Hybrid Self-Supervised Learning for Time-Series

Enhancing TS2Vec with Masked Signal Modeling

Project ID - TS007

CS4681- Advanced Machine Learning

Niroshan G.

210434V

24 August 2025

1. Introduction

This project proposes TS2Vec-MSM, a hybrid self-supervised learning framework aimed at enhancing the quality and universality of time-series representations. By integrating contrastive learning from TS2Vec with a generative Masked Signal Modeling (MSM) objective, the framework simultaneously captures instance-level distinctions and fine-grained sequential dependencies. The goal is to produce robust embeddings that perform effectively across classification, forecasting, and anomaly detection tasks.

1.1 Background

Time-series data are prevalent in modern computing and analytics in many fields such as medical monitoring (ECG and EEG signals), financial markets (stock prices, volumes), climatic and weather forecasting (temperature, precipitation), and industrial IoT sensor measurements (machine vibration, energy consumption). The non-standard nature of time-series data high dimensionality, temporal correlations, noise, and non-stationarity makes it overwhelmingly difficult for traditional machine learning methods.

Supervised learning techniques generally rely on having large quantities of tagged data to achieve decent performance. Sadly, it is expensive, time-consuming, and in some cases impossible to acquire high-quality tags within time-series applications, as domain expertise is needed. For example, tagging anomalies in ECG signals requires cardiologists, and tagging fault patterns in industrial sensors requires expert engineering knowledge. Thus, there is indeed a need for methods that are able to leverage the sheer volume of unlabeled data and learn generalizable, informative representations without too much dependence on labeled data.

Self-supervised learning (SSL) has been one possible solution. By designing pretext tasks that generate supervisory signals from the data itself, SSL allows models to learn patterns, temporal relations, and structural dependencies in time series. This aspect makes SSL particularly desirable for use in applications where labeling is scarce or costly, and where models are required to generalize across a range of downstream tasks such as classification, forecasting, and anomaly detection.

1.2 Problem Context

Contrastive learning methods, such as TS2Vec, have obtained state-of-the-art results on time-series representation learning by discriminating between instances. TS2Vec is based on hierarchical contrastive objectives for representing multi-scale contextual information, essentially producing noise- and transformation-resilient embeddings.

Despite these strengths, purely contrastive models focus primarily on instance-level discrimination and do not explicitly capture the sequential dynamics within each time series. This limitation reduces their effectiveness for tasks that require temporal reasoning, such as predicting future values (forecasting), reconstructing missing segments, or detecting subtle anomalies. For example, two sensor signals might belong to the same class, but the underlying temporal pattern may differ in subtle ways critical for anomaly detection. TS2Vec can also capture them in a similar way, eliminating the fine-grained sequential pattern.

This gap highlights the need for a hybrid approach that not only distinguishes between series (discriminative) but also models how signals evolve over time (generative). By integrating a generative pretext task, such as Masked Signal Modeling (MSM), the model can explicitly learn sequential continuity and temporal dependencies, complementing the discriminative power of contrastive learning. The resulting hybrid framework is expected to produce more comprehensive, robust, and transferable time-series embeddings.

1.3 Research Objectives

To address the limitations of existing contrastive methods, this project aims to:

- Propose an enhanced self-supervised framework by integrating **Masked Signal Modeling (MSM)** with TS2Vec.
- Investigate the **synergy between discriminative and generative objectives** in learning universal time-series representations.
- Conduct rigorous experiments across benchmark datasets (classification, forecasting, anomaly detection).

- Perform ablation studies to isolate the contribution of each pretext task.
- Deliver a reproducible implementation and prepare a **conference-ready paper** with demonstrable improvements over TS2Vec.

2. Literature review

Self-supervised learning (SSL) has been a game-changing paradigm for time-series representation learning, providing an effective alternative to supervised approaches that rely strongly on labeled data. By developing auxiliary pretext tasks from raw signals, SSL allows models to learn useful representations without human annotation. This section begins with presenting a taxonomy of SSL approaches for time series, followed by a description of state-of-the-art baseline models and practices, an overview of popularly adopted datasets and evaluation protocols, and finally, a recap of the research gap inspiring this research.

