TFEC: Multivariate Time-Series Clustering via Temporal-Frequency Enhanced Contrastive Learning

Anonymous Authors

***Abstract*—Multivariate Time-Series (MTS) clustering is crucial for signal processing and data analysis. Although deep learning approaches, particularly those leveraging Contrastive Learning (CL), are prominent for MTS representation, existing CL-based models face two key limitations: 1) neglecting clustering information during positive/negative sample pair construction, and 2) introducing unreasonable inductive biases, e.g., destroying time dependence and periodicity through augmentation strategies, compromising representation quality. This paper, therefore, proposes a Temporal-Frequency Enhanced Contrastive (TFEC) learning framework. To preserve temporal structure while generating low-distortion representations, a temporal-frequency Co-EnHancement (CoEH) mechanism is introduced. Accordingly, a synergistic dual-path representation and cluster distribution learning framework is designed to jointly optimize cluster structure and representation fidelity. Experiments on six real-world benchmark datasets demonstrate TFEC's superiority, achieving 4.48% average NMI gains over SOTA methods, with ablation studies validating the design. The code of the paper is available at: https://github.com/yueliangy/TFEC.**

***Index Terms*—Time-Series Data, Clustering, Frequency-Domain, Contrastive Learning, Data Augmentation.**

I. Introduction

Multivariate Time-Series (MTS) clustering [1]–[3] represents a critical analytical paradigm [4], [5] with diverse applications in healthcare monitoring [6], [7], industrial diagnostics [8], [9], and financial analytics [10], [11]. Conventional distance-based [12], [13] and statistical feature methods [14], [15] rely on handcrafted features derived from superficial temporal characteristics (e.g., seasonality/periodicity), yielding shallow representations. These inherently lack the capacity for modeling complex cross-variate dependencies and hierarchical structures while exhibiting pronounced sensitivity to noise regimes and temporal misalignments, fundamentally compromising clustering fidelity in high-dimensional spaces.

Deep representation learning [16], [17] overcomes these limitations by autonomously extracting discriminative embeddings from MTS structures. Contemporary methodologies can be categorized into three paradigms: reconstruction-based approaches [18], [19] employ autoencoders for latent representation learning; generative frameworks [20], [21] utilize adversarial/variational architectures for distribution modeling; while contrastive learning (CL) [22], [23] explicitly optimizes relative similarity through instance discrimination. This direct discrimination enables CL to learn superior discriminative and inherently cluster-friendly embeddings by eliminating the necessity to model complex data distributions or irrelevant input specifics.

Contrastive learning has established itself as a powerful paradigm for learning representations from MTS by optimizing embedding similarity through instance discrimination, offering inherent benefits for cluster structure formation [24]. Nevertheless, existing CL methods are constrained by two major drawbacks: many approaches either depend extensively on latent-space manipulations, such as those leveraging factorized orthogonal spaces [25], information-aware augmentations [26], or soft assignment strategies [27]; or they construct views based on hand-crafted mechanisms, which include predefined augmentations [28], [29] and pre-training tasks tailored for specific data properties like irregularity [30], [31]. A common shortcoming among these strategies is their frequent oversight of intrinsic cluster-discriminative information during contrastive pair construction. Moreover, such hand-crafted mechanisms, especially suboptimal augmentations, can introduce temporal distortions that compromise the fidelity of the representations.

To address these challenges, this paper proposes TFEC: a Temporal-Frequency Enhanced Contrastive learning framework. Our core innovation integrates: (1) A temporal-frequency Co-EnHancement (CoEH) mechanism generating low-distortion representations through adaptive spectral mixing while preserving temporal topology, and (2) A synergistic dual-path architecture where cluster-aware contrastive learning explicitly leverages inherent categorical structures for discriminative embedding formation, while reconstruction-based stabilization maintains representation integrity. This unified paradigm establishes an end-to-end self-supervised framework for optimized MTS clustering, effectively mitigating temporal distortions while fully exploiting underlying cluster structures to improve representation quality. The three main contributions of TFEC are summarized as follows:

- **Temporal-Frequency Co-Enhancement Mechanism**: This paper introduces a dual-domain co-enhancement strategy that preserves temporal coherence through aligned cropping and enriches representations via adaptive frequency mixing with semantically proximate neighbors, enabling the generation of a low-distortion Enhanced MTS Embedding (EME) with improved discriminability and fidelity.
- **Cluster-Aware Contrastive Learning with High-Confidence Sampling**: A pseudo-label guided paradigm is proposed to fuse dual-view representations for cluster assignment, followed by the selection of high-confidence intra-cluster samples to form reliable contrastive pairs.

