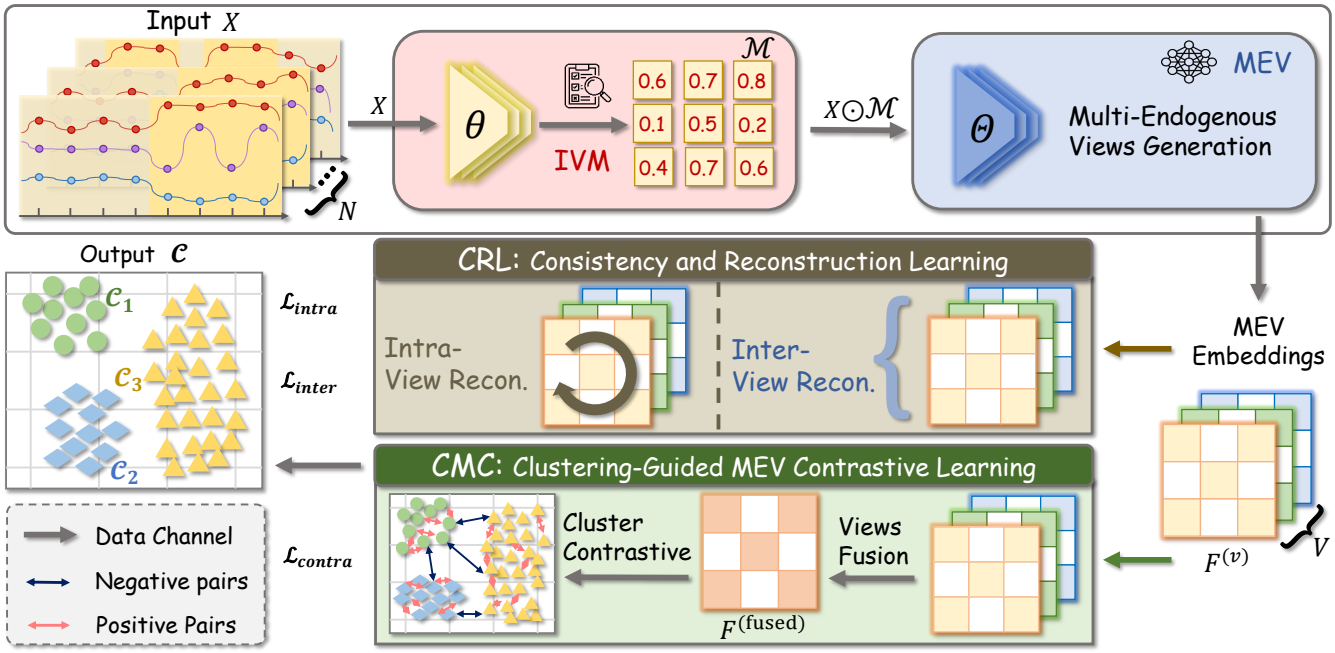

Figure 2: Overview of the proposed EMTC framework.

autoencoding for forecasting alignment, PrimeNet (Chowdhury et al. 2023) implements density-aware masking for irregular sampling, and TS-MVP (Zhong et al. 2023) employs probability-based masking for multi-view alignment. Others including STCR (Lee, Kim, and Son 2024) and TS2Vec (Yue et al. 2022) incorporate random masking in contrastive frameworks. However, these MTS methods predominantly rely on static or predetermined masking policies that lack adaptability to co-evolve with learning, hindering dynamic suppression of evolving redundant patterns or prioritization of cluster-salient features.

Proposed Method

Existing masking strategies are limited by their inability to adapt to MTS clustering tasks. To address this, this paper proposes **EMTC** with two core components: **IVM** and **MEV** modules. As illustrated in Figure 2, learning of the model involves two pathways:

- **CRL**: Dynamically suppresses temporal redundancy while learning robust multi-view representations.
- **CMC**: Structures embedding space under cluster-aware contrastive objectives to enhance cluster separability.

Let $X_i = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T] \in \mathbb{R}^{T \times D}$ denotes an MTS sample of length T with D variates, where each $\mathbf{x}_t \in \mathbb{R}^D$ represents the multivariate observation at timestamp t. Given a dataset $X = \{X_i\}_{i=1}^N$, the objective is to partition the N samples into g disjoint clusters $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_g\}$.

IVM: Importance-aware Variate-wise Masking

To explicitly facilitate dynamic redundancy suppression, learnable IVM is designed to adapt the masking to the MTS clustering tasks. The clustering-friendly evolving masking is realized in three stages: single-variate view generation, content-aware importance evaluation, and redundant timestamp masking.

Single-Variate View Generation: To facilitate content-aware importance evaluation, we first generate single-variate view embeddings from the original input X:

$$Z^{(d)} = f_\theta^{(d)}(X), \quad \forall d \in \{1, 2, \ldots, D\}, \tag{1}$$

where $Z^{(d)}$ provides the single-variate embedding that preserves the integrity of individual variate information.

Content-Aware Importance Evaluation: Based on the single-variate embeddings, the importance of each timestamp of sample $Z_i^{(d)}$ is evaluated by content-aware attention:

$$S_i^{(d)} = \text{Softmax}\left(\frac{Q_i^{(d)}(K_i^{(d)})^T}{\sqrt{d_k}}\right) Z_i^{(d)}, \tag{2}$$

where the attention mechanism optimized with the clustering objective can dynamically refine its focus on discriminative timestamps to enhance inter-cluster separation. Unlike conventional attention-weighted MTS that are often misled by highly activated yet non-informative patterns, our attention-based masking directly excludes the timestamps that are relatively clustering-irrelevant, ensuring dynamic alignment with learned cluster semantics, achieving precise and task-aware MTS redundancy suppression with the CRL and CMC dual-learning pathways.

