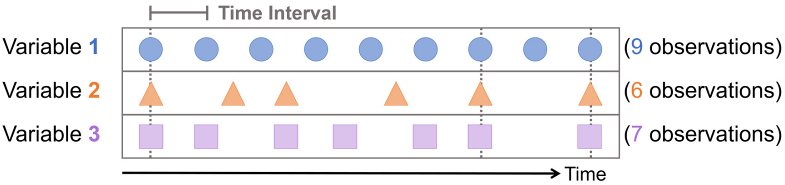
Adaptive Time Encoding for Irregular Multivariate Time-Series Classification

Sangho Lee[1], Kyeongseo Min[2], Youngdoo Son[2], Hyungrok Do[3]

[1]School of Industrial and Systems Engineering, Gyeongsang National University
[2]Department of Industrial and Systems Engineering, Dongguk University-Seoul
[3]Department of Population Health, NYU Grossman School of Medicine

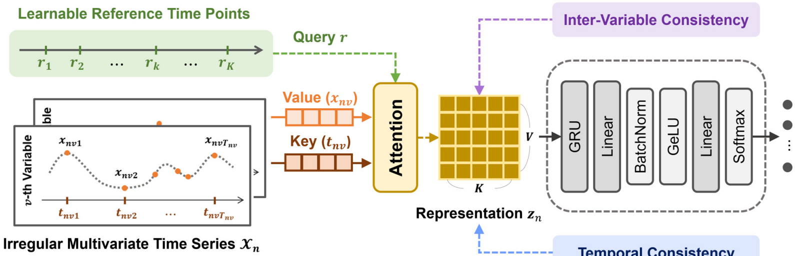
Motivations

- Why Irregular Time Series?
  - In many domains, observations are often recorded at **irregular intervals** [5, 6, 2]
    - ★ Healthcare(medical interventions), finance(non-periodic transactions), sensor failures, etc.
  - In multivariate time series,
    - ★ Observations across variables may not be aligned.
    - ★ The number of observations in each variable can differ.
  - These irregularities hinder the capture of intrinsic patterns.
    $\Rightarrow$ Standard deep learning models often perform poorly in classification tasks [7].



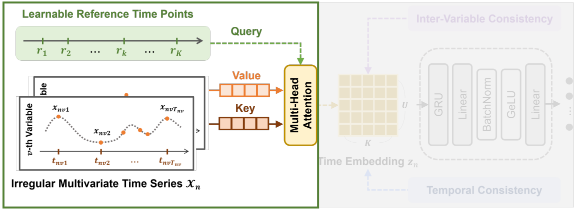
Proposed Method

- Main Contributions
  - Designing a novel **interpolation-based encoder-classifier framework** that learns effective representations for *irregular multivariate time-series classification*
  - Directly **learning optimal reference points** to capture underlying patterns within time series
  - Incorporating **temporal and intervariable consistency regularization** terms to explicitly consider intricate temporal dynamics and relationships across variables
  - Achieving *state-of-the-art performance* with *high computational efficiency*



Proposed Method

- Learnable Reference Time Points
  - Learning a *globally shared set of reference points*, which are not fixed but jointly optimized with model parameters to capture *task-relevant temporal structures across training data*
    - ★ These reference points serve as soft anchors that reflect representative temporal patterns, such as common event timings, even when sequences are irregular or misaligned.
  - While the reference points are shared, *the attention-based interpolation is computed individually for each sample*.



Adaptive alignment of irregular time series to shared temporal structures for robust and efficient representation learning

Proposed Method

Learnable time embedding function:

$$\phi_h(t_{v\tau})[\ell] = \begin{cases} w_{h\ell} \cdot t_{v\tau} + b_{h\ell}, & \text{if } \ell = 1 \\ \sin(w_{h\ell} \cdot t_{v\tau} + b_{h\ell}) & \text{otherwise} \end{cases}$$

Attention mechanism:

$$\kappa_h(r_k, t_{v\tau}) = \frac{e^{\phi_h(r_k)\phi_h(t_{v\tau})^\top/\sqrt{\epsilon}}}{\sum_{\tau'}^{T_v} e^{\phi_h(r_k)\phi_h(t_{v\tau'})^\top/\sqrt{\epsilon}}}$$

