Journal of KIIT. Vol. 23, No. 6, pp. 55-63, Jun. 30, 2025. pISSN 1598-8619, eISSN 2093-7571 
http://dx.doi.org/10.14801/jkiit.2025.23.6.55

# Self-Supervised Contrastive Representation Learning for Time Series

Kunpeng Li*, Jong-Chul Lee**

This research was financially supported by Kwangwoon University

## Abstract

Obtaining labeled samples often requires substantial domain expertise and involves complex, time-consuming procedures, limiting the scalability of supervised learning methods. To address this, we propose a representation learning framework based on time series information contrastive learning (TICL) for learning from unlabeled data. Specifically, two distinct augmentation strategies generate semantically related views for each sample, and a latent information constraint is introduced to eliminate redundancy between sample pairs, promoting the learning of discriminative representations. Experiments on time series datasets from three domains demonstrate that TICL consistently achieves superior performance, particularly under few-shot conditions and in transfer learning scenarios, outperforming standard baselines by over 2% in accuracy on three benchmark datasets.

## 요 약

라벨이 지정된 샘플을 확보하려면 상당한 도메인 전문 지식과 복잡하고 시간 소모적인 절차가 필요하므로, 지도 학습 방식의 확장성이 제한됩니다. 이를 해결하기 위해, 우리는 레이블이 없는 시계열 데이터로부터 학습하기 위한 시계열 정보 대비 학습(TICL, Time Series Information Contrastive Learning) 기반 표현 학습 프레임워크를 제안합니다. 구체적으로, 두 가지 서로 다른 데이터 증강 전략을 통해 각 샘플에 대해 의미적으로 관련된 두 개의 뷰를 생성하고, 샘플 쌍 간의 중복 정보를 제거하기 위해 잠재 정보 제약 기법을 도입하여 판별력 있는 표현 학습을 촉진합니다. 세 개의 서로 다른 도메인에서의 시계열 데이터셋 실험 결과, TICL은 특히 few-shot 조건과 전이 학습 시나리오에서 일관되게 우수한 성능을 보였으며, 세 개의 벤치마크 데이터셋에서 표준 기준 모델 대비 2% 이상 높은 정확도를 기록하였습니다.

Keywords
time series, contrastive learning, data augmentation, transfer learning

* PhD candidate, Dept. of Electronic Convergence Eng. Kwangwoon University
- ORCID: https://orcid.org/0009-0004-9782-3253
** Professor, Dept. of Electronic Convergence Eng., Kwangwoon University
- ORCID: https://orcid.org/0000-0002-7854-6180

· Received: May 02, 2025, Revised: May 21, 2025, Accepted: May 24, 2025
· Corresponding Author: Jong-Chul Lee
Dept. of Electronic Convergence Eng., Kwangwoon University
20 Kwangwoon-ro, Nowon-Gu, Seoul 01897, South Korea
Tel.: +82-2-940-5203, Email: jclee@kw.ac.kr

## Ⅰ. Introduction

Time series classification (TSC) is a critical task in various domains such as healthcare, finance, industrial monitoring, and human activity recognition[1]. Traditionally, supervised learning approaches have been widely used in TSC, requiring substantial volumes of accurately labeled data to achieve high performance. However, acquiring extensive labeled datasets is often expensive, time-consuming, and sometimes impractical due to privacy constraints or the complexity of labeling processes[2].

In In recent years, self-supervised learning has emerged as a promising strategy to mitigate these challenges by leveraging unlabeled data to learn meaningful representations[3][4]. Among self-supervised methods, contrastive learning has gained notable attention due to its ability to extract robust, discriminative features without reliance on explicit annotations[5]. Contrastive learning operates by pulling representations of similar or augmented instances closer together while simultaneously pushing representations of dissimilar instances apart in the embedding space. Despite the extensive research on contrastive learning in computer vision and natural language processing, its application to time series data remains relatively under-explored[6]-[8]. Time series data pose unique challenges, including temporal dependencies, varying sequence lengths, and noise interference, which complicate the straightforward adaptation of methods developed for other data modalities. To address these issues, recent studies have focused on designing specialized contrastive frameworks tailored for time series classification tasks [9]-[11]. These approaches typically incorporate domain-specific data augmentation techniques, novel loss functions, and effective encoding schemes to enhance representation learning. By doing so, they have demonstrated the potential of self-supervised contrastive methods to match or even surpass supervised benchmarks, particularly in scenarios with limited labeled data. For example, Dai et al. and Kim et al. proposed contrastive learning frameworks for EEG and biosignal classification, respectively—leveraging brain-area-specific features or class-subject dual labels to enhance generalizability and achieve state-of-the-art performance[12][13]. However, these methods are designed for specific application scenarios and lack generalizability to other forms of time series data.