Redundant Timestamp Masking: With the attention-based importance evaluation, thresholding is employed to mask less-important timestamps:

$$M_i^{(d)}(t) = \begin{cases} 1 & \text{if } S_i^{(d)}(t) \geq \epsilon \\ 0 & \text{otherwise} \end{cases}, \quad \forall t \in [1, T], \tag{3}$$

where ϵ is a predefined threshold to filter out redundant timestamps. $M_i^{(d)}(t)$ denotes the binary mask value at timestamp t. The computed mask set $\mathcal{M}_i = \{M_i^{(1)}, M_i^{(2)}, \ldots, M_i^{(D)}\} \in \mathbb{R}^{T\times D}$ is then applied to D variates of the i-th original sample X_i through element-wise multiplication $X \odot \mathcal{M}$ to obtain the masked input $\widetilde{X}$ in the next epoch. In the first learning epoch, a randomly initialized mask set $\mathcal{M} \in \mathbb{R}^{N\times T\times D}$ is adopted to obtain $\widetilde{X}$.

MEV: Multi-Endogenous Views Generation

To overcome the limitations of single-view representations in capturing complex multivariate patterns, this paper proposes to perform MEV generation to obtain complementary MTS perspectives. That is, learning upon the multi-view MTS ensures robust representation learning while providing natural regularization for the evolving masking. Specifically, the masked input $\widetilde{X}$ obtained from IVM is processed through MEV embeddings generation:

$$F^{(v)} = f_\Theta^{(v)}\left(\widetilde{X}\right), \quad \forall v \in \{1, 2, \ldots, V\}, \tag{4}$$

where the $F^{(v)}$ denotes the v-th endogenous view. Such an MEV design ensures comprehensive pattern capture while maintaining the integrity of view-specific information for masking decisions.

With the endogenous views obtained based on the IVM masking, the model performs a dual-path representation learning, comprising CRL for IVM consolidation and CMC for cluster discriminative representation enhancement, which are introduced below.

CRL: Consistency and Reconstruction Learning

CRL involves two complementary mechanisms: intra-view reconstruction learning helps the MEV retain the semantic structure of the original time-series, while inter-view reconstruction learning establishes semantic bridges between different views. For each view $F^{(v)}$, the model reconstructs the original MTS X through $X^{(v)} = R^{(v)}(F^{(v)})$, where the decoder $R^{(v)}$ maps the endogenous view embedding $F^{(v)}$ to the reconstructed sequence $X^{(v)}$. The intra-view reconstruction learning can then be formulated as:

$$\mathcal{L}_{\text{intra}} = \sum_{v=1}^{V} \left\| X - X^{(v)} \right\|_F^2. \tag{5}$$

Such self-reconstruction acts as a regularizer to retain the semantic structure during modal training and preserve discriminative temporal features for clustering.

By contrast, inter-view reconstruction establishes inter-view semantic consistency to enhance representation robustness by measuring the discrepancy between the original and transformed embeddings across views:

$$\mathcal{L}_{\text{inter}} = \sum_{i=1}^{V} \sum_{j=i+1}^{V} \mathcal{L}_{\text{inter}}^{i\to j} + \mathcal{L}_{\text{inter}}^{j\to i}, \tag{6}$$

specifically, for each pair of endogenous view embeddings $(F^{(i)}, F^{(j)})$, the loss between $F^{(j)}$ and the transformed embedding $F^{(i\to j)} = \mathcal{H}_{i\to j}(F^{(i)})$ is computed, where $\mathcal{H}_{i\to j}$ is a transformation decoder that maps embeddings from view $F^{(i)}$ to view $F^{(j)}$ as:

$$\mathcal{L}_{\text{inter}}^{i\to j} = \left\| F^{(j)} - F^{(i\to j)} \right\|_F^2. \tag{7}$$

CRL simultaneously considers single-view fidelity and multi-view consistency to learn representations that are both consistent with the key features of the masked MTS and semantically coherent across endogenous view embeddings. More importantly, the reconstruction learning and consistency learning based on MEV mitigate the risk of overfitting to variate-specific knowledge and losing critical single temporal information caused by the crisp IVM masking.

CMC: Clustering-Guided MEV Contrastive Learning

To explicitly structure the embedding space for cluster separation, CMC is also introduced to leverage cluster assignments as supervisory signals. The process begins by aggregating the complementary information from all MEV embeddings. Specifically, temporal pooling is applied to each endogenous view embedding $F^{(v)}$, followed by an MEV fusion operation, yielding a fused representation:

$$F^{(\text{fused})} = F^{(1)} \circledast F^{(2)} \circledast \cdots \circledast F^{(V)}, \tag{8}$$

where $\circledast$ denotes the fusion operation that integrates information across all views. The integrated representation $F^{(\text{fused})}$ combines discriminative patterns from different views, forming a robust basis for clustering. For simplicity but without losing generality, $F^{(\text{fused})}$ is denoted as F hereinafter. Subsequently, cluster labels $\mathcal{C}$ are obtained by performing k-means clustering on F at each training epoch:

$$\mathcal{C} = \text{Cluster}(F). \tag{9}$$

These dynamically updated cluster labels guide the construction of pairs for contrastive learning. For an anchor sample F_i, samples sharing the same cluster label form positive pairs, with all the positive samples denoted as $\mathcal{P}(i^+)$. The other samples outside the cluster of F_i serve as negative pairs, which can be written as $\mathcal{N}(i^-)$.

The discriminative contrastive objective, adapted from the foundation laid in (Yang et al. 2023; Peng et al. 2024b), is utilized to maximize the cosine similarity within each cluster (positive pairs) and enforce robust separation between distinct clusters (negative pairs), which can be expressed as:

$$\mathcal{L}_{\text{contra}} = -\sum_{i=1}^{N} \log \frac{L_i^+}{L_i^+ + L_i^-}. \tag{10}$$

L_i^+ is the loss contributed by the positive pairs w.r.t. F_i, which can be written as:

$$L_i^+ = \sum_{\mathcal{P}(i^+)} \exp\left(\frac{\text{sim}(i, i^+)}{\tau}\right), \tag{11}$$

where τ serves as a temperature parameter controlling the sharpness of the similarity distribution, and $\text{sim}\,(i, i^+)$ denotes the cosine similarity between positive sample pairs,

Algorithm 1: EMTC: Evolving-masked MTS Clustering.