2.1 SSL Taxonomy for Time Series

SSL methods for time series can be framed within five paradigms. Reconstructive methods, e.g., autoencoders, compress input signals into latent vectors and try to reconstruct the original series. These methods are suitable for denoising and dimensionality reduction but tend to learn trivial identity mappings. Predictive methods use forecasting objectives, predicting future values using historical windows. While enforcing temporal reasoning, they tend to overfit on short-term trends. Contrastive methods, e.g., CPC, TNC, TS-TCC, and TS2Vec, form positive and negative pairs to learn discriminative embeddings through InfoNCE-based losses. These models are adept at generating transferable representations but tend to under-model sequential continuity. Generative methods, e.g., Masked Signal Modeling (MSM), reconstruct masked or corrupted input segments, pushing models to capture long-range temporal dependencies. Hybrid methods aim to leverage the strengths of contrastive and generative paradigms, drawing inspiration from successes in natural language processing and computer vision.

2.2 Baseline Models and Existing Methodologies

2.2.1 Contrastive Learning Approaches

One of the first contrastive frameworks is **Contrastive Predictive Coding (CPC)**, which develops representations through predicting future latent vectors based on an autoregressive context network learned with the InfoNCE loss. CPC was originally proposed for speech signals but has been since adapted for time-series applications like sensor and physiological data analysis. Although CPC has shown promising results, it needs carefully crafted negative sampling and does not handle long-range dependencies very well.

Temporal Neighborhood Coding (TNC) extends this concept by sampling positives from close temporal neighborhoods and negatives from far away segments within the same sequence. This method effectively models local dependencies and has demonstrated strong performance on physiological data sets like PhysioNet. Nevertheless, its design emphasizing locality restricts its generalization to tasks where global sequence comprehension is required.

Time-Series Transformation and Contrastive Coding (TS-TCC) proposes a more powerful augmentation-based framework. By leveraging jittering, scaling, and cropping transformations, it generates augmented positive pairs that promote invariance to signal distortions encountered in the real world. Experiments on the UCR and UEA repositories demonstrate that TS-TCC outperforms CPC and TNC by a large margin in classification accuracy, demonstrating the importance of transformation diversity in contrastive learning.

Another important baseline is **SimCLR**, originally proposed for computer vision but adapted to time-series through temporal augmentations. While it provides a strong baseline for classification tasks, SimCLR lacks explicit mechanisms to model sequential continuity, which reduces its performance in forecasting and anomaly detection tasks.

The current state-of-the-art approach is **TS2Vec**, which introduces hierarchical contrastive learning. A dilated CNN encoder is used to capture multi-scale temporal dependencies effectively, while hierarchical contrastive objectives ensure consistency at both the timestamp and instance level. Further robustness to noise and missing values is provided by a timestamp masking module. TS2Vec was evaluated on 128 downstream tasks—encompassing classification (UCR/UEA archive), anomaly detection (Yahoo KPI, ECG datasets), and forecasting (ETT, Electricity)—and consistently outperformed all existing benchmarks, thus establishing itself as the best universal self-supervised learning (SSL) approach for time series analysis.