To address the aforementioned limitations, we propose a time-series information contrastive learning (TICL) framework. This framework employs two distinct data augmentation strategies to enhance the model's representational capability by generating two different yet semantically related views from the original time series, thereby facilitating the learning of robust and discriminative representations. Subsequently, we introduce an auxiliary latent information constraint (LIC) mechanism to enhance the diversity of latent representations and eliminate redundant information among samples. Specifically, LIC encourages the uniqueness of each sample representation, ultimately contributing to improved representation clarity and cross-domain adaptability. Extensive experiments conducted on benchmark datasets validate the proposed method's performance, highlighting its capability to achieve high accuracy with significantly reduced labeling requirements. Our approach thus provides a viable pathway towards efficient and scalable time series classification in diverse real-world applications.

## Ⅱ. Related Wokrs

### 2.1 Self-supervised contrastive learning

The self-supervised contrastive learning (SSCL) was initially introduced and widely adopted in the field of computer vision as a promising alternative to fully supervised learning, especially in scenarios where labeled data is limited or costly to acquire. Pioneering methods such as SimCLR[14], MoCo[15], and

BYOL[16] introduced contrastive and non-contrastive paradigms that enable models to learn powerful visual representations from unlabeled data. These approaches rely on instance discrimination tasks, where the objective is to maximize agreement between augmented views of the same image while minimizing similarity to other images in the batch. SimCLR leverages a simple contrastive loss and large batch sizes to learn semantically meaningful embeddings using strong data augmentations. MoCo improves training efficiency by maintaining a momentum-updated encoder and a dynamic queue of negative samples. In contrast, BYOL eliminates the need for negative pairs by employing a dual-network architecture with momentum updates, achieving state-of-the-art performance on ImageNet without contrastive loss. Although all these methods have been successfully applied to visual representation learning, they may lack the capability to effectively model the temporal dependencies and dynamic characteristics inherent in time series data.