Input: MTS dataset X, cluster number g.

Initialize IVM and MEV components.
repeat
  IVM: Update mask set via Eqs. (1-3);
  MEV: Generate embedding $F^{(v)}$ via Eq. (4);
  CRL: Compute $\mathcal{L}_{\text{intra}}$ and $\mathcal{L}_{\text{inter}}$ via Eqs. (5-7);
  Update cluster labels $\mathcal{C}$ via Eq. (9);
  CMC: Compute $\mathcal{L}_{\text{contra}}$ via Eqs. (10) and (12);
  Compute total loss $\mathcal{L}_{\text{total}}$ for Adam optimization;
until convergence

Output: Final MTS clustering results $\mathcal{C}$.

which can be instantiated as:

$$\text{sim}(i, i^+) = \frac{F_i^\top F_{i^+}}{\|F_i\| \, \|F_{i^+}\|} \text{ s.t. } F_{i^+} \in \mathcal{P}(i^+). \tag{12}$$

The definitions for negative pairs, i.e., L_i^- and $\text{sim}(i, i^-)$, are analogous.

Such a dynamic clustering-guided representation learning strategy ensures a close connection between the contrastive learning process and the clustering objective. By continuously updating the cluster labels throughout the training, the model can effectively refine the MTS representation to fit newly learned cluster distributions.

Model Training and Clustering

The overall loss $\mathcal{L}_{\text{total}}$ with previously introduced loss terms integrated through balancing coefficients, i.e., α for $\mathcal{L}_{\text{intra}}$ and β for $\mathcal{L}_{\text{inter}}$, can be written as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{contra}} + \alpha \mathcal{L}_{\text{intra}} + \beta \mathcal{L}_{\text{inter}}. \tag{13}$$

The model parameters are optimized using the Adam optimizer. The full EMTC algorithm for MTS representation learning and clustering is summarized as Algorithm 1.

Experiments

Eight experiments have been designed to comprehensively evaluate the proposed EMTC: 1) **Clustering Performance Comparison** with eight SOTA methods on 15 datasets; 2) **Significance Test** using BD-test with 95% and 99% confidence intervals; 3) **Key Component Ablation** validating IVM and MEV modules; 4) **Loss Ablation** validating each loss term; 5) **Masking Strategies Comparison** illustrating the rationality of our IVM design; 6) **Convergence Analysis** tracking loss dynamics; 7) **Efficiency Evaluation** measuring computational requirements and scalability; 8) **Cluster Effect Visualization** using t-SNE for qualitative assessment. Due to space constraints, partial results of 1) **Clustering Performance Comparison** and 3) **Key Component Ablation**, and the whole results of 2) **Significance Test** and 4) **Loss Ablation**, are provided in the Extended Version.

Extended Version — http://arxiv.org/abs/2511.17008

15 real-world benchmark datasets from the UEA MTS archive (Bagnall et al. 2018), spanning diverse domains with varying number of samples, sequence lengths and feature dimensions. Detailed statistics are presented in Table 1. **Four standard metrics** are employed: Accuracy (ACC), F1-Score (F1), Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI). **Eight SOTA counterparts** are compared, including deep MTS clustering methods FEI (Fu and Hu 2025), FCACC (Wang et al. 2025), TimesURL (Liu and Chen 2024), UNITS (Gao et al. 2024), USLA (Zhang and Sun 2023), Ti-MAE (Li et al. 2023b), MHCCL (Meng et al. 2023), and conventional feature engineering method T-GMRF (Ding et al. 2023). All counterparts are tuned following their original papers, with parameters optimized for each dataset individually. Our architecture utilizes dilated convolution for encoders $f_\Theta^{(v)}$ and $f_\theta^{(d)}$, while employing MLP decoder with specific configurations: six-layer networks for intra-view reconstruction $(R^{(v)})$ and four-layer networks for inter-view transformation $(\mathcal{H}_{i\to j})$. The fused representation $F^{(\text{fused})}$ is obtained through the fusion operation $\circledast$ that averages the sum of all $F^{(v)}$, which can be extended to weighted combinations. We set the batch size to 64 and train for a maximum of 200 epochs. All experiments are implemented in PyTorch 1.8.0 on a NVIDIA RTX4090 GPU, 20GB RAM.

No.	Datasets	N	T	D	g
1	BasicMotions	40	100	6	4
2	Cricket	72	1197	6	12
3	DuckDuckGeese	40	270	1345	5
4	EigenWorms	131	17984	6	5
5	Epilepsy	138	206	3	4
6	FingerMovements	100	50	28	2
7	HandMovementDirection	147	400	10	4
8	Heartbeat	205	405	61	2
9	MotorImagery	100	3000	64	2
10	NATOPS	180	51	24	6
11	PEMS-SF	173	144	963	7
12	RacketSports	152	30	6	4
13	SelfRegulationSCP1	293	896	6	2
14	SelfRegulationSCP2	180	1152	7	2
15	StandWalkJump	15	2500	4	3

Table 1: Statistics of datasets. g indicates the 'true' number of clusters provided by the labels of the datasets.

Clustering Performance Evaluation

As Table 2 shown, EMTC demonstrates superior clustering performance compared to eight baselines across 15 benchmark datasets, evaluated by ACC and F1 metrics. Achieving the highest average rank, EMTC consistently outperforms all baselines on most datasets and metrics. Notably, it exhibits substantial improvements over the strongest competitors, i.e., FCACC and FEI. On long-sequence datasets like StandWalkJump, EMTC achieves significant performance gains. For datasets with balanced class distributions such as MotorImagery, it delivers notable enhancements. On multi-class datasets such as Cricket (12 clusters), EMTC maintains competitive performance, further validating its robustness across diverse clustering scenarios.