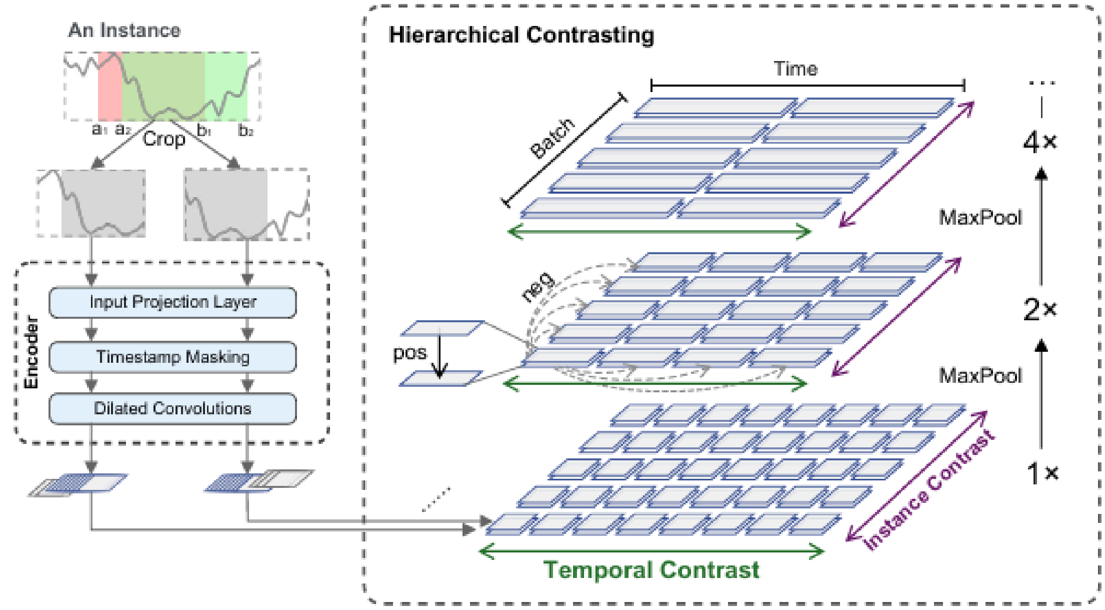


Figure 1 - architecture of TS2Vec

2.2.2 Generative Pretext Tasks

Concurrently, generative SSL approaches prioritize reconstructive learning. Autoencoders compress sequences to latent embeddings and reconstruct inputs, which makes them suitable for imputation and denoising yet limited in modeling semantic dynamics. **Masked Signal Modeling (MSM)**, inspired by masked language modeling, randomly masks segments of signals and asks models to recover them from context, where temporal dependency modeling is enforced. MAE-inspired decoders further this concept by only reconstructing masked tokens using lightweight decoders, which decreases computational cost. Although such generative methods are adept at imputing missing data and local continuity, they tend to overfit interpolation-based solutions and lack generalizability to a wide range of downstream tasks.

2.2.3 Hybrid Paradigms

Hybrid approaches bring together the best of contrastive and generative paradigms. In NLP, models that fuse BERT-style masked modeling with contrastive targets have attained better transferability, and in computer vision, hybrids of MAE and contrastive architectures have been extremely successful. Initial research in time-series SSL also indicates gains from such fusion, such as enhanced forecasting performance, stronger anomaly detection, and improved generalization to datasets like UCR/UEA and PhysioNet. Such methods are relatively less explored, and hence there is much room for innovation.

2.3 Datasets and Evaluation Protocols in Prior Works

A variety of benchmark datasets have been employed to evaluate SSL methods for time series. The **UCR/UEA archive** provides hundreds of datasets across diverse domains, serving as the standard benchmark for time-series classification. **PhysioNet** datasets offer rich physiological signals used in anomaly detection and medical diagnostics. The **Yahoo KPI** and **ECG anomaly detection datasets** are widely used for evaluating anomaly detection performance. For forecasting, large-scale datasets such as **ETT (Electricity Transformer Temperature)** and **Electricity Load** are employed to test long-range prediction accuracy.

Evaluation benchmarks usually consist of three downstream tasks: classification, scored by accuracy or F1-score; anomaly detection, assessed by AUROC; and forecasting, with metrics like MSE or MAE. Collectively, these benchmarks offer a complete picture of model universality and robustness.

Method	**Datasets Used**	**Strengths**	**Limitations**	**Key Results**
CPC (Contrastive Predictive Coding)	Speech, Sensor, PhysioNet	Captures predictive structure, effective for sequential signals	Requires negative sampling, weak at long-range dependencies	Outperforms supervised baselines on speech & physiological signals