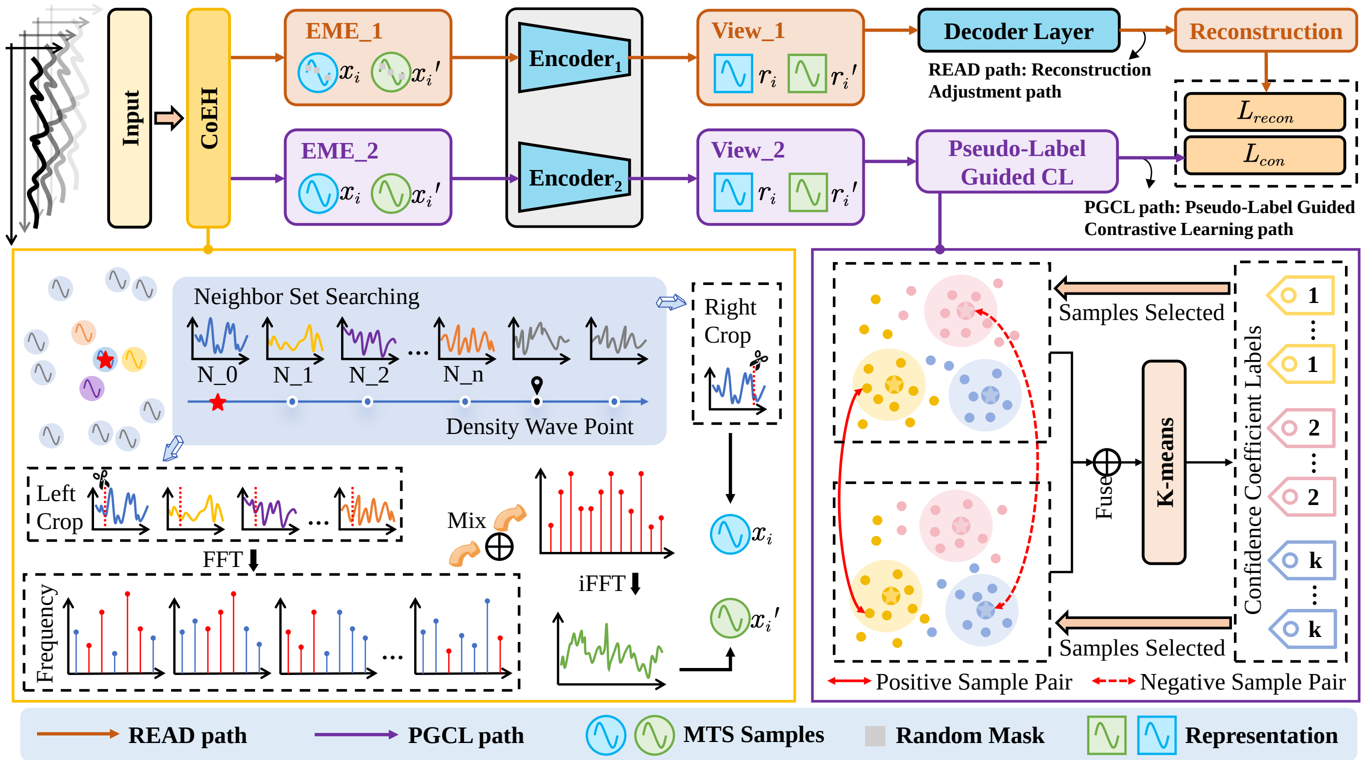


Fig. 1: **Overall framework of TFEC**. CoEH generates low-distortion EME. The dual-path architecture processes EME: the PGCL path performs pseudo-label guided contrastive learning on cluster structures based on high-confidence samples, while the READ path stabilizes representations via mask reconstruction.

This approach explicitly harnesses inherent cluster structures, substantially improving the quality and stability of learned representations.

- **Synergistic Dual-Path Learning Architecture**: This paper comes with a dual-path architecture comprising a Pseudo-label Guided Contrastive Learning (PGCL) path and a REconstruction ADjustment (READ) path. The PGCL path refines cluster compactness and separation, while the READ path enhances representation fidelity through masked EME reconstruction. Their synergistic interaction ensures discriminative and cluster-friendly representations generation.

II. Methodology

As shown in Fig. 1, the TFEC framework addresses the challenges of temporal distortion and underutilized clustering information through two core innovations: (1) A **Temporal-Frequency Co-Enhancement Mechanism** for generating low-distortion EME, and (2) A **Synergistic Dual-Path Learning Architecture** for learning discriminative and low-distortion representations.