Datasets	Metrics	Methods								
		EMTC (ours)	FEI (AAAI'25)	FCACC (arxiv'25)	TimesURL (AAAI'24)	UNITS (NeurIPS'24)	USLA (TPAMI'23)	Ti-MAE (arxiv'23)	MHCCL (AAAI'23)	T-GMRF (TKDE'23)
BasicMotions	ACC	0.9083 ± 0.0656	0.4667 ± 0.0589	0.7250 ± 0.0000	0.6917 ± 0.0118	0.7667 ± 0.1662	0.8750 ± 0.0000	0.9750 ± 0.0000	0.1750 ± 0.0354	0.3250 ± 0.0000
	F1	0.9057 ± 0.0675	0.4975 ± 0.1143	0.8713 ± 0.0000	0.8446 ± 0.0103	0.8665 ± 0.0961	0.3801 ± 0.1374	0.9749 ± 0.0000	0.2002 ± 0.0311	0.2375 ± 0.0000
Cricket	ACC	0.5972 ± 0.0227	0.4537 ± 0.0755	0.4259 ± 0.0429	0.4722 ± 0.0196	0.3287 ± 0.0236	0.3611 ± 0.0000	0.4832 ± 0.0226	0.0648 ± 0.0398	0.1389 ± 0.0000
	F1	0.6317 ± 0.0366	0.5136 ± 0.0501	0.4301 ± 0.0392	0.4913 ± 0.0134	0.3556 ± 0.0187	0.0573 ± 0.0367	0.4903 ± 0.0124	0.0600 ± 0.0321	0.1102 ± 0.0000
DuckDuckGeese	ACC	0.4800 ± 0.0163	0.2933 ± 0.0249	0.3550 ± 0.0087	0.3133 ± 0.0340	0.3667 ± 0.0094	0.3200 ± 0.0000	0.3733 ± 0.0189	0.1933 ± 0.0249	0.2600 ± 0.0000
	F1	0.4917 ± 0.0357	0.2445 ± 0.0185	0.3300 ± 0.0070	0.3065 ± 0.0294	0.3980 ± 0.0095	0.1457 ± 0.0169	0.3602 ± 0.0224	0.1447 ± 0.0241	0.1758 ± 0.0000
EigenWorms	ACC	0.4707 ± 0.0095	0.4580 ± 0.0125	0.4530 ± 0.0125	0.4633 ± 0.0134	0.3410 ± 0.0401	0.3486 ± 0.0036	0.4540 ± 0.0221	0.3486 ± 0.0036	0.4540 ± 0.0225
	F1	0.3713 ± 0.0549	0.3404 ± 0.0779	0.3333 ± 0.0224	0.3204 ± 0.0789	0.2894 ± 0.0168	0.2245 ± 0.0625	0.3401 ± 0.0023	0.2245 ± 0.0625	0.3404 ± 0.0023
Epilepsy	ACC	0.5556 ± 0.0239	0.4082 ± 0.0208	0.5236 ± 0.0107	0.4517 ± 0.0034	0.4179 ± 0.0304	0.5338 ± 0.0034	0.5145 ± 0.0000	0.2053 ± 0.0754	0.4928 ± 0.0000
	F1	0.5519 ± 0.0276	0.4077 ± 0.0065	0.5193 ± 0.0093	0.5226 ± 0.0048	0.4116 ± 0.0414	0.1491 ± 0.0204	0.5133 ± 0.0001	0.2017 ± 0.0712	0.3467 ± 0.0000
Finger Movements	ACC	0.5967 ± 0.0094	0.5100 ± 0.0141	0.5000 ± 0.0000	0.5100 ± 0.0000	0.5800 ± 0.0535	0.5000 ± 0.0000	0.5800 ± 0.0000	0.5133 ± 0.0377	0.5200 ± 0.0163
	F1	0.5892 ± 0.0151	0.5071 ± 0.0146	0.3503 ± 0.0000	0.4954 ± 0.0000	0.5771 ± 0.0514	0.4770 ± 0.0001	0.5758 ± 0.0000	0.4867 ± 0.0422	0.4028 ± 0.0679
HandMovement Direction	ACC	0.4640 ± 0.0064	0.3333 ± 0.0169	0.4696 ± 0.0112	0.3514 ± 0.0191	0.3514 ± 0.0110	0.3649 ± 0.0000	0.3243 ± 0.0000	0.2523 ± 0.0337	0.4189 ± 0.0000
	F1	0.3737 ± 0.0262	0.3071 ± 0.0601	0.3993 ± 0.0192	0.3434 ± 0.0194	0.3222 ± 0.0164	0.2717 ± 0.0475	0.3186 ± 0.0047	0.2293 ± 0.0502	0.2661 ± 0.0000
Heartbeat	ACC	0.7463 ± 0.0040	0.7138 ± 0.0023	0.5646 ± 0.0040	0.7122 ± 0.0000	0.7122 ± 0.0000	0.7138 ± 0.0023	0.7122 ± 0.0000	0.4341 ± 0.0000	0.6911 ± 0.0506
	F1	0.6134 ± 0.0329	0.8330 ± 0.0016	0.5549 ± 0.0051	0.8319 ± 0.0000	0.8319 ± 0.0000	0.6014 ± 0.0011	0.4160 ± 0.0000	0.4325 ± 0.0000	0.5952 ± 0.0303
MotorImagery	ACC	0.6500 ± 0.0216	0.5100 ± 0.0000	0.5700 ± 0.0122	0.5200 ± 0.0000	0.5500 ± 0.0327	0.5100 ± 0.0000	0.5100 ± 0.0000	0.5233 ± 0.0660	0.5500 ± 0.0141
	F1	0.6452 ± 0.0213	0.3924 ± 0.0263	0.5572 ± 0.0232	0.3912 ± 0.0000	0.5428 ± 0.0391	0.3552 ± 0.0000	0.3552 ± 0.0000	0.5175 ± 0.0668	0.4812 ± 0.0499
NATOPS	ACC	0.6185 ± 0.0094	0.4593 ± 0.0723	0.6056 ± 0.0000	0.4167 ± 0.1179	0.3833 ± 0.0552	0.3944 ± 0.0000	0.5278 ± 0.0000	0.0815 ± 0.0498	0.2167 ± 0.0045
	F1	0.6205 ± 0.0130	0.5502 ± 0.0698	0.5821 ± 0.0000	0.4102 ± 0.1252	0.4029 ± 0.0367	0.1579 ± 0.0518	0.5294 ± 0.0000	0.0943 ± 0.0607	0.1316 ± 0.0045
PEMS-SF	ACC	0.5780 ± 0.0164	0.4355 ± 0.0196	0.4697 ± 0.0085	0.6012 ± 0.0000	0.5453 ± 0.0607	0.2370 ± 0.0000	0.4682 ± 0.0000	0.1715 ± 0.0027	0.2004 ± 0.0027
	F1	0.6215 ± 0.0398	0.4310 ± 0.0260	0.5328 ± 0.0085	0.6419 ± 0.0000	0.6408 ± 0.0433	0.1118 ± 0.0186	0.4506 ± 0.0021	0.1313 ± 0.0111	0.1150 ± 0.0052
RacketSports	ACC	0.4715 ± 0.0135	0.4254 ± 0.0265	0.3766 ± 0.0028	0.3750 ± 0.0000	0.3772 ± 0.0499	0.3553 ± 0.0000	0.4276 ± 0.0000	0.2851 ± 0.0967	0.2895 ± 0.0000
	F1	0.4657 ± 0.0695	0.3949 ± 0.0301	0.3757 ± 0.0034	0.3489 ± 0.0000	0.3748 ± 0.0465	0.3078 ± 0.0467	0.4347 ± 0.0001	0.2911 ± 0.0972	0.1396 ± 0.0000
SelfRegulation SCP1	ACC	0.8510 ± 0.0126	0.8385 ± 0.0621	0.5742 ± 0.0044	0.7247 ± 0.0903	0.7270 ± 0.1401	0.6007 ± 0.0000	0.8840 ± 0.0000	0.4790 ± 0.0595	0.5040 ± 0.0016
	F1	0.8500 ± 0.0135	0.8337 ± 0.0687	0.5038 ± 0.0098	0.6979 ± 0.1211	0.7234 ± 0.1418	0.3154 ± 0.0000	0.8838 ± 0.0000	0.4671 ± 0.0611	0.3415 ± 0.0018
SelfRegulation SCP2	ACC	0.6000 ± 0.0253	0.5630 ± 0.0114	0.5972 ± 0.0028	0.5222 ± 0.0000	0.5315 ± 0.0159	0.5481 ± 0.0026	0.5796 ± 0.0026	0.4907 ± 0.0262	0.5000 ± 0.0000
	F1	0.5850 ± 0.0244	0.5446 ± 0.0045	0.5972 ± 0.0028	0.5184 ± 0.0000	0.4656 ± 0.0248	0.4546 ± 0.0483	0.5533 ± 0.0034	0.4837 ± 0.0266	0.4327 ± 0.0000
StandWalkJump	ACC	0.7556 ± 0.0831	0.5556 ± 0.0629	0.5333 ± 0.0000	0.4667 ± 0.0000	0.5111 ± 0.1133	0.4000 ± 0.0000	0.4667 ± 0.0000	0.1333 ± 0.0000	0.6667 ± 0.0000
	F1	0.7382 ± 0.1048	0.5729 ± 0.0049	0.5238 ± 0.0000	0.5422 ± 0.0616	0.4582 ± 0.1094	0.3119 ± 0.0000	0.4550 ± 0.0000	0.0784 ± 0.0000	0.5342 ± 0.0000