Univariate time function:

$$\psi_{hv}(r_k, \mathcal{X}_v) = \sum_{\tau=1}^{T_v} \kappa_h(r_k, t_{v\tau}) \cdot x_{v\tau}$$

Linear combination:

$$\boldsymbol{z}_k[v'] = \sum_{h=1}^{H} \sum_{v=1}^{V} \psi_{hv}(r_k, \mathcal{X}) \cdot W_{hvv'}$$

Following these procedures for all reference points in parallel, we obtain the final representation:

$$\boldsymbol{z} = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_K\} \text{ for } \mathcal{X}.$$

Proposed Method

- Temporal Consistency Regularization

Random masking-based temporal contrastive loss for capturing intricate temporal patterns

Instance-wise contrastive loss function:

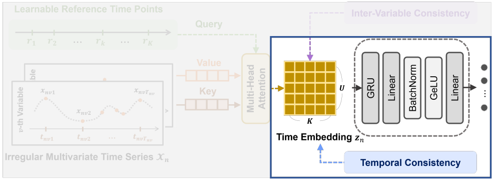
$$\mathcal{L}_{TCI_n} = -\frac{1}{K}\sum_{k=1}^{K}\log\frac{e^{\mathbf{z}_{nk}\cdot\mathbf{z}'_{nk}}}{\sum_{b=1}^{B}\left(e^{\mathbf{z}_{nk}\cdot\mathbf{z}'_{bk}} + \mathbb{1}_{[n\neq b]}e^{\mathbf{z}_{nk}\cdot\mathbf{z}_{bk}}\right)}$$

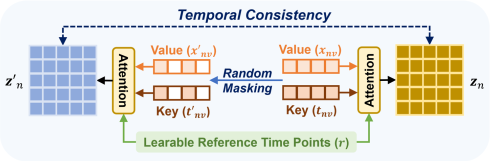
Point-wise contrastive loss function:

$$\mathcal{L}_{TCP_n} = -\frac{1}{K}\sum_{k=1}^{K}\log\frac{e^{\mathbf{z}_{nk}\cdot\mathbf{z}'_{nk}}}{\sum_{k'=1}^{K}\left(e^{\mathbf{z}_{nk}\cdot\mathbf{z}'_{nk'}} + \mathbb{1}_{[k\neq k']}e^{\mathbf{z}_{nk}\cdot\mathbf{z}_{nk'}}\right)}$$

Temporal consistency regularization:

$$\mathcal{L}_{TC_n} = \frac{1}{2}\left(\mathcal{L}_{TCI_n} + \mathcal{L}_{TCP_n}\right)$$





Proposed Method

- Intervariable Consistency Regularization

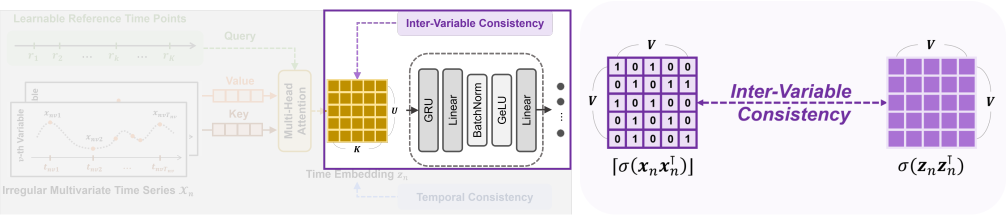
Efficient preservation of cross-variable dependencies via outer product

Two outer product matrices:

$$\mathcal{P}_n = \lceil \sigma(\boldsymbol{x}_n \boldsymbol{x}_n^\top) \rfloor, \;\; \mathcal{Q}_n = \sigma(\boldsymbol{z}_n \boldsymbol{z}_n^\top)$$

Intervariable consistency regularization:

$$\mathcal{L}_{VC_n} = \sum_{(p_{ij}, q_{ij}) \in (\mathcal{P}_n, \mathcal{Q}_n)} p_{ij} \log q_{ij} + (1 - p_{ij}) \log(1 - q_{ij})$$