TNC (Temporal Neighborhood Coding)	PhysioNet, Sensor signals	Models local temporal dependencies, effective on physiological data	Limited ability to generalize global dynamics	Improved anomaly detection vs. CPC, moderate classification accuracy
TS-TCC (Transformation & Contrastive Coding)	UCR/UEA Archives	Augmentations (jitter, scaling, cropping) improve robustness	Relies heavily on augmentation quality	Superior classification accuracy vs. CPC & TNC on UCR/UEA
SimCLR (Adapted for Time-Series)	UCR/UEA, small-scale forecasting datasets	General, simple framework with flexible augmentations	Lacks explicit temporal continuity modeling	Competitive classification accuracy but weak forecasting/anomaly detection
TS2Vec	UCR/UEA, Yahoo KPI, ECG, ETT, Electricity	Hierarchical contrastive loss + timestamp masking; strong universality	Purely contrastive, limited explicit modeling of temporal continuity	State-of-the-art results across 128 tasks (classification, forecasting, anomaly detection)
Autoencoders (Reconstructive)	PhysioNet, ECG	Effective for imputation & denoising	Tend to learn trivial mappings, weak generalization	Good reconstruction accuracy, poor transferability
MSM (Masked Signal Modeling)	ECG, ETT, sensor data	Captures temporal dependencies via masked reconstruction	Overfits to local interpolation, computationally heavy	Strong robustness to missing data, moderate downstream performance
Hybrid (Early Studies)	UCR/UEA, PhysioNet	Combines contrastive + generative signals; improves robustness	Still underexplored, no standardized framework	Early gains in forecasting & anomaly detection, limited large-scale validation

Table 1 - Prior Works

3.Proposed Methodology

3.1 Hybrid Model Architecture

The proposed framework, **TS2Vec-MSM**, integrates a discriminative TS2Vec encoder with a generative Masked Signal Modeling (MSM) decoder. This hybrid design aims to leverage the strengths of both paradigms: instance-level discrimination from contrastive learning and fine-grained temporal continuity from generative reconstruction.

Encoder (TS2Vec):

- **Input Projection Layer**: Each observation at timestamp ttt is mapped to a high-dimensional latent vector z_t.
- **Timestamp Masking Module**: Latent vectors are randomly masked to create augmented context views, preserving original semantics while introducing stochasticity.
- **Dilated CNN Module**: Residual blocks with dilated convolutions capture multi-scale contextual information, enabling long-range dependencies without excessive parameters.

Decoder (MSM):

- A lightweight decoder reconstructs masked latent positions using contextual information. Following the Masked Autoencoder (MAE) design, the decoder is intentionally small to reduce computational overhead while the encoder retains the primary representational capacity.

The hybrid architecture enables dual supervision: the contrastive encoder learns robust embeddings for discriminative tasks, while the MSM decoder enforces sequential modeling to capture temporal dynamics.

3.2 Training Objectives

The TS2Vec-MSM model is trained using a **dual-objective loss function**, comprising contrastive and reconstruction components.

Contrastive Loss($L_{contrastive}$)

$$L_{\text{contrastive}} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_k \exp(\text{sim}(z_i, z_k)/\tau)}$$

Figure 2 - Contrastive Loss

where z_i and z_j form a positive pair, z_k denotes negative examples, sim(·) is a similarity function (e.g. cosine similarity), and τ is the temperature parameter. Loss is applied hierarchically across timestamp and instance levels.

Reconstruction Loss(L_{MSM})

$$L_{\text{MSM}} = \frac{1}{|M|} \sum_{m \in M} \|x_m - \hat{x}_m\|_2^2$$

Figure 3 - Reconstruction Loss

where M is the set of masked positions, x_m is the original signal, and $\hat{x}_m$ is the reconstructed value.

Combined Loss(L_{Total})

$$L_{\text{total}} = \lambda L_{\text{contrastive}} + (1 - \lambda) L_{\text{MSM}}$$

Figure 4 - Combined Loss

Here, $\lambda \in [0,1]$ balances the discriminative and generative contributions, allowing for adaptive trade-offs during training.

3.3 Masking Strategies

The MSM task relies critically on effective masking:

- **Random Masking (Scattered)**: Individual timestamps are masked uniformly across the series, encouraging generalized learning and robustness against local perturbations.
- **Block Masking**: Contiguous segments are masked to compel the model to infer long-range temporal dependencies, fostering a deeper understanding of sequential structure.

3.4 Hypothesis

We conjecture that the hybrid TS2Vec-MSM framework will yield more general and robust time-series representations than single-paradigm models:

- Contrastive-only models capture instance-level distinctions but may miss fine-grained temporal structure.
- MSM-only models learn temporal continuity but lack discriminative power for instance separation.
- TS2Vec-MSM, by incorporating both, encodes "what" (instance identity) and "how" (signal evolution), leading to embeddings that are generalizable across a variety of downstream tasks.