A. Temporal-Frequency Co-Enhancement Mechanism

Given a MTS dataset $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ where each $\mathbf{x}_i \in \mathbb{R}^{T \times F}$, this paper first preserves temporal integrity through aligned cropping. This ensures that local temporal structures and phase coherence are maintained, avoiding the introduction of spurious inductive biases.

To further enhance the embeddings without introducing unreasonable inductive biases, this paper incorporates frequency-domain information from semantically proximate neighbors. These neighbors are identified via a density-aware selection process that ensures physically meaningful associations, avoiding arbitrary or noisy pairings. For each sample $\mathbf{x}_i$, this work applies the Fast Fourier Transform (FFT) to its temporally aligned segments:

$$\mathbf{q}_i^L = \mathcal{F}(x_i^L), \quad \mathbf{q}_{\langle i,p\rangle}^L = \mathcal{F}(x_{\langle i,p\rangle}^L), \tag{1}$$

where x_i^L and $x_{\langle i,p\rangle}^L$ denote cropped segments of the processed MTS and its neighbor, respectively. The frequency embeddings are then adaptively mixed using importance weights δ_p derived from feature-space distances, resulting in a blended frequency embedding:

$$\mathbf{F} = \mathbf{q}_i^L + \sum_p \delta_p \cdot \mathbf{q}_{\langle i,p\rangle}^L. \tag{2}$$

Finally, the inverse FFT is applied to synthesize the enhanced time-series sample, preserving both temporal structure and enriched frequency characteristics. This dual-domain strategy effectively augments the data while minimizing distortion, facilitating the learning of more robust representations.

Algorithm 1 TFEC: Temporal-Frequency Enhanced Contrastive Learning Framework

Require: Multivariate time-series dataset $\mathcal{X}$, number of clusters k, hyperparameters α, β.
Ensure: Cluster assignments $\mathbf{Y}$, representation $\mathbf{R}$.

```
Temporal-Frequency Co-Enhancement:
for each x_i ∈ X do
    Identify proximate neighbors N_i via density wave detection
    Apply aligned cropping to obtain segments x_i^L, x_(i,p)^L
    Compute FFT: q_i^L = F(x_i^L), q_(i,p)^L = F(x_(i,p)^L)
    Mix frequencies: F = q_i^L + Σ_p δ_p · q_⟨i,p⟩^L
    Synthesize enhanced sample via inverse FFT
end for
Form Enhanced MTS Embedding (EME)
Dual-Path Learning:
repeat
    PGCL Path:
    Extract encodings r, r' from dual views
    Fuse views: R = (r + r')/2
    Initialize clusters via K-means on R and update cluster assignments Y
    Compute confidence scores CONF_i (Eq. 3)
    Construct contrastive pairs P, N using high-confidence samples
    Update embeddings via L_con (Eq. 4)
    READ Path:
    Reconstruct masked EME via autoencoder
    Compute reconstruction loss L_recon
    Update model parameters by minimizing L_total = βL_con + (1 − β)L_recon
until convergence
```

B. Synergistic Dual-Path Learning Architecture

The dual-path architecture jointly improves representation stability and cluster distribution fidelity. The READ path processes EME which has been masked randomly through autoencoding to mitigate training drift. Concurrently, the PGCL path extracts discriminative encodings $\mathbf{r}, \mathbf{r}'$ from dual representation views. Cluster assignments emerge from fused representations $\mathbf{R} = (\mathbf{r} + \mathbf{r}')/2$ via K-means initialization. Crucially, high-confidence samples are selected using cluster-aware pseudo-labels:

$$\mathrm{CONF}_i = \exp\left(-\|\mathbf{R}_i - \mathbf{c}_p\|^2\right), \tag{3}$$

where $\mathbf{c}_p$ denotes the p-th cluster centroid. This confidence metric quantifies representation-cluster alignment reliability, enabling the construction of unambiguous contrastive pairs. Positive pairs ($\mathcal{P}$) comprise same-cluster high-confidence samples from different views, while negative pairs ($\mathcal{N}$) are formed between distinct central samples of high-confidence samples. This strategy explicitly leverages inherent cluster structures rather than relying on arbitrary latent space manipulations, ensuring contrastive signals align with genuine categorical boundaries. The READ path's reconstruction stability and PGCL's cluster distribution improvement form complementary objectives that mutually reinforce representation quality.

TABLE I: Statistics of datasets. k indicates the 'true' number of clusters provided by the labels of the datasets.