Table 2: Clustering performance of different methods on 15 MTS datasets evaluated by ACC (↑) and F1 (↑) metrics. The **Best** and **second-best** results are highlighted by cell background.

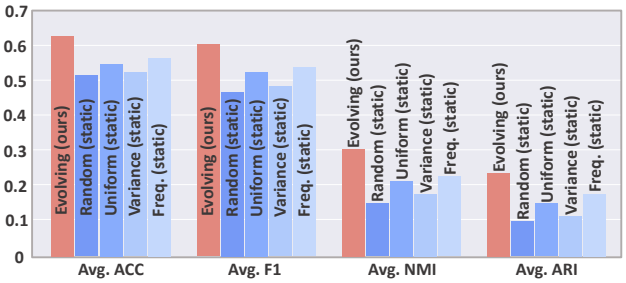

Figure 3: Clustering performance of EMTC+Evolving Masking vs. EMTC+Static Masking.

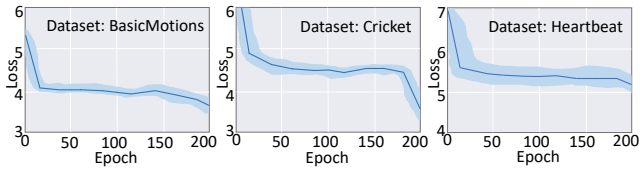

Figure 4: Loss function values during EMTC training.

Masking Strategies Comparison

Efficacy of the proposed evolving masking is validated by comparing it with four conventional static masking variants, including Random, Uniform, Variance-based, and Frequency-based masking, through replacement of the IVM module of EMTC. As shown in Figure 3, evolving masking consistently outperforms all static counterparts across all evaluation metrics, with average performance computed over all 15 benchmark datasets. Although Variance-based and Frequency-based methods incorporate data statistics, their static nature limits their adaptability to different clustering tasks. In contrast, our approach progressively refines its focus on cluster-salient timestamps through content-aware importance evaluation described bu Eq. (2), enabling precise redundancy suppression while preserving discriminative and clustering-friendly patterns.

Convergence Analysis

Figure 4 illustrates the convergence behavior across three representative datasets, showing smooth and stable optimization with the total loss consistently decreasing throughout training. The convergence characteristics vary appropriately with dataset properties: BasicMotions with balanced classes exhibits faster initial convergence, Cricket with more classes shows gradual stabilization, and Heartbeat with longer sequences maintains steady decline. This stable convergence validates the effectiveness of our joint optimization strategy with the corresponding total loss described by Eq. (13), where the synergy between CRL and CMC ensures progressive refinement of both representation quality and cluster structures.