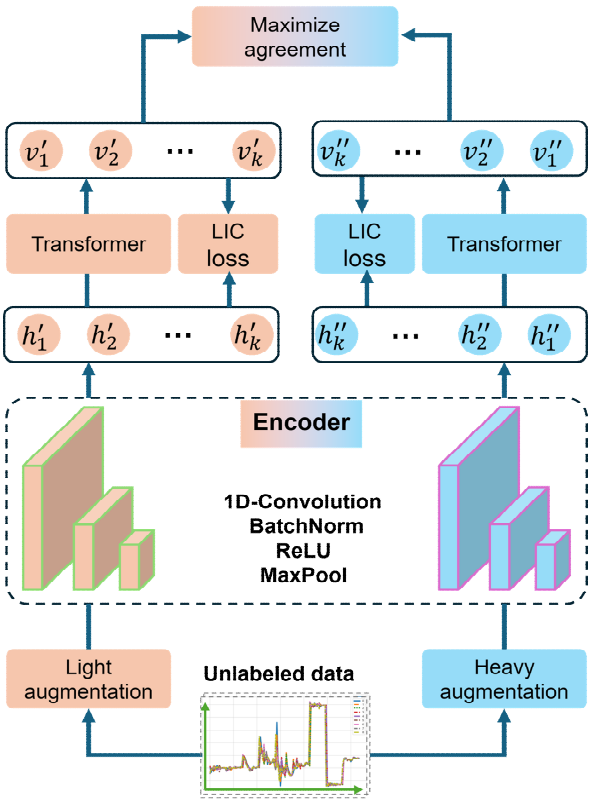


Fig. 1. Detailed architecture of TICL model

## 2.2 Self-supervised contrastive learning for time series classification

Representation learning has emerged as a powerful approach for time series representation learning, addressing the challenge of limited labeled data. Early adaptations of contrastive frameworks to time series data demonstrated the potential of SSCL in capturing meaningful representations. For instance, TimeCLR introduced a framework for univariate time series, employing data augmentations tailored to temporal structures to learn robust representations[17]. Building upon these foundations, Zhang et al. proposed the Time-Frequency Consistency (TF-C) model, which leverages the inherent duality between time and frequency domains in time series data[18]. By aligning representations from both domains through contrastive objectives, TF-C achieved significant improvements in transfer learning tasks across diverse datasets, including EEG and EMG signals. Further advancements include the TimesURL framework by Liu and Chen, which introduces frequency-temporal-based augmentations and constructs "double universes" as hard negatives to enhance the learning of universal time series representations[19]. Despite these advancements, significant challenges remain in designing effective data augmentation strategies that preserve temporal dependencies and in constructing semantically meaningful sample pairs, particularly due to the unique characteristics of time series data. Addressing these challenges is essential for the continued advancement of SSCL methods tailored to time series applications.

# III. Methods

We proposed a TICL framework, with its detailed architecture illustrated in Figure 1. Initially, two distinct data augmentation strategies are applied to the unlabeled time-series signals to generate a pair of different yet correlated views.

To further enhance representation diversity and suppress redundant information across samples, a LIC module is introduced. Finally, the model is optimized by maximizing the similarity between positive pairs (augmented from the same instance) while minimizing the similarity between negative pairs, thereby enabling the encoder to learn highly discriminative and generalizable representations.

## 3.1 Data augmentation

In the context of SSCL for temporal signals, the formulation of effective data augmentation techniques is pivotal to improving the model's robustness and generalization ability. Conventional approaches typically rely on subtle modifications to the input signals to create multiple views, ensuring that essential temporal patterns and structural properties are maintained. To further enhance adaptability across diverse feature domains, we introduce a hierarchical augmentation strategy that targets both local and global signal attributes. Specifically, two distinct augmentation schemes—referred to as light and heavy—are designed to generate complementary signal views. Light augmentation applies mild distortions such as signal scaling and jittering, effectively preserving local structural features and creating minimally altered signal variants. In contrast, heavy augmentation incorporates more aggressive transformations, including jitter and permutation, to significantly disrupt the global signal structure, thereby exposing the model to a broader range of feature variations. This approach simulates signal variations in complex scenarios. For each input sample $x_i = [d_1, d_2, \cdots, \mathrm{d_n}]$, its lightly augmented view is denoted as $x_i^{'}$, as defined in Formula (1):

$$x_i^{'} = x_i \bullet f_i, \quad f_i \sim N(\mu, \sigma^2) \tag{1}$$

The heavily augmented view is represented $x_i^{''}$, as defined in Formula (2):

$$x_i^{''} = x_{\pi(i,j)} \bullet s_i, \quad s_i \sim N(\mu, \sigma^2) \tag{2}$$

where $\pi(i,j)$ is a randomly shuffled index sequence that rearranges the time steps. Here, $N$ denotes a normal distribution.

## 3.2 Details of TICL

For any input sample $x$, the model initially encodes the raw signal into a high-dimensional representation $h$, capturing preliminary feature information. This transformation is achieved via an embedding module composed of three one-dimensional convolutional blocks, which project the low-dimensional input into a transferable latent motion space, effectively extracting local patterns while retaining temporal characteristics. The resulting embedding $h$ is subsequently fed into a Transformer-based module (Figure. 2), which produces semantically enriched view-level features $v$. These features are employed for similarity-based contrastive learning across augmented views. While maximizing the similarity between positive pairs is critical during similarity loss computation, we observe that excessive alignment of latent features may result in redundancy, thereby degrading the model's capacity to discriminate between different samples. To mitigate this, we introduce a LIC. Specifically, we incorporate a regularization term derived from information divergence into the original similarity loss to penalize over-similar latent encodings across samples. Let $z_i$ and $z_i^{'}$ denote the projected latent vectors corresponding to the embedded representation $h_i$ and its associated view $v_i$, where $B$ represents the batch size and $\theta$ is a tunable hyperparameter.

$$LIC = \frac{1}{B \quad (B-1)} \times \log\left[ \sum_{i=1, j=1 \cdots B, j \neq i}^{B} \sum e^{\theta |z_i - z_j|^2} \right] \tag{3}$$