No.	Datasets	N	T	F	k
1	AtrialFibrillation (UEA 2018)	15	640	2	3
2	ERing (UEA 2018)	30	65	4	6
3	RacketSports (UEA 2018)	152	30	6	4
4	Libras (UEA 2018)	180	45	2	15
5	StandWalkJump (UEA 2018)	15	2500	4	3
6	NATOPS (UEA 2018)	180	51	24	6

C. Loss Function and Model Training

The hybrid loss function synergistically integrates contrastive learning and reconstruction objectives. Pseudo-label guided contrastive loss minimizes intra-cluster variance while maximizing inter-cluster separation:

$$\mathcal{L}_{\text{con}} = \underbrace{\frac{1}{|\mathcal{P}|}\sum_{(i,j)\in\mathcal{P}} \|\mathbf{r}_i - \mathbf{r}'_j\|^2}_{\text{Positive alignment}} + \alpha \underbrace{\frac{1}{|\mathcal{N}|}\sum_{(p,q)\in\mathcal{N}} \frac{\langle \mathbf{c}_p, \mathbf{c}_q\rangle}{\|\mathbf{c}_p\|\|\mathbf{c}_q\|}}_{\text{Negative separation}}, \tag{4}$$

where α balances attraction and repulsion forces. Simultaneously, the READ path computes the reconstruction loss $\mathcal{L}_{\text{recon}}$ for the reconstructed MTS data. The joint optimization objective integrates both components through a trade-off hyperparameter β:

$$\mathcal{L}_{\text{total}} = \beta\mathcal{L}_{\text{con}} + (1-\beta)\mathcal{L}_{\text{recon}}. \tag{5}$$

This formulation enables end-to-end training where reconstruction stabilizes feature learning while contrastive signals sharpen cluster boundaries. The dual-path learning ensures representations retain temporal fidelity while achieving separability, fulfilling critical requirements for effective MTS clustering. The Adam optimizer minimizes $\mathcal{L}_{\text{total}}$ through gradient updates that progressively refine both cluster distribution and representation efficacy. The full TFEC algorithm for MTS representation learning and clustering is summarized as Algorithm 1.

III. Experiment

Experimental Settings. This paper evaluates TFEC comprehensively through **six types of experiments**: *Clustering Performance Evaluation*, *Ablation Study*, *Significance Test*, *Comparison with Other Augmentation Mechanisms*, *Efficiency Evaluation* and *visualization*. Experiments are conducted on **six diverse UEA datasets** [32]: *AtrialFibrillation*, *ERing*, *RacketSport*, *Libras*, *StandWalkJump* and *NATOPS*. As shown in TABLE I, these datasets exhibit varied lengths ($T \in [30, 2500]$) and label types ($Labels \in [3, 15]$). This paper compares against **four SOTA deep clustering methods**: *TimesURL* [22], *UNITS* [33], *DropPatch* [17], and *FCACC* [24], with *K-means* [34] as baseline. Performance is

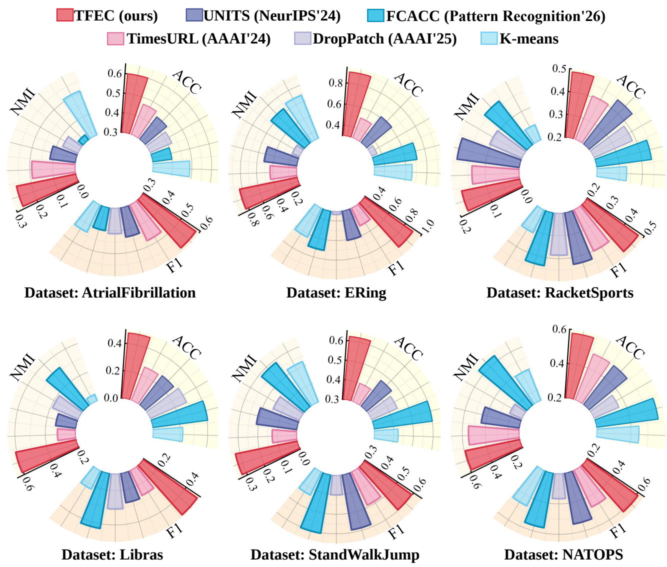


Fig. 2: Clustering performance comparison on six UEA datasets using ACC, F1 and NMI as evaluation metrics.

measured using **three metrics**: Clustering Accuracy (ACC), Normalized Mutual Information (NMI), and F1-Score (F1). All experiments are implemented in Python 3.10. Due to space constraints, the extended version of this paper is available as online Extended Version.