Ablation Study of Key Components

The results in Table 3 demonstrates that both IVM and MEV are essential for EMTC's superior performance. The complete model consistently outperforms all ablated variants, with the absence of either component causing significant performance degradation. IVM generally contributes more substantially than MEV, particularly on long-sequence

Datasets	Components		Metrics	
	IVM	MEV	ACC	F1
BasicMotions	✓	✓	0.9083 ± 0.0656	0.9057 ± 0.0656
		✓	0.7000 ± 0.1080‡	0.7831 ± 0.0549‡
	✓		0.8750 ± 0.0354‡	0.8749 ± 0.0363‡
			0.3167 ± 0.0236‡	0.2958 ± 0.0464‡
Cricket	✓	✓	0.5972 ± 0.0227	0.6317 ± 0.0366
		✓	0.5139 ± 0.0113‡	0.5423 ± 0.0524‡
	✓		0.5556 ± 0.0300‡	0.5664 ± 0.0179‡
			0.5417 ± 0.1195‡	0.5895 ± 0.1296‡
DuckDuckGeese	✓	✓	0.4800 ± 0.0163	0.4917 ± 0.0357
		✓	0.4400 ± 0.0163‡	0.4259 ± 0.0241‡
	✓		0.4333 ± 0.0094‡	0.4449 ± 0.0238‡
			0.2667 ± 0.0094‡	0.2227 ± 0.0041‡
NATOPS	✓	✓	0.6185 ± 0.0094	0.6205 ± 0.0130
		✓	0.5093 ± 0.0664‡	0.4951 ± 0.0726‡
	✓		0.5296 ± 0.0502‡	0.5605 ± 0.0912‡
			0.4204 ± 0.1089‡	0.4145 ± 0.1189‡
PEMS-SF	✓	✓	0.5780 ± 0.0164	0.6215 ± 0.0398
		✓	0.4566 ± 0.0000‡	0.4149 ± 0.0012‡
	✓		0.4509 ± 0.0324‡	0.5077 ± 0.0133‡
			0.2543 ± 0.0544‡	0.2037 ± 0.0695‡
SelfRegulation SCP1	✓	✓	0.8510 ± 0.0126	0.8500 ± 0.0135
		✓	0.7782 ± 0.0121‡	0.7757 ± 0.0105‡
	✓		0.7713 ± 0.0255‡	0.7706 ± 0.0254‡
			0.7531 ± 0.1754‡	0.7006 ± 0.2497‡
SelfRegulation SCP2	✓	✓	0.6000 ± 0.0253	0.5850 ± 0.0244
		✓	0.5963 ± 0.0105‡	0.5953 ± 0.0095
	✓		0.5815 ± 0.0069‡	0.5658 ± 0.0143‡
			0.5241 ± 0.0233‡	0.4710 ± 0.0921‡
StandWalkJump	✓	✓	0.7556 ± 0.0831	0.7382 ± 0.1048
		✓	0.6000 ± 0.0544‡	0.5681 ± 0.0646‡
	✓		0.6444 ± 0.0629‡	0.6165 ± 0.0688‡
			0.4222 ± 0.0314‡	0.4321 ± 0.0175‡

Table 3: Ablation study of key components. The symbol "✓" denotes included components, and the symbol "‡" indicates performance degradation compared to the complete EMTC.

datasets like *StandWalkJump* where temporal redundancy is prominent. It is worth noting that the performance gap between the full model and the version without both MEV and IVM became more significant, confirming the synergistic effect of the two key components. Overall, the ablation validates the coupling design: IVM dynamically suppresses redundant timestamps while MEV comprehensively captures multi-view patterns to consolidate masking effectiveness and enhance cluster separation.

Efficiency Analysis

Efficiency and scalability of EMTC are evaluated by recording its execution time with respect to three key factors: number of dimensions (D), sequence length (T), and number of samples (N). As shown in Figure 5, EMTC demonstrates favorable scalability across different dimensions. When D and T scale, EMTC exhibits near-linear growth, maintaining low computational overhead.

Cluster Effect Visualization

To qualitatively assess representation quality from the perspective of clustering, Figure 6 presents t-SNE visualizations of learned representations on the Epilepsy dataset,

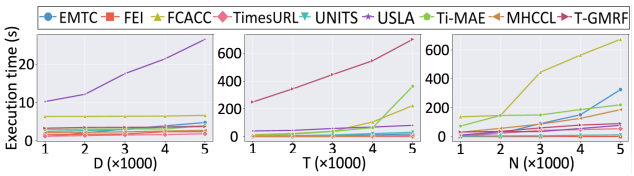

Figure 5: Execution time (s) of model training on synthetic datasets with varying D, T, and N.

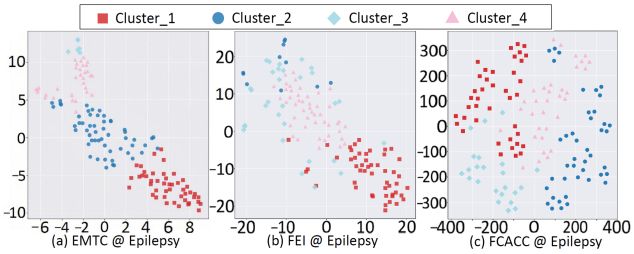

Figure 6: t-SNE visualization of the Epilepsy dataset represented by different methods.

comparing EMTC with FEI and FCACC. EMTC demonstrates superior cluster separation with relatively prominent boundaries and compact groupings, while FEI exhibits substantial inter-cluster overlap and FCACC shows moderate overlap but with its clusters scattered. These visual results confirm that the EMTC effectively captures clustering-friendly patterns from MTS, yielding semantically meaningful embeddings for MTS clustering.