The light and heavy augmentations correspond to $LIC_{light}$ and $LIC_{heavy}$, respectively. The self-supervised similarity comparison loss function is $LIC_{self}$, as defined in Formula (4):

$$L_{self} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(s(v_i, v_i^+)/\tau)}{\sum_{j=1}^{2N}\exp(s(v_i, v_j)/\tau)} \quad (4)$$

Where $s(\alpha,\beta) = (\alpha^T\beta)/\|\alpha\|\|\beta\|$ denotes the cosine similarity between $\alpha$ and $\beta$. $\tau$ symbolizes a temperature parameter. The TICL loss function can be formulated as shown in Formula (5):

$$L_{TICL} = L_{self} + \lambda(LIC_{light} + LIC_{heavy}) \quad (5)$$

Where $\lambda$ is is a hyperparameter used to balance the weight between contrastive learning loss and LIC.

## IV. Experiments

We conduct experiments on three publiclyavailable time series datasets (Table 1). All datasets were randomly divided into three subsets: 80% for training, 10% for validation, and 10% for testing. To better reflect real-world scenarios, the training set was first used for self-supervised pretraining, after which 20% of the training data was randomly selected to fine-tune the pre-trained model. (a) The epilepsy seizure detection dataset contains EEG recordings from 500 subjects, with each recording capturing 23.6 seconds of brain activity per subject. (b) Human Activity Recognition (HAR) is a widely used benchmark dataset for time series classification. It consists of 10,299 multivariate time series samples, each with 9 sensor channels and a fixed length of 128 time steps. The data were collected from 30 different individuals using accelerometers and gyroscopes embedded in a smartphone, sampled at 50 Hz. (c) The Sleep-EDF dataset was used for sleep stage classification. We utilized daily life recordings from four subjects, which were collected using a modified cassette recorder.

Table 1. Description of dataset used in our experiments

| Dataset | Train set | Test set | Shape | Class |
|---|---|---|---|---|
| Epilepsy | 9200 | 2300 | [1,178] | 2 |
| HAR | 8239 | 2059 | [9,128] | 6 |
| Sleep-EDF | 18108 | 4527 | [3,3000] | 8 |

## V. Results

### 5.1 Comparison with baseline approaches

To evaluate the effectiveness of the proposed TICL framework, we used two metrics: accuracy and macro-averaged F1 score (MF1) to assess model performance.

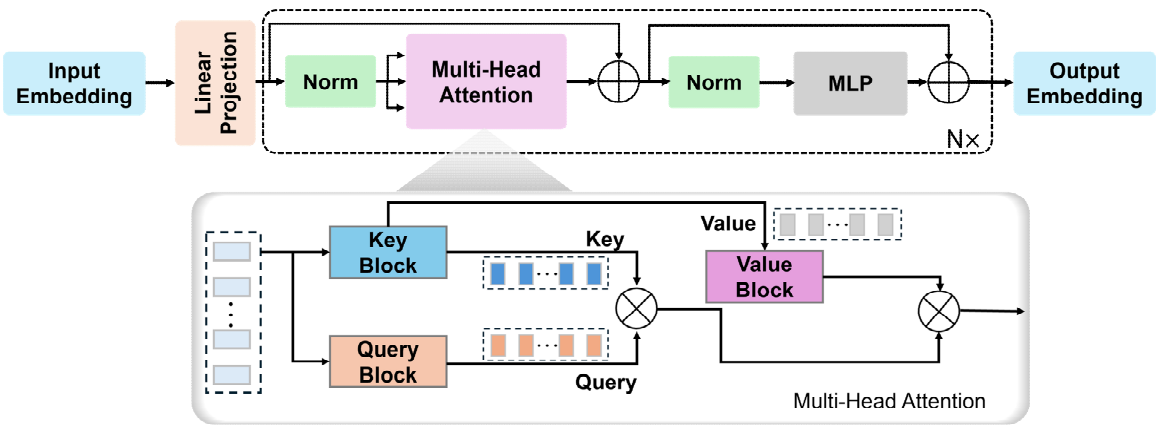


Fig. 2. Detailed architecture of transformer

We also compared TICL with several baseline methods: (1) SSL-ECG[10]; (2) CPC[20]; (3) SimCLR[14]; and (4) TS-TCC[21]. After fine-tuning the model with 20% of the labeled data, the results are summarized in Table 2. Overall, our proposed TICL framework outperformed all four state-of-the-art methods, demonstrating its strong representational learning capability and more stable performance on downstream tasks. is worth noting that contrastive methods such as CPC, SimCLR, and our proposed TICL generally outperform prompt-based approaches like SSL-ECG, highlighting the effectiveness of augmentation-based contrastive learning in capturing invariant and discriminative features.