Clustering Performance Evaluation (Fig. 2). TFEC consistently achieves superior performance across all six datasets, with particularly notable gains on *AtrialFibrillation* and *ERing*, where it substantially outperforms the strongest baselines. The CoEH mechanism effectively preserves structural integrity while enriching discriminative features, contributing to these improvements. The framework also maintains robust performance on long-duration sequences such as *StandWalkJump* and complex multi-class datasets like *Libras*, demonstrating its adaptability to diverse temporal characteristics. Although the advantage is relatively narrower on datasets like *RacketSports* and *NATOPS*, TFEC still achieves the highest scores across all metrics, confirming its ability to produce low-distortion, cluster-friendly representations that significantly surpass existing SOTA methods.

Ablation Study (TABLE II). The study confirms the necessity of all three core modules (CoEH, PGCL, READ), as removing any component generally degrades performance. Exceptions include *RacketSports*, where ablating PGCL yields a marginal NMI gain, suggesting limited contrastive benefits in its discriminative space; and *NATOPS*, where excluding either PGCL or READ maintains stability. Notably, jointly ablating PGCL and READ causes the most severe performance drop, underscoring their complementary roles. The full TFEC achieves optimal performance via synergistic temporal-frequency enhancement and dual-path learning.

Significance Test (Fig. 3). We also conducted BD tests to the results of clustering performance in Fig. 2. The lengths of the CD, when comparing six approaches on six real-world benchmark datasets, are 3.035 and 2.884 with $\alpha = 0.05$ and 0.1, respectively. It can be seen that TFEC significantly outperforms most of its counterparts, including four state-of-the-art methods, which indicates the competitiveness of TFEC in multivariate time-series clustering.

TABLE II: Ablation study. Checkmarks "✓" denote included components. The red dagger "‡" indicates performance degradation relative to the complete TFEC model.

Datasets	Key Components			Evaluation Metrics		
	CoEH	PGCL	READ	ACC	F1	NMI
AtrialFibrillation	✓	✓	✓	0.6000	0.5841	0.2661
		✓	✓	0.5556‡	0.5322‡	0.2646‡
	✓		✓	0.4889‡	0.4462‡	0.1926‡
	✓	✓		0.5556‡	0.5136‡	0.2800
	✓			0.4889‡	0.4336‡	0.2086‡
ERing	✓	✓	✓	0.9062	0.9068	0.8339
		✓	✓	0.8938‡	0.8950‡	0.8209‡
	✓		✓	0.8519‡	0.8852‡	0.8272‡
	✓	✓		0.8926‡	0.8932‡	0.8226‡
	✓			0.6593‡	0.6320‡	0.6795‡
RacketSports	✓	✓	✓	0.4868	0.4758	0.1737
		✓	✓	0.4583‡	0.4416‡	0.2082
	✓		✓	0.4561‡	0.4430‡	0.1808
	✓	✓		0.4189‡	0.4330‡	0.2105
	✓			0.3311‡	0.3115‡	0.0483‡
Libras	✓	✓	✓	0.4778	0.4886	0.6074
		✓	✓	0.4537‡	0.4759‡	0.5895‡
	✓		✓	0.4796	0.4838‡	0.5884‡
	✓	✓		0.4537‡	0.4759‡	0.5895‡
	✓			0.2185‡	0.2220‡	0.2465‡
StandWalkJump	✓	✓	✓	0.6222	0.5961	0.3231
		✓	✓	0.5778‡	0.5566‡	0.3034‡
	✓		✓	0.5556‡	0.5291‡	0.2579‡
	✓	✓		0.5778‡	0.5562‡	0.3027‡
	✓			0.4222‡	0.4321‡	0.2229‡
NATOPS	✓	✓	✓	0.5778	0.6114	0.5898
		✓	✓	0.5778	0.6114	0.5898
	✓		✓	0.5778	0.6114	0.5898
	✓	✓		0.4763‡	0.4678‡	0.4433‡
	✓			0.4463‡	0.4612‡	0.3925‡

Comparison with Other Augmentation Mechanism (Fig. 4). This paper compares the proposed temporal-frequency enhancement against five common augmentation strategies: jitter, scale, warp, permute, and frequency mask (freq_mask). Results across all six datasets demonstrate the consistent superiority of TFEC. It achieves the highest or competitive performance across all metrics, with notable advantages on challenging datasets such as *AtrialFibrillation* and *NATOPS*. While certain augmentations (e.g., warp on *RacketSports*) occasionally perform well, they exhibit instability and performance degradation on more complex data. These findings confirm that TFEC's enhancement mechanism effectively preserves temporal structure and enriches discriminative features, outperforming heuristic augmentations that often introduce unrealistic distortions or break inherent periodicities.

