Causal-oriented representation learning for time-series forecasting based on the spatiotemporal information transformation

Sihua Cai [1,3] Hao Peng [1,2,3] Rui Liu [1] Pei Chen [1]

[1]School of Mathematics, South China University of Technology [2]School of Future Technology, South China University of Technology [3]These authors contributed equally.

TL;DR

CReP: Causal-oriented Representation Learning Predictor

is a self-supervised framework that jointly performs multi-step time-series forecasting and causal analysis by decomposing high-dimensional observations into orthogonal cause, effect, and non-causal representations for a chosen target variable.

Introduction

Objectives

- From high-dimensional time-series observations:
  - **Predict** the future trajectory of a selected target variable over multiple time steps.
  - **Identify** which observed variables are likely **drivers** of the target and which are likely **responses**.

Problem

- Existing approaches forecast well but remain non-causal, or infer causality separately.

Importance of the paper

- Enables **simultaneous forecasting and causal analysis** in high-dimensional time-series systems.
- Provides **interpretable causal insights** (causes vs. effects) rather than black-box predictions.

Grounding

- Dynamic causation
  - If variable a drives b in a dynamical system, then information about a is embedded in the time series of b.
  - Using delay embeddings, there exists an implicit mapping (Takens):

$$A_t = (a_t, \dots, a_{t+L-1})^\top, \quad B_t = (b_t, \dots, b_{t+L})^\top, \quad A_t \approx h(B_t)$$

- Spatiotemporal information transformation (delay $\leftrightarrow$ non-delay)
  - STI transformation allows a learned non-delay embedding C_t to be topologically conjugate to B_t: $A_t \approx \tilde{h}(C_t)$
  - This motivates learning a latent Z_t from which Y_t can be predicted: $Y_t \approx g(Z_t)$

Method

The self-supervised framework is characterized by three key features:

1. Dynamic causation detection with the STI transformation mechanism.
2. Causal-oriented representation learning for multi-step predictions through the CReP.
3. Causal analysis of the target variable via $\alpha\beta$-LRP.