Concluding Remarks

This paper proposes EMTC, an evolving-masking guided MTS clustering framework that explicitly addresses temporal redundancy through learnable masking and multi-view representation learning. It introduces IVM module to dynamically suppress redundant timestamps via attention-based scoring, ensures that the representation learning focuses on discriminative temporal regions. A dual-path MEV learning is also adopted, integrating the CRL and CMC learning to ensure the effectiveness of IVM and facilitate clustering-friendly representation learning, respectively. EMTC is comprehensively validated on 15 benchmark datasets with comparative, ablation, efficiency, and qualitative studies, demonstrating its superior performance over the SOTA methods. To our knowledge, EMTC is the first dynamic masking attempt under the scenario of unsupervised MTS analysis, breaking the clustering performance bottleneck caused by the MTS redundancy, and providing a mask learning paradigm to unsupervised MTS analysis. Despite the above-mentioned merits, this work is also not exempt from limitations. IVM relies on determining a to-be-masked threshold, while MEV generation assumes that all data is known. Extending EMTC to automatically determine the masking threshold and continuously adapt to nonstationary streaming data would be the next avenue.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 62476063 and the Natural Science Foundation of Guangdong Province under Grant 2025A1515011293.

References

Alwan, A.; Brimicombe, A. J.; Ciupala, M. A.; Ghorashi, S. A.; Baravalle, A.; and Falcarin, P. 2022. Time-Series Clustering for Sensor Fault Detection in Large-scale Cyber–Physical Systems. *Computer Networks*, 218(3): 109384.

Ashok, A.; Étienne Marcotte; Zantedeschi, V.; Chapados, N.; and Drouin, A. 2024. TACTiS-2: Better, Faster, Simpler Attentional Copulas for Multivariate Time Series. arXiv preprint arXiv:2310.01327.

Bagnall, A.; Dau, H. A.; Lines, J.; Flynn, M.; Large, J.; Bostrom, A.; Southam, P.; and Keogh, E. 2018. The UEA Multivariate Time Series Classification Archive, 2018. arXiv preprint arXiv:1811.00075.

Barragan, J. F.; Fontes, C. H.; and Embiruçu, M. 2016. A Wavelet-based Clustering of Multivariate Time Series Using A Multiscale SPCA Approach. *Computers & Industrial Engineering*, 95: 144–155.

Ben-Shaul, I.; Shwartz-Ziv, R.; Galanti, T.; Dekel, S.; and LeCun, Y. 2023. Reverse Engineering Self-Supervised Learning. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 58324–58345.

Boniol, P.; Tiano, D.; Bonifati, A.; and Palpanas, T. 2025. k-Graph: A Graph Embedding for Interpretable Time Series Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 37(5): 2680–2694.

Cai, Q.; Chen, L.; Shao, J.; and Chen, L. 2025. Data-Adaptive Dynamic Time Warping-based Multivariate Time Series Fuzzy Clustering. *IEEE Access*, 13: 75525–75534.

Chang, C.; Chan, C.-T.; Wang, W.-Y.; Peng, W.-C.; and Chen, T.-F. 2024. TimeDRL: Disentangled Representation Learning for Multivariate Time-Series. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*, 625–638.

Chowdhury, R. R.; Li, J.; Zhang, X.; Hong, D.; Gupta, R. K.; and Shang, J. 2023. PrimeNet: Pre-training for Irregular Multivariate Time Series. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 37, 7184–7192.

Ding, C.; Sun, S.; and Zhao, J. 2023. MST-GAT: A Multimodal Spatial–Temporal Graph Attention Network for Time Series Anomaly Detection. *Information Fusion*, 89: 527–536.

Ding, W.; Li, W.; Zhang, Z.; Wan, C.; Duan, J.; and Lu, S. 2023. Time-Varying Gaussian Markov Random Fields Learning for Multivariate Time Series Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(11): 11950–11966.

Eldele, E.; Ragab, M.; Chen, Z.; Wu, M.; and Li, X. 2024. TSLANet: Rethinking Transformers for Time Series Representation Learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, 494, 12409–12428.

Fu, E.; and Hu, Y. 2025. Frequency-Masked Embedding Inference: A Non-Contrastive Approach for Time Series Representation Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, 16639–16647.

Fu, Y.; and Xue, F. 2022. MAD: Self-Supervised Masked Anomaly Detection Task for Multivariate Time Series. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Gao, S.; Koker, T.; Queen, O.; Hartvigsen, T.; Tsiligkaridis, T.; and Zitnik, M. 2024. UniTS: A Unified Multi-Task Time Series Model. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, volume 37, 140589–140631.

Hamerly, G.; and Elkan, C. 2003. Learning The k in k-means. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, volume 16, 281–288.

Han, S.; and Woo, S. S. 2022. Learning Sparse Latent Graph Representations for Anomaly Detection in Multivariate Time Series. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD)*, 2977–2986.

Ienco, D.; and Interdonato, R. 2020. Deep Multivariate Time Series Embedding Clustering via Attentive-Gated Autoencoder. In *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, volume 12084, 318–329.

Ienco, D.; and Interdonato, R. 2023. Deep Semi-Supervised Clustering for Multi-Variate Time-Series. *Neurocomputing*, 516: 36–47.

Ikotun, A. M.; Ezugwu, A. E.; Abualigah, L.; Abuhaija, B.; and Heming, J. 2023. K-means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in The Era of Big Data. *Information Sciences*, 622: 178–210.

Lafabregue, B.; Weber, J.; Gancarski, P.; and Forestier, G. 2022. End-to-End Deep Representation Learning for Time Series Clustering: A Comparative Study. *Data Mining and Knowledge Discovery*, 36(1): 29–81.

Lee, S.; Kim, W.; and Son, Y. 2024. Spatio-Temporal Consistency for Multivariate Time-Series Representation Learning. *IEEE Access*, 12: 30962–30975.

Li, H. 2019. Multivariate Time Series Clustering Based on Common Principal Component Analysis. *Neurocomputing*, 349: 239–247.

Li, H.; and Liu, Z. 2021. Multivariate Time Series Clustering Based on Complex Network. *Pattern Recognition*, 115: 107919.