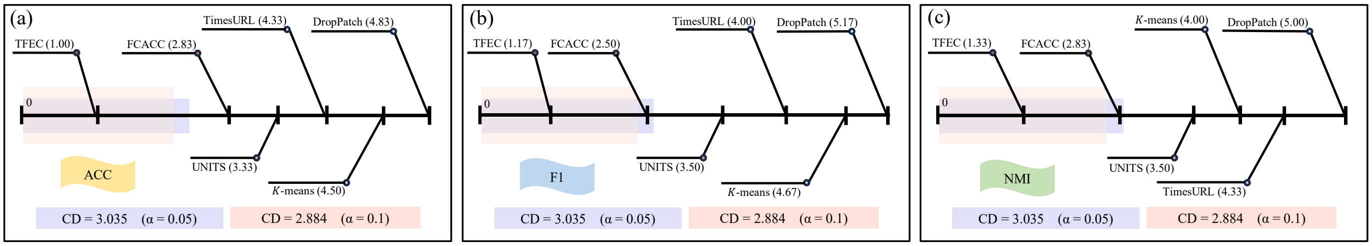
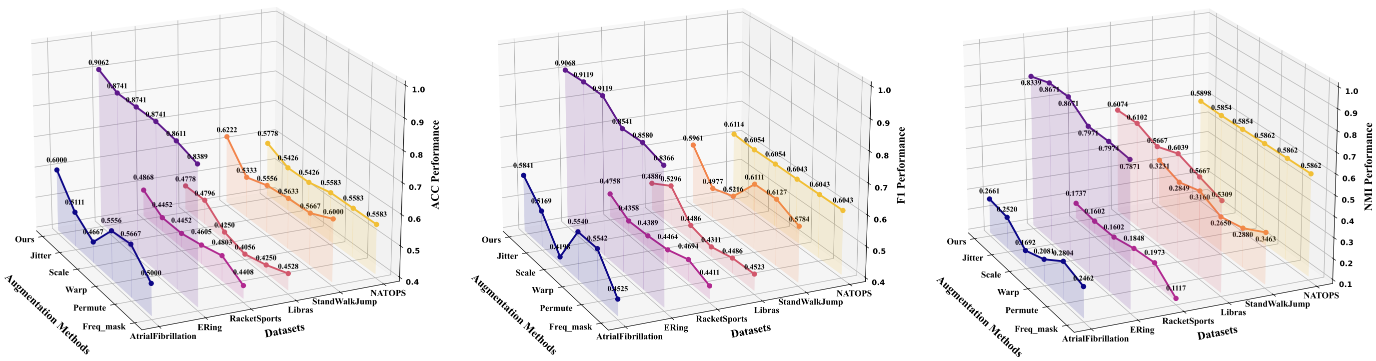


Fig. 3: BD test (a), (b) and (c) based on the average ranks of ACC, F1 and NMI in the Fig. 2. Methods ranked outside the CD intervals are believed to perform significantly differently from TFEC.

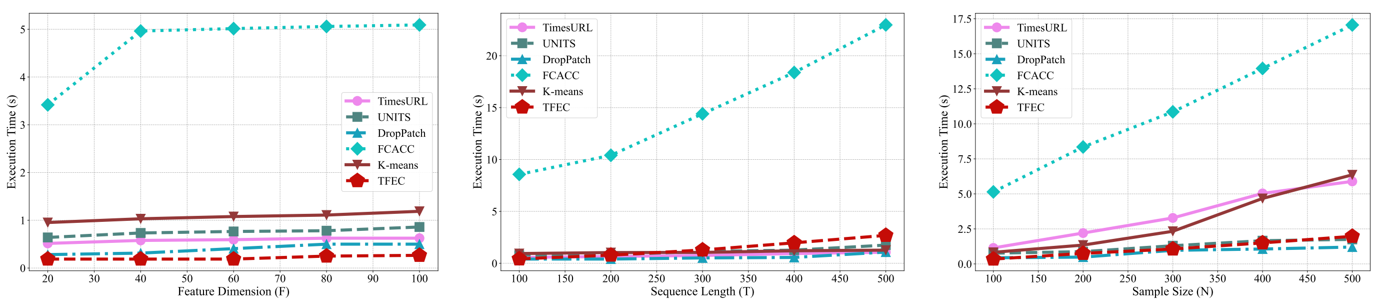


(a) ACC Performance

(b) F1 Performance

(c) NMI Performance

Fig. 4: Comparison of the proposed temporal-frequency enhancement against five common augmentation strategies: jitter, scale, warp, permute, and frequency mask (freq_mask).