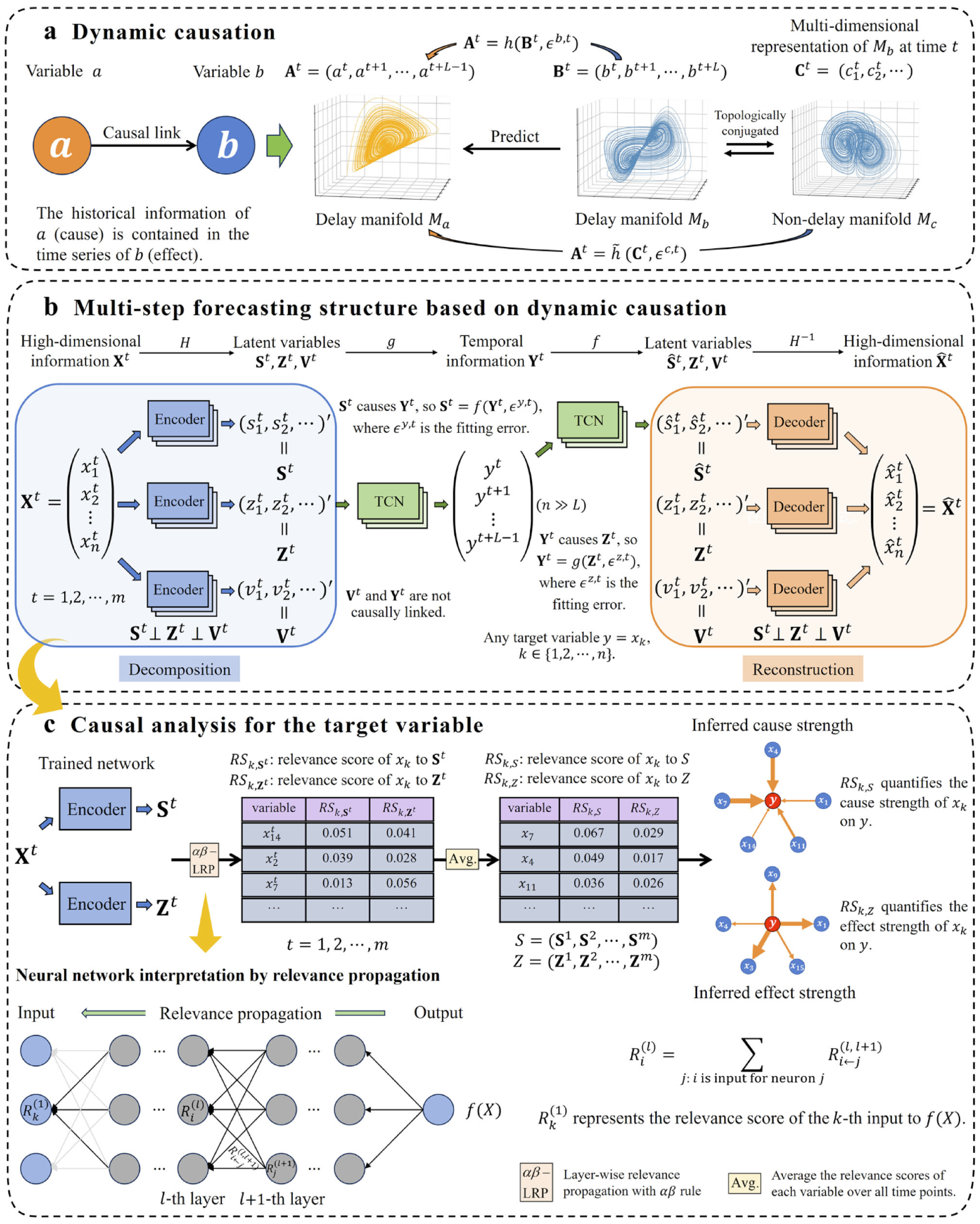

Figure 1. Schematic illustration of the CReP framework.

Loss

$$\mathcal{L} = \lambda_1 \mathcal{L}_{DS} + \lambda_2 \mathcal{L}_{FC} + \lambda_3 \mathcal{L}_{REC} + \lambda_4 \mathcal{L}_{ORTH}$$

- **Determined-State Loss**: RMSE on known historical y.
- **Future-Consistency Loss**: RMSE between overlapping future estimates.
- **Reconstruction Loss**: assesses information recovery
  - $\mathcal{L}_{REC\_X}$: spatiotemporal information X from (S, Z, V)
  - $\mathcal{L}_{REC\_Y}$: latent cause representation S recovered by Y
- **Orthogonality Loss**: enforces orthogonality among S, Z, and V.

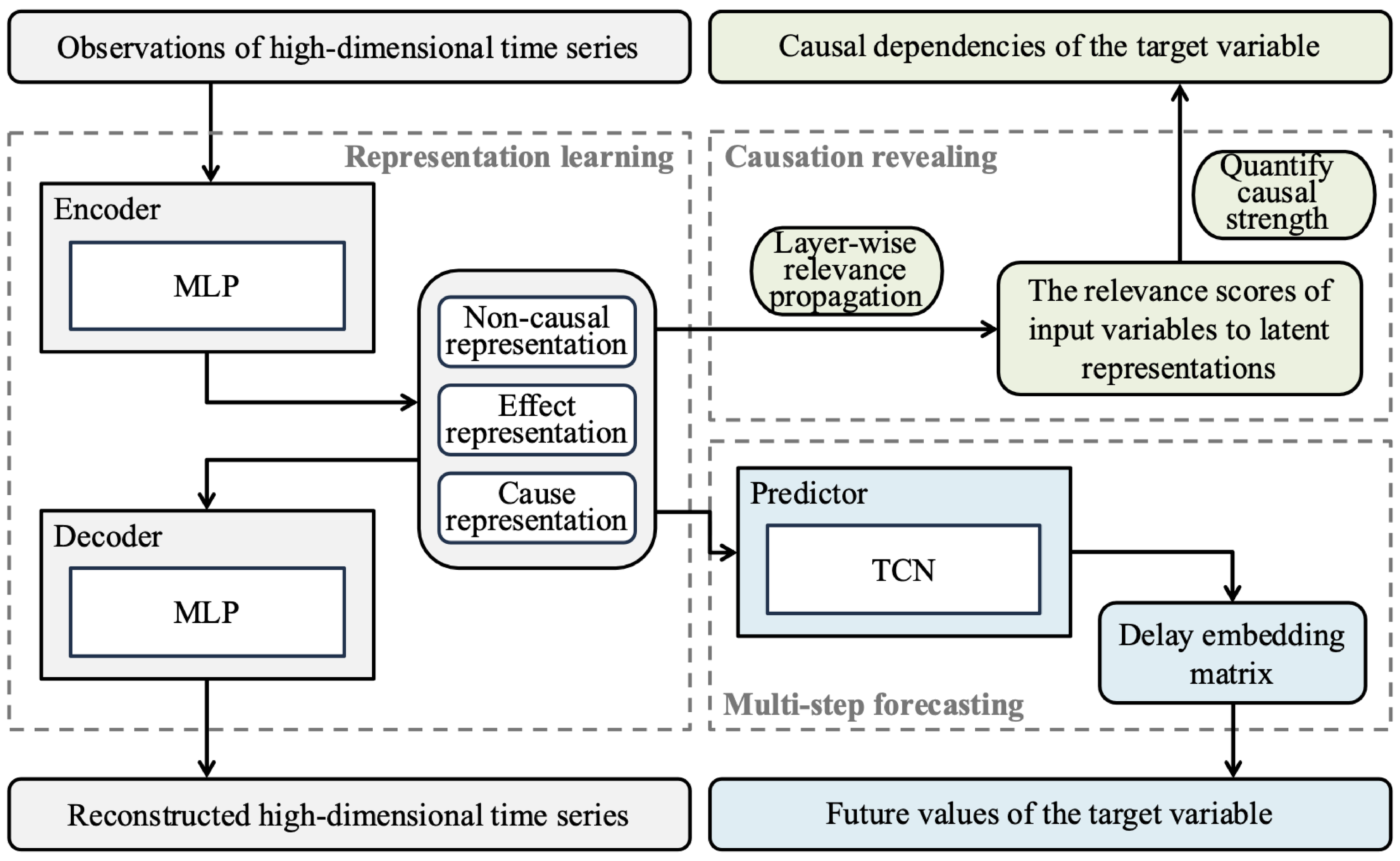

Figure 2. The flowchart of CReP method.