## 5.2 Ablation study

We evaluated the effectiveness of each component in the proposed TICL model by conducting a series of ablation studies. Specifically, we constructed several model variants for comparison. First, we trained a baseline model without using the Transformer architecture, where the encoder output was directly projected to compute similarity between samples—this variant is referred to as TICLa). Second, we removed the LIC module, resulting in the variant denoted as TICLb). Additionally, we investigated the impact of data augmentation strategies by generating both views using either only light augmentation or only heavy augmentation. Specifically, the LIC module is designed to reduce feature redundancy by projecting input representations into a compact latent space, where only the most discriminative and task-relevant information is preserved. This compression is achieved through a contrastive learning objective that encourages representations of augmented views of the same sample to remain close, while pushing apart unrelated samples. By minimizing mutual information between different views in a controlled manner, the LIC module suppresses modality- or dataset-specific noise and spurious correlations. This redundancy reduction is particularly beneficial for cross-dataset generalization, as it helps the model focus on invariant features that are consistently informative across domains. In other words, the LIC module acts as a regularizer that encourages the network to discard superficial patterns that may overfit to a particular dataset's distribution, and instead retain robust features that generalize better to unseen data.

The results are summarized in Table 3. Clearly, the incorporation of the Transformer module, along with the redundancy-reducing LIC strategy, significantly improved the model's cross-dataset generalization and overall accuracy. Furthermore, our analysis of augmentation strategies revealed that using identical augmentation types for both views (either only lightor only heavy) resulted in suboptimal performance. In contrast, the best results were achieved when two distinct augmentation methods—light and heavy—were used to generate diverse but semantically related views. The “identical” setting refers to using the same view of a sample as both the anchor and the positive pair in the contrastive learning process. This effectively removes the augmentation-induced variability that is critical for learning invariant and discriminative features. Without meaningful transformations, the model is not challenged to align representations of semantically equivalent but appearance-varied inputs, which weakens its ability to generalize across diverse conditions. From a theoretical perspective, contrastive learning benefits from augmentations that introduce controlled variability while preserving semantic consistency. These augmentations help the model learn to ignore superficial differences and focus on core signal patterns. The absence of such variability in the “identical” strategy leads to trivial representations and limits the model’s ability to capture invariant features, resulting in poorer performance.

Table 2. Comparison between our proposed models against baseline

| | Dataset | | | | | |
|---|---|---|---|---|---|---|
| | Epilepsy | | HAR | | Sleep-EDF | |
| Baseline | ACC | MF1 | ACC | MF1 | ACC | MF1 |
| SSL-ECG | 93.71 ±0.42 | 89.20 ±0.97 | 65.32 ±1.60 | 63.62 ±1.31 | 65.32 ±1.12 | 30.44 ±1.80 |
| CPC | 96.62 ±0.43 | 94.41 ±0.72 | 83.82 ±2.01 | 84.47 ±2.13 | 88.72 ±1.30 | 39.71 ±3.40 |
| SimCLR | 96.01 ±0.38 | 93.50 ±0.67 | 86.18 ±2.30 | 86.67 ±2.24 | 87.33 ±1.73 | 41.83 ±3.33 |
| TS-TCC | 97.21 ±0.86 | 95.55 ±1.11 | 87.90 ±1.73 | 88.77 ±1.42 | 86.05 ±3.23 | 39.01 ±4.53 |
| Ours | 98.59 ±0.13 | 97.38 ±0.21 | 91.62 ±0.71 | 90.66 ±0.81 | 91.38 ±0.41 | 45.34 ±1.25 |