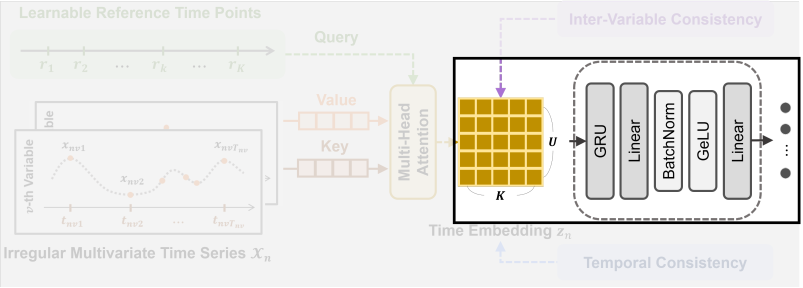


Proposed Method

- Optimization

$$\mathcal{L} = \frac{1}{N}\sum_{n=1}^{N}\left(\mathcal{L}_{CL_n} + \alpha\mathcal{L}_{TC_n} + \beta\mathcal{L}_{VC_n}\right)$$

  - $\mathcal{L}_{CL}$: Classification loss (cross-entropy loss)
  - $\mathcal{L}_{TC}$: Temporal consistency regularization
  - $\mathcal{L}_{VC}$: Intervariable consistency regularization
    - ★ α and β control the temporal and intervariable regularization.



Experiments

- **Classification Performance**
  - How well does ATENet perform compared to existing methods on *irregular multivariate time-series classification*?
- **Robustness to Missing Variables**
  - Can ATENet maintain stable performance even when some variables are missing?
  - Does *intervariable consistency* help capture structural relationships under missingness?
  - *The results on robustness to missing observations, which demonstrate the effectiveness of capturing temporal patterns via temporal consistency, are provided in the manuscript.*
- **Computational Efficiency**
  - Is ATENet more *computationally efficient* than baselines that rely on graph neural networks or complex attention mechanisms?
  - How do its parameter size and processing time compare?
- **Ablation Studies**
  - How much does each key component contribute to ATENet's overall performance?
    - What is the effect of the *learnable reference time points*?
    - How does *temporal consistency regularization* affect temporal stability?
    - How does *intervariable consistency regularization* enhance robustness?

Experimental Settings

- Datasets
  - P12-M (In-hospital mortality) [3], P12-L (Hospitalization length) [3]
  - P19 (Occurrence of sepsis) [9]
  - PAM (Human activity recognition) [8]
- Evaluation metrics
  - AUROC: Area under the receiver operating characteristic curve
  - AUPRC: Area under the precision-recall curve
- Baselines
  - *mTAND* [10], *DGM*^{2} [12], *GRU-D* [1], *MTGNN* [13], *Transformer* [11], *Trans-mean*, *SeFT* [4], *Raindrop* [15], *Warpformer* [14], and *MTSformer* [16]

* *We repeated each experiment five times and reported the averages and standard deviations.*

* *Further details and results of our experiments are provided in the manuscript.*