Li, H.; and Wei, M. 2020. Fuzzy Clustering Based on Feature Weights for Multivariate Time Series. *Knowledge-Based Systems*, 197: 105907.

Li, R.; He, C.; Li, S.; Zhang, Y.; and Zhang, L. 2023a. DynaMask: Dynamic Mask Selection for Instance Segmentation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, 11279–11288.

Li, Y.; Du, M.; Jiang, X.; and Zhang, N. 2025. Contrastive Learning-based Multi-View Clustering for Incomplete Multivariate Time Series. *Information Fusion*, 117: 102812.

Li, Z.; Rao, Z.; Pan, L.; Wang, P.; and Xu, Z. 2023b. Ti-MAE: Self-Supervised Masked Time Series Autoencoders. arXiv preprint arXiv:2301.08871.

Liu, J.; and Chen, S. 2024. TimesURL: Self-Supervised Contrastive Learning for Universal Time Series Representation Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, 13918–13926.

Ma, H.; Zhang, Z.; Li, W.; and Lu, S. 2021. Unsupervised Human Activity Representation Learning with Multi-task Deep Clustering. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1): 1–25.

MacQueen, J. 1967. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability (Berkeley Symp.)*, volume 1, 281–297.

Meng, Q.; Qian, H.; Liu, Y.; Cui, L.; Xu, Y.; and Shen, Z. 2023. MHCCL: Masked Hierarchical Cluster-Wise Contrastive Learning for Multivariate Time Series. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 9153–9161.

Peng, F.; Luo, J.; Lu, X.; Wang, S.; and Li, F. 2024a. Cross-Domain Contrastive Learning for Time Series Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, 8921–8929.

Peng, F.; Luo, J.; Lu, X.; Wang, S.; and Li, F. 2024b. Cross-Domain Contrastive Learning for Time Series Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, 8921–8929.

Shen, D. S.; and Chi, M. 2023. TC-DTW: Accelerating Multivariate Dynamic Time Warping Through Triangle Inequality and Point Clustering. *Information Sciences*, 621: 611–626.

Shi, J.; Wu, Y.; Peng, Y.; Wu, B.; Wang, L.; Liu, G.; and Luo, Y. 2025. Trainable Dynamic Mask Sparse Attention. arXiv preprint arXiv:2508.02124.

Sinaga, K. P.; and Yang, M.-S. 2020. Unsupervised K-Means Clustering Algorithm. *IEEE Access*, 8: 80716–80727.

Tan, Z.; Xie, T.; Sun, B.; Zhang, X.; Zhang, Y.; and Cheung, Y.-M. 2025. MEET-Sepsis: Multi-Endogenous-View Enhanced Time-Series Representation Learning for Early Sepsis Prediction. arXiv preprint arXiv:2510.15985.

Trosten, D. J.; Strauman, A. S.; Kampffmeyer, M.; and Jenssen, R. 2019. Recurrent Deep Divergence-based Clustering for Simultaneous Feature Learning and Clustering of Variable Length Time Series. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3257–3261.

Wang, C.; Du, M.; Jiang, X.; and Dong, Y. 2025. Fuzzy Cluster-Aware Contrastive Clustering for Time Series. arXiv preprint arXiv:2503.22211.

Wang, X.; Wirth, A.; and Wang, L. 2007. Structure-Based Statistical Features and Multivariate Time Series Clustering. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, 351–360.

Wu, Y.; Meng, X.; He, Y.; Zhang, J.; Zhang, H.; Dong, Y.; and Lu, D. 2024. Multi-view Self-Supervised Contrastive Learning for Multivariate Time Series. In *Proceedings of the ACM International Conference on Multimedia (MM)*, 9582–9590.

Xie, T.; Tan, Z.; Xiao, H.; Sun, B.; and Zhang, Y. 2025. DE3S: Dual-Enhanced Soft-Sparse-Shape Learning for Medical Early Time-Series Classification. arXiv preprint arXiv:2510.12214.

Yang, K.; and Shahabi, C. 2004. A PCA-based Similarity Measure for Multivariate Time Series. In *Proceedings of the ACM International Workshop on Multimedia Databases (MMDB)*, 65—74.

Yang, X.; Liu, Y.; Zhou, S.; Wang, S.; Tu, W.; Zheng, Q.; Liu, X.; Fang, L.; and Zhu, E. 2023. Cluster-Guided Contrastive Graph Clustering Network. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 37, 10834–10842.

Yue, Z.; Wang, Y.; Duan, J.; Yang, T.; Huang, C.; Tong, Y.; and Xu, B. 2022. TS2Vec: Towards Universal Representation of Time Series. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 36, 8980–8987.

Zhang, N.; and Sun, S. 2023. Multiview Unsupervised Shapelet Learning for Multivariate Time Series Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4981–4996.

Zhang, W.; Yang, L.; Geng, S.; and Hong, S. 2024. Self-Supervised Time Series Representation Learning via Cross Reconstruction Transformer. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11): 16129–16138.

Zhang, Y.; Feng, S.; Wang, P.; Tan, Z.; Luo, X.; Ji, Y.; Zou, R.; and ming Cheung, Y. 2025. Learning Self-Growth Maps for Fast and Accurate Imbalanced Streaming Data Clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36(9): 16049–16061.

Zhang, Z.; Zhang, Y.; Zeng, A.; Pan, D.; Ji, Y.; and Lin, J. 2023a. Time-Series Data Imputation via Realistic Masking-guided Tri-attention Bi-GRU. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, 3074–3082.

Zhang, Z.; Zhang, Y.; Zeng, A.; Pan, D.; and Zhang, X. 2023b. Learning Hierarchical Representations in Temporal and Frequency Domains for Time Series Forecasting. In *Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, 91–103.

Zhong, B.; Wang, P.; Pan, J.; and Wang, X. 2023. TS-MVP: Time-Series Representation Learning by Multi-view Prototypical Contrastive Learning. In *Proceedings of the International Conference on Advanced Data Mining and Applications (ADMA)*, 278–292.