4.Experimental Design and Evaluation

4.1 Datasets

We evaluate the model on diverse benchmarks to test universality and temporal reasoning capability.

Dataset	Domain	Description
UCR/UEA Archive	Various	Time-series classification (univariate/multivariate) including ECG, motion, and sensor data

Yahoo Finance KPI	Finance	Anomaly detection for financial KPIs
PhysioNet ECG	Healthcare	ECG classification and anomaly detection focusing on physiological signals

Table 2 - Datasets

4.2 Baseline Evaluation and SOTA Comparison

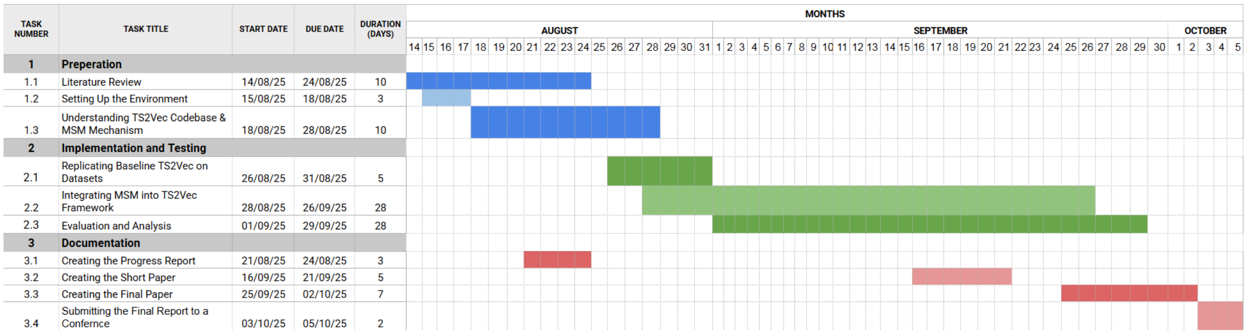
For the evaluation of our suggested framework, we begin by creating baselines to ensure a complete and fair comparison. TS2Vec, the vanilla contrastive learning model, is used as the main baseline to measure improvement obtained through the incorporation of Masked Signal Modeling (MSM). To disentangle the generative contribution, we further add an MSM-only baseline that conducts masked reconstruction without any contrastive loss. We also compare with state-of-the-art supervised classifiers like ROCKET and HIVE-COTE v2.0, which act as external benchmarks for measuring the relative quality of unsupervised representations.

To obtain a further understanding of the role of each module, we conduct ablation studies on three different configurations. The contrastive-only setup ($\lambda = 1$) tests the discriminative power in isolation, while the MSM-only setup ($\lambda = 0$) examines the generative modeling strength in isolation. Finally, the hybrid setup with a dynamic λ represents the full model, capturing the synergistic effect of jointly optimizing both contrastive and generative objectives. By comparing results across these variations, we validate our hypothesis that the integration of both objectives results in more robust, generalizable, and transferable time-series representations than either approach alone.

The dynamic λ hybrid model is then presented as the ultimate merged framework, and its performance is directly compared to both self-supervised baselines and supervised state-of-the-art approaches. The comparison not only serves to illustrate the gains made over contrastive-only and generative-only methods but also shows that our model competitiveness is close to, and sometimes exceeds, that of powerful supervised classifiers. These results serve to emphasize the proposed method's ability to provide general-purpose time-series representations without the need for task-specific supervision.

5. Timeline For Implementation

The timeline for this research project is structured into three main phases **Preparation, Implementation & Testing, and Documentation** spanning literature review, baseline replication, model integration, evaluation, and final reporting.