Experimental Results

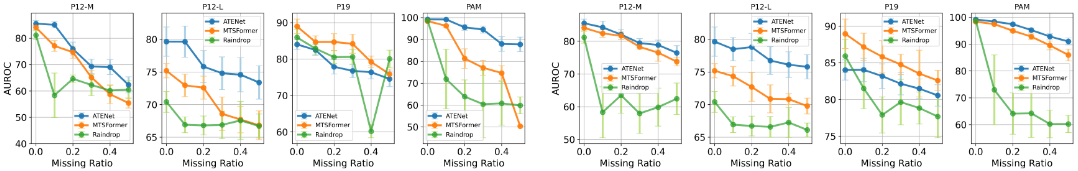
- Classification Performance

Metric	Dataset	mTAND	DGM^2	GRU-D	MTGNN	Transformer	Trans-mean	SeFT	Raindrop	Warpformer	MTSFormer	ATENet
AUROC	P12-M	84.18±1.20	71.08±2.30	48.62±2.41	61.59±5.79	82.92±0.72	83.39±0.56	68.05±1.49	81.19±1.76	79.35±1.65	84.11±0.71	**85.54**±1.26
	P12-L	49.60±3.16	69.46±1.47	49.82±3.85	68.36±6.09	59.05±1.81	61.64±1.54	64.70±2.01	70.40±1.60	74.57±2.28	75.17±1.09	**79.64**±2.24
	P19	80.00±1.23	81.96±2.05	87.16±1.34	85.07±3.54	77.56±3.06	78.57±3.02	77.89±2.62	85.93±2.24	85.41±2.39	**88.96**±2.01	84.02±1.38
	PAM	92.21±0.70	96.87±0.50	91.72±0.59	96.95±0.32	96.61±1.27	97.64±0.25	74.46±6.70	98.73±0.25	97.94±0.45	98.39±0.28	**99.18**±0.15
AUPRC	P12-M	52.89±2.27	29.99±2.24	14.83±1.55	24.25±5.49	46.35±2.81	48.54±2.24	24.43±3.10	42.14±3.32	41.98±1.30	48.53±2.55	**53.31**±2.02
	P12-L	92.42±1.09	96.42±0.41	93.41±0.93	96.41±1.08	94.16±0.99	94.65±0.80	95.28±0.24	96.57±0.51	96.99±0.34	97.43±0.28	**97.70**±0.38
	P19	31.24±4.15	31.12±5.25	47.37±2.97	41.13±8.01	29.60±6.26	28.05±6.23	30.34±1.80	50.63±3.32	41.12±3.30	**57.96**±4.10	41.16±3.02
	PAM	74.95±2.68	88.28±1.28	75.78±2.02	88.85±2.00	86.73±4.21	91.50±0.61	36.43±12.23	95.48±0.91	92.75±1.43	94.21±0.71	**97.61**±0.26
Average Rank		7.50	6.88	8.25	6.75	8.13	6.63	9.13	3.75	4.63	2.38	**2.00**

Table 1: Classification performance of ATENet and baselines. The best score in each dataset is shown in bold.

Experimental Results

- Robustness to Missing Variables
  - *Leave-fixed-sensors-out*: The most informative variables determined by information gain analysis are dropped. The dropped variables are fixed across every sample.
  - *Leave-random-sensors-out*: Missing variables are not fixed but are selected randomly from each sample.



(a) Leave-fixed-sensors-out (b) Leave-random-sensors-out

Figure 4: AUROC scores of ATENet, MTSFormer, and Raindrop when dropping variables with various ratios $\in [0.1, 0.5]$ in (a) *leave-fixed-sensors-out* and (b) *leave-random-sensors-out* scenarios

Experimental Results

- Computational Efficiency

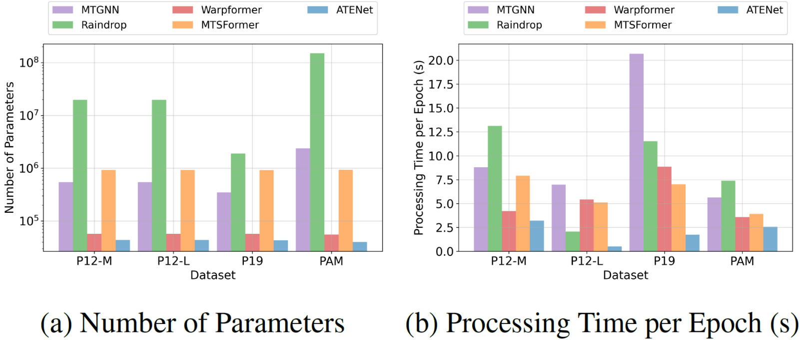


Figure 5: (a) Number of parameters and (b) processing time per epoch for MTGNN, Raindrop, Warpformer, MTSFormer, and ATENet

Experimental Results

- Ablation Studies
  - Fixed reference point vs. **Learnable reference points**
  - *w/o* **Temporal consistency regularization**
  - *w/o* **Intervariable consistency regularization**

Metric	Dataset	Regular	Sparse	Dense	ATENet
AUROC	P12-M	85.49	85.61	**85.65**	85.54
	P12-L	77.26	75.35	77.35	**79.64**
	P19	83.58	82.67	83.46	**84.02**
	PAM	99.00	99.10	98.97	**99.18**
AUPRC	P12-M	51.26	51.22	51.20	**53.31**
	P12-L	97.55	97.54	97.55	**97.70**
	P19	38.06	36.44	38.85	**41.16**
	PAM	96.99	97.13	96.97	**97.61**