(a) Execution time w.r.t. F

(b) Execution time w.r.t. T

(c) Execution time w.r.t. N

Fig. 5: Comparison of execution time with respect to different dimensions

Efficiency Evaluation (Fig. 5). The computational efficiency of TFEC against five baselines has been evaluated by us under three scaling scenarios. When fixing $N = 50, T = 2$ and increasing feature dimension F from 20 to 100, TFEC maintains low runtime (0.1875s to 0.2656s), significantly outperforming FCACC. With fixed $N = 50, F = 2$ and increasing sequence length T from 100 to 500, TFEC scales linearly (0.4219s to 2.6875s), comparable to TimesURL and UNITS. Under fixed $T = 2, F = 2$ with sample size N increasing from 100 to 500, TFEC demonstrates sub-linear growth (0.3438s to 1.9688s), substantially more efficient than FCACC and TimesURL. Overall, TFEC achieves competitive scalability across all dimensions, offering practical efficiency for multivariate time-series clustering.

IV. Concluding Remarks

This paper presents TFEC, a novel temporal-frequency enhanced contrastive learning framework for multivariate time-series clustering that mitigates augmentation-induced distortions and leverages latent cluster structures. The co-enhancement mechanism produces representations via adaptive frequency mixing while preserving temporal structure. A synergistic dual-path architecture improves cluster distribution and representation robustness through pseudo-label guided contrastive learning path and reconstruction adjustment path. Extensive evaluations show SOTA performance on six UEA datasets, with an average gain of 4.48% on NMI metric over counterparts. As this work is precision-oriented, future directions include enhancing noise robustness and computational efficiency via lightweight enhancement.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62476063, the NSFC/Research Grants Council (RGC) Joint Research Scheme under Grant N_HKBU214/21, the Natural Science Foundation of Guangdong Province under Grant 2025A1515011293, the General Research Fund of RGC under Grants 12202622 and 12201323, the RGC Senior Research Fellow Scheme under Grant SRFS23242S02, the General Projects of Shenzhen Science and Technology Program under Grant JCYJ20240813115124032, and the Shenzhen Maternity and Child Healthcare Hospital under Grant FYA2022018.

References

[1] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering – a decade review," *Information Systems*, vol. 53, pp. 16–38, 2015.

[2] Q. Meng, H. Qian, Y. Liu, L. Cui, Y. Xu, and Z. Shen, "Mhccl: Masked hierarchical cluster-wise contrastive learning for multivariate time series," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2023, pp. 9153–9161.

[3] Z. Tan, X. Luo, Y. Liu, and Y. Zhang, "Mask the redundancy: Evolving masking representation learning for multivariate time-series clustering," arXiv preprint arXiv:2511.17008, 2025.

[4] L. Xie, C. Li, X. Zhang, S. Zhai, Y. Fang, and Q. Shen, "Trls: A time series representation learning framework via spectrogram for medical signal processing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 1591–1595.

[5] J. Chen, M. Feng, and T. S. Wirjanto, "Harnessing contrastive learning and neural transformation for time series anomaly detection," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[6] Z. Tan, T. Xie, B. Sun, X. Zhang, Y. Zhang, and Y.-M. Cheung, "Meetsepsis: Multi-endogenous-view enhanced time-series representation learning for early sepsis prediction," arXiv preprint arXiv:2510.15985, 2025.

[7] T. Xie, Z. Tan, H. Xiao, B. Sun, and Y. Zhang, "De3s: Dual-enhanced soft-sparse-shape learning for medical early time-series classification," arXiv preprint arXiv:2510.12214, 2025.

[8] C. Xu, H. Huang, and S. Yoo, "A deep neural network for multivariate time series clustering with result interpretation," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2021, pp. 1–8.

[9] W. H. Suh, S. Oh, and C. W. Ahn, "Metaheuristic-based time series clustering for anomaly detection in manufacturing industry," *Applied Intelligence*, vol. 53, pp. 21 723–21 742, 2023.

[10] M. C. A. Neto, G. Tavares, V. M. Alves, G. D. Cavalcanti, and T. I. Ren, "Improving financial time series prediction using exogenous series and neural networks committees," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2010, pp. 1–8.

[11] P. D'Urso, L. D. Giovanni, and R. Massari, "Trimmed fuzzy clustering of financial time series based on dynamic time warping," *Annals of Operations Research*, vol. 299, pp. 1379–1395, 2021.