Table 3. Ablation experiments of effect of different components in TICL

| | Dataset | | | | | |
|---|---|---|---|---|---|---|
| | Epilepsy | | HAR | | Sleep-EDF | |
| Component | ACC | MF1 | ACC | MF1 | ACC | MF1 |
| TICLa) | 90.41 ±1.26 | 86.90 ±1.58 | 78.02 ±1.54 | 70.93 ±1.60 | 75.62 ±1.30 | 32.11 ±1.22 |
| TICLb) | 96.62 ±0.24 | 95.62 ±1.27 | 89.81 ±1.34 | 88.17 ±1.18 | 90.12 ±1.76 | 44.72 ±1.14 |
| Ours | 98.51 ±0.1.8 | 97.32 ±0.28 | 91.62 ±0.78 | 90.60 ±0.88 | 91.30 ±0.42 | 45.32 ±1.25 |
| TICL (Heavy only) | 97.11 ±0.27 | 95.43 ±0.38 | 76.51 ±3.36 | 61.66 ±3.50 | 79.33 ±3.08 | 40.12 ±2.19 |
| TICL (Light only) | 97.41 ±0.42 | 95.71 ±0.89 | 77.14 ±2.70 | 73.72 ±3.44 | 80.01 ±2.29 | 42.02 ±1.80 |

## VI. Conclusion

In this work, we propose a TICL framework for self-supervised representation learning from unlabeled time series data. Specifically, the framework first constructs positive sample pairs for each input by applying two distinct data augmentation strategies. A Transformer-based encoder is then employed to extract temporal features, while the LIC module is integrated to eliminate redundant information across samples, thereby facilitating the learning of robust and discriminative representations. Experimental results demonstrate that the representations learned via TICL when fine-tuned with a small number of labeled samples, achieve high accuracy across various domains and tasks. For instance, TICL achieves an average accuracy improvement of 2% over standard baselines on three widely-used benchmark datasets.

## References

[1] M. A. Farahani, M. R. McCormick, R. Harik, and T. Wuest, "Time-series classification in smart manufacturing systems: An experimental evaluation of state-of-the-art machine learning algorithms", Robotics and Computer-Integrated Manufacturing, Vol. 91, pp. 102839, Feb. 2025. https://doi.org/10.1016/j.rcim.2024.102839.

[2] E. Eldele, M. Ragab, Z. Chen, M. Wu, C.-K. Kwoh, and X. Li, "Label-Efficient Time Series Representation Learning: A Review", IEEE Trans. Artif. Intell., Vol. 5, No. 12, pp. 6027-6042, Dec. 2024. https://doi.org/10.1109/TAI.2024.3430236.

[3] E. Eldele, M. Ragab, Z. Chen, M. Wu, C.-K. Kwoh, and X. Li, "Self-Supervised Learning for Label- Efficient Sleep Stage Classification: A Comprehensive Evaluation", IEEE Trans. Neural Syst. Rehabil. Eng., Vol. 31, pp. 1333-1342, Feb. 2023. https://doi.org/10.1109/TNSRE.2023.3245285.

[4] J. Ye, Q. Xiao, J. Wang, H. Zhang, J. Deng, and Y. Lin, "CoSleep : A Multi-View Representation Learning Framework for Self-Supervised Learning of Sleep Stage Classification", IEEE Signal Process. Lett., Vol. 29, pp. 189-193, Nov. 2021. https://doi.org/10.1109/LSP.2021.3130826.

[5] S. Tipirneni and C. K. Reddy, "Self-Supervised Transformer for Sparse and Irregularly Sampled Multivariate Clinical Time-Series", ACM Trans. Knowl. Discov. Data, Vol. 16, No. 6, pp. 1-17, Dec. 2022. https://doi.org/10.1145/3516367.

[6] P. Sermanet, et al., "Time-Contrastive Networks: Self-Supervised Learning from Video", 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, pp.

1134-1141, May 2018. https://doi.org/10.1109/ICRA.2018.8462891.

[7] R. Güldenring and L. Nalpantidis, "Self-supervised contrastive learning on agricultural images", Computers and Electronics in Agriculture, Vol. 191, pp. 106510, Dec. 2021. https://doi.org/10.1016/j.compag.2021.106510.

[8] M. Jing, Y. Zhu, T. Zang, and K. Wang, "Contrastive Self-supervised Learning in Recommender Systems: A Survey", ACM Trans. Inf. Syst., Vol. 42, No. 2, pp. 1-39, Mar. 2024. https://doi.org/10.1145/3627158.