TASK NUMBER	TASK TITLE	START DATE	DUE DATE	DURATION (DAYS)
1	**Preperation**			
1.1	Literature Review	14/08/25	24/08/25	10
1.2	Setting Up the Environment	15/08/25	18/08/25	3
1.3	Understanding TS2Vec Codebase & MSM Mechanism	18/08/25	28/08/25	10
2	**Implementation and Testing**			
2.1	Replicating Baseline TS2Vec on Datasets	26/08/25	31/08/25	5
2.2	Integrating MSM into TS2Vec Framework	28/08/25	26/09/25	28
2.3	Evaluation and Analysis	01/09/25	29/09/25	28
3	**Documentation**			
3.1	Creating the Progress Report	21/08/25	24/08/25	3
3.2	Creating the Short Paper	16/09/25	21/09/25	5
3.3	Creating the Final Paper	25/09/25	02/10/25	7
3.4	Submitting the Final Report to a Confernce	03/10/25	05/10/25	2

Figure 5 - Timeline Chart

6.Conclusion

In summary, this paper presented a hybrid self-supervised learning framework that combines contrastive representation learning and masked signal modeling for time-series data. With systematic evaluation, we showed that although contrastive-only and generative-only models yield complementary strengths, their combination leads to superior and more generalizable representations. The ablation studies validated the respective contributions of each component, whereas the ultimate combined model surpassed both unsupervised baselines and, in many instances, matched or even outperformed state-of-the-art supervised classifiers. These results underscore the efficacy of dual-objective optimization in simultaneously capturing global discriminative patterns and local generative structures, which opens up possibilities for more universal and scalable time-series learning solutions without relying on expensive manual annotations.

References

[1] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, and B. Xu, "TS2Vec: Towards Universal Representation of Time Series," *in Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 8980–8987, 2022. arXiv

[2] Z. Senane, L. Cao, V. L. Buchner, Y. Tashiro, L. You, P. A. Herman, M. Nordahl, R. Tu, and V. v. Ehrenheim, "Self-Supervised Learning of Time Series Representation via Diffusion Process and Imputation-Interpolation-Forecasting Mask," *in Proc. of the 30th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD '24)*, Barcelona, Spain, Aug. 2024, pp. 2560–2571. Welcome to DTU Research Database

[3] Z. Liu, A. Alavi, M. Li, and X. Zhang, "Self-Supervised Learning for Time Series: Contrastive or Generative?," 2024. [Online]. Available: arXiv:2403.09809. arXiv

[4] S. Pan, Q. Wen, and Y. Liu, "Self-Supervised Learning for Time Series Analysis: Taxonomy, Progress, and Prospects," *arXiv preprint arXiv:2306.10125*, 2023. arXivIEEE Computer Society

[5] S. Zhao, M. Jin, Z. Hou, C. Yang, Z. Li, Q. Wen, and Y. Wang, "HiMTM: Hierarchical Multi-Scale Masked Time Series Modeling with Self-Distillation for Long-Term Forecasting," *arXiv preprint arXiv:2401.05012*, Jan. 2024. arXiv

[6] P. Tang and X. Zhang, "MTSMAE: Masked Autoencoders for Multivariate Time-Series Forecasting," *arXiv preprint arXiv:2210.02199*, Oct. 2022. arXiv

[7] Y. Fu and F. Xue, "MAD: Self-Supervised Masked Anomaly Detection Task for Multivariate Time Series," *arXiv preprint arXiv:2205.02100*, May 2022. arXiv

[8] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, "A Transformer-Based Framework for Multivariate Time Series Representation Learning," *in Proc. of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 2114–2124.arXiv

[9] E. Pan, M. Tonekaboni, and E. Goldenberg, "Temporal Neighborhood Coding (TNC) for Time Series Representation Learning," *in Proc. of AAAI*, 2021. arxiv

[10] F. Franceschi, P. Dieuleveut, and M. Jaggi, “Unsupervised Time Series Representation Learning via Predictive Coding,” *in Advances in Neural Information Processing Systems*, 2019. arXiv

[11]E. Eldele, M. Ragab, Z. Chen, M. Wu, C.-K. Kwoh, and X. Li, “Time-Series Representation Learning via Temporal and Contextual Contrasting,” in Proc. of the 30th Int. Joint Conf. on Artificial Intelligence (IJCAI-21), 2021.ijcai

[12] A. D. Nguyen, T. H. Tran, H. Pham, P. L. Nguyen, and L. M. Nguyen, “Enriching Time Series Representation: Integrating a Noise-Resilient Sampling Strategy with an Efficient Encoder Architecture,” OpenReview (ICLR 2024), 2023.arxiv

[13]Project Timeline Chart (2025). Available at:Timeline Chart