Results

Simulation systems

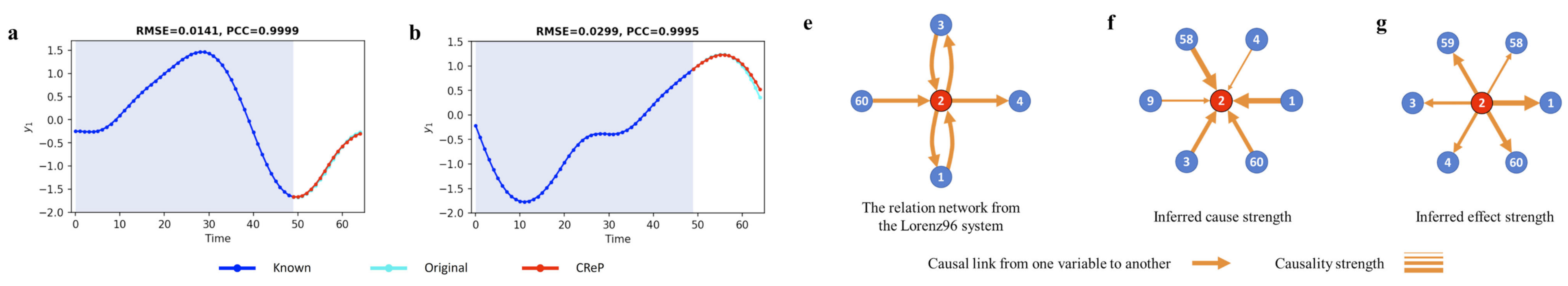

Figure 3. CReP on Lorenz 96 (50-step history, 15-step-ahead prediction, 60D system).

	Lorenz 96		Power grid		Dream4	
Method	RMSE ↓	PCC ↑	RMSE ↓	PCC ↑	RMSE ↓	PCC ↑
CReP	**0.1086**	**0.9908**	**0.1127**	**0.9868**	**0.0984**	0.8894
ARNN	0.2328	0.8206	0.1501	0.9784	0.2281	0.8167
STICM	0.1822	0.9752	0.1530	0.9854	0.1080	**0.9928**
LSTM	0.4883	0.8303	0.3916	0.9659	0.8063	0.0615
ARIMA	0.4204	0.6156	0.2558	0.9583	0.2650	0.1218
SVR	0.8311	0.1618	0.4856	0.9540	0.6779	0.0566
Informer	0.1543	0.9005	0.1163	0.9837	0.4716	0.0489

Table 1. Comparison to other forecasting methods.

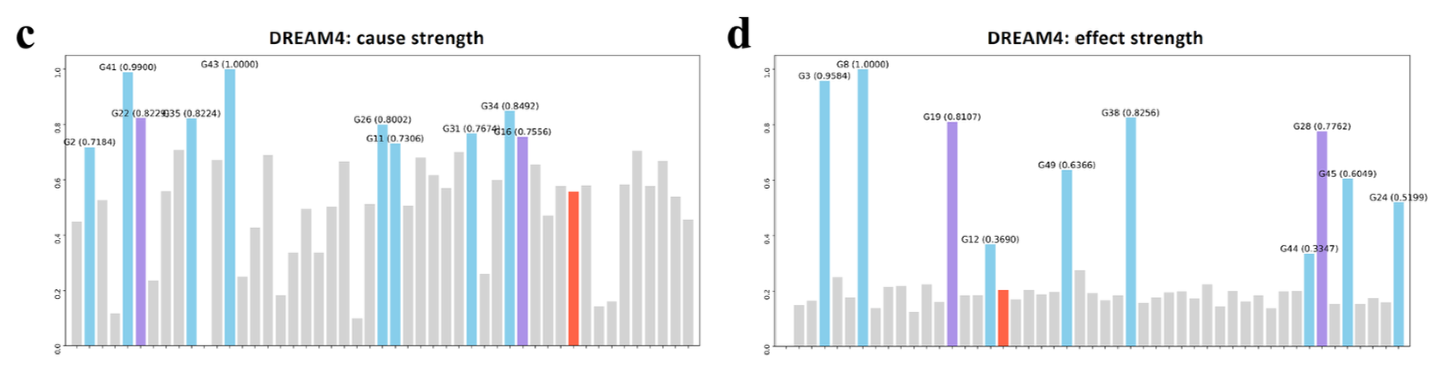

Figure 4. The normalized causal results of CReP on Dream4. Blue bars denote the top 10 variables by causal strength, purple bars highlight true causes or effects among them, and red bars indicate true causes or effects that were missed.

Real datasets