[9] F. Fahimi, S. Dosen, K. K. Ang, N. Mrachacz-Kersting, and C. Guan, "Generative Adversarial Networks-Based Data Augmentation for Brain–Computer Interface", IEEE Trans. Neural Netw. Learning Syst., Vol. 32, No. 9, pp. 4039-4051, Sep. 2021. https://doi.org/10.1109/TNNLS.2020.3016666.

[10] P. Sarkar and A. Etemad, "Self-Supervised ECG Representation Learning for Emotion Recognition", IEEE Trans. Affective Comput., Vol. 13, No. 3, pp. 1541-1554, Jul. 2022. https://doi.org/10.1109/TAFFC.2020.3014842.

[11] G. Zhang, V. Davoodnia, and A. Etemad, "PARSE: Pairwise Alignment of Representations in Semi-Supervised EEG Learning for Emotion Recognition", IEEE Trans. Affective Comput., Vol. 13, No. 4, pp. 2185-2200, Oct. 2022. https://doi.org/10.1109/TAFFC.2022.3210441.

[12] S. Dai, et al., "Contrastive Learning of EEG Representation of Brain Area for Emotion Recognition", IEEE Trans. Instrum. Meas., Vol. 74, pp. 1-13, Jan. 2025. https://doi.org/10.1109/TIM.2025.3533618.

[13] H. Kim, J. Kim, and S. B. Kim, "Cross-subject generalizable representation learning with class-subject dual labels for biosignals", Knowledge-Based Systems, Vol. 295, pp. 111855, Jul. 2024. https://doi.org/10.1016/j.knosys.2024.111855.

[14] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations", Proc. of the 37th International Conference on Machine Learning, Online, Vol. 119, Jul. 2020.

[15] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum Contrast for Unsupervised Visual Representation Learning", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 9729-9738, Jun. 2020.

[16] J.-B. Grill, et al. "Bootstrap your own latent: A new approach to self-supervised Learning", NIPS'20, Vancouver BC Canada, No. 1786, pp. 21271-21284, Dec. 2020.

[17] X. Yang, Z. Zhang, and R. Cui, "TimeCLR: A self-supervised contrastive learning framework for univariate time series representation", Knowledge-Based Systems, Vol. 245, pp. 108606, Jun. 2022. https://doi.org/10.1016/j.knosys.2022.108606.

[18] X. Zhang, Z. Zhao, T. Tsiligkaridis, and M. Zitnik, "Self-Supervised Contrastive Pre-Training For Time Series via Time-Frequency Consistency in Advances in Neural Information Processing Systems", NIPS'22, New Orleans LA USA, No. 288, pp. 3988-4003, Nov. 2022.

[19] J. Liu and S. Chen, "TimesURL: Self-Supervised Contrastive Learning for Universal Time Series Representation Learning", AAAI, Vol. 38, No. 12, pp. 13918-13926, Feb. 2024. https://doi.org/10.1609/aaai.v38i12.29299.

[20] A. van den Oord, Y. Li, and O. Vinyals, "Representation Learning with Contrastive Predictive Coding", arXiv:1807.03748, Jul. 2018. https://doi.org/10.48550/arXiv.1807.03748.

[21] E. Eldele, et al., "Time-Series Representation Learning via Temporal and Contextual Contrasting", arXiv:2106.14112, Jun. 2021. https://doi.org/10.48550/arXiv.2106.14112.

## Authors


### Kunpeng Li

2019. 6 : BS degree, Dept. of Business Administration, Qilu University of Technology
2022. 6 : MS degree, Dept. of Computer Science Eng., University of Jinan
2022. 9 ~ Present : PhD candidate, Dept. of Electronic Convergence Eng. Kwangwoon University
Research interests : Deep learning/Human Machine Interaction

### Jong-Chul Lee

1983. 2 : BS degree, Dept. of Electronic Eng, Hanyang University
1985. 2 : MS degree, Dept. of Electronic Eng, Hanyang University
1990. 12 : MS degree, Dept. of Electrical Eng, Arizona State University
1994. 5 : PhD degree, Dept. of Electrical Eng, Texas A&M University
1996. 3 ~ Present : Professor, Dept. of Electronic Convergence Eng., Kwangwoon University
Research interests : RF/Microwave Circuit Design/Sensors/Machine Learning