Table 2: Classification performance of ATENet and ablation models

Metric	w/o $\mathcal{L}_{TC}$	w/o $\mathcal{L}_{TCP}$	w/o $\mathcal{L}_{TCI}$	w/o $\mathcal{L}_{VC}$
AUROC	0.20	0.13	0.91	3.22
AUPRC	1.28	0.57	1.63	3.33

Table 3: Average performance drop rate (%) of ablation models without consistency regularization compared to ATENet

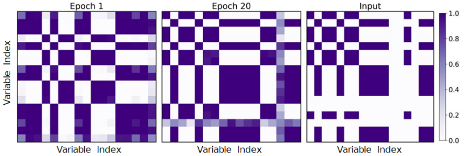


Figure 7: Visualization of intervariable relations from the input and from the learned representations at epochs 1 and 20

Conclusion

- ATENet: Adaptive Time Encoding Network
  - A novel end-to-end framework designed to enhance classification performance on *irregular multivariate time series* by learning their effective representations
    - ★ *Directly learning reference time points* and *generating representations* at these reference points
      → Successfully capturing missingness patterns without information loss caused by disregarding uneven time intervals
      → Finding optimal reference points without the need for an expensive tuning process
    - ★ Introducing *temporal and intervariable consistency regularization terms*
      → Ensuring the enrichment of temporal information
      → Efficiently reflecting intervariable relationships
  - Achieving *state-of-the-art performance* with *high computational efficiency*

Thank You

GitHub OpenReview

Email: sangholee@gnu.ac.kr, youngdoo@dongguk.edu

References

[1] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):6085, 2018.

[2] E. Choi, Z. Xu, Y. Li, M. Dusenberry, G. Flores, E. Xue, and A. Dai. Learning the graphical structure of electronic health records with graph convolutional transformer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 606–613, 2020.

[3] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

[4] M. Horn, M. Moor, C. Bock, B. Rieck, and K. Borgwardt. Set functions for time series. In *International Conference on Machine Learning*, pages 4353–4363. PMLR, 2020.

[5] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4):917–963, 2019.

[6] S. Lee, W. Kim, and Y. Son. Spatio-temporal consistency for multivariate time-series representation learning. *IEEE Access*, 12:30962–30975, 2024.

[7] Z. Li, S. Li, and X. Yan. Time series as images: Vision transformer for irregularly sampled time series. *Advances in Neural Information Processing Systems*, 36, 2024.

[8] A. Reiss and D. Stricker. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th international symposium on wearable computers*, pages 108–109. IEEE, 2012.

[9] M. A. Reyna, C. S. Josef, R. Jeter, S. P. Shashikumar, M. B. Westover, S. Nemati, G. D. Clifford, and A. Sharma. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. *Critical care medicine*, 48(2):210–217, 2020.

[10] S. N. Shukla and B. Marlin. Multi-time attention networks for irregularly sampled time series. In *International Conference on Learning Representations*, 2021.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[12] Y. Wu, J. Ni, W. Cheng, B. Zong, D. Song, Z. Chen, Y. Liu, X. Zhang, H. Chen, and S. B. Davidson. Dynamic gaussian mixture based deep generative model for robust forecasting on sparse multivariate time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 651–659, 2021.

[13] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 753–763, 2020.

[14] J. Zhang, S. Zheng, W. Cao, J. Bian, and J. Li. Warpformer: A multi-scale modeling approach for irregular clinical time series. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3273–3285, 2023.

[15] X. Zhang, M. Zeman, T. Tsiligkaridis, and M. Zitnik. Graph-guided network for irregularly sampled multivariate time series. In *International Conference on Learning Representations*, 2022.

[16] L. N. Zheng, Z. Li, C. G. Dong, W. E. Zhang, L. Yue, M. Xu, O. Maennel, and W. Chen. Irregularity-informed time series analysis: Adaptive modelling of spatial and temporal dynamics. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 3405–3414, 2024.