[12] R. Ma and R. Angryk, "Distance and density clustering for time series data," in *Proceedings of the IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017, pp. 25–32.

[13] H. He and Y. Tan, "Unsupervised classification of multivariate time series using vpca and fuzzy clustering with spatial weighted matrix distance," *IEEE Transactions on Cybernetics*, vol. 50, no. 3, pp. 1096–1105, 2020.

[14] X. Wang, K. Smith, and R. Hyndman, "Characteristic-based clustering for time series data," *Data Mining and Knowledge Discovery*, vol. 13, no. 3, pp. 335–364, 2006.

[15] X. Wang, A. Wirth, and L. Wang, "Structure-based statistical features and multivariate time series clustering," in *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, 2007, pp. 351–360.

[16] N. Zhang and S. Sun, "Multiview unsupervised shapelet learning for multivariate time series clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4981–4996, 2023.

[17] T. Qiu, Y. Xie, H. Niu, Y. Xiong, and X. Gao, "Enhancing masked time-series modeling via dropping patches," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2025, pp. 20 077–20 085.

[18] H. Li, "Multivariate time series clustering based on common principal component analysis," *Neurocomputing*, vol. 349, pp. 239–247, 2019.

[19] G. He, H. Wang, S. Liu, and B. Zhang, "Csmvc: A multiview method for multivariate time-series clustering," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13 425–13 437, 2022.

[20] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks," in *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, vol. 32, no. 494, 2019, pp. 5508–5518.

[21] E. Brophy, Z. Wang, Q. She, and T. Ward, "Generative adversarial networks in time series: A systematic literature review," *ACM Computing Surveys*, vol. 55, no. 10, pp. 1–31, 2023.

[22] J. Liu and S. Chen, "Timesurl: Self-supervised contrastive learning for universal time series representation learning," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2024, pp. 13 918–13 926.

[23] C. Chang, C.-T. Chan, W.-Y. Wang, W.-C. Peng, and T.-F. Chen, "Timedrl: Disentangled representation learning for multivariate time-series," in *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*, 2024, pp. 625–638.

[24] C. Wang, M. Du, X. Jiang, and Y. Dong, "Fuzzy cluster-aware contrastive clustering for time series," *Pattern Recognition*, vol. 173, p. 112899, 2026.

[25] S. Liu, T. Kimura, D. Liu, R. Wang, J. Li, S. Diggavi, M. Srivastava, and T. Abdelzaher, "Focal: contrastive learning for multimodal time-series sensing signals in factorized orthogonal latent space," in *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, 2023, pp. 47 309–47 338.

[26] D. Luo, W. Cheng, Y. Wang, D. Xu, J. Ni, W. Yu, X. Zhang, Y. Liu, Y. Chen, H. Chen, and X. Zhang, "Time series contrastive learning with information-aware augmentations," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2023, pp. 4534–4542.

[27] S. Lee, T. Park, and K. Lee, "Soft contrastive learning for time series," in *Proceedings of the International Conference on Learning Representations (ICLR)*, vol. 2024, 2024, pp. 46 815–46 839. [Online]. Available: https://openreview.net/forum?id=pAsQSWlDUf

[28] E. Eldele, M. Ragab, Z. Chen, M. Wu, C. Keong, Kwoh, X. Li, and C. Guan, "Time-series representation learning via temporal and contextual contrasting," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2021, pp. 2352–2359.

[29] B. U. Demirel and C. Holz, "Finding order in chaos: a novel data augmentation method for time series in contrastive learning," in *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, 2023, pp. 30 750–30 783.

[30] Y. Jiao, K. Yang, S. Dou, P. Luo, S. Liu, and D. Song, "Timeautoml: Autonomous representation learning for multivariate irregularly sampled time series," arXiv preprint arXiv:2010.01596, 2020.

[31] R. R. Chowdhury, J. Li, X. Zhang, D. Hong, R. K. Gupta, and J. Shang, "Primenet: Pre-training for irregular multivariate time series," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2023, pp. 7184–7192.

[32] A. Bagnall, H. A. Dau, J. Lines, M. Flynn, J. Large, A. Bostrom, P. Southam, and E. Keogh, "The uea multivariate time series classification archive," arXiv preprint arXiv:1811.00075, 2018.

[33] S. Gao, T. Koker, O. Queen, T. Hartvigsen, T. Tsiligkaridis, and M. Zitnik, "Units: a unified multi-task time series model," in *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, vol. 37, 2024, pp. 140 589–140 631.

[34] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability (Berkeley Symp.)*, vol. 1, 1967, pp. 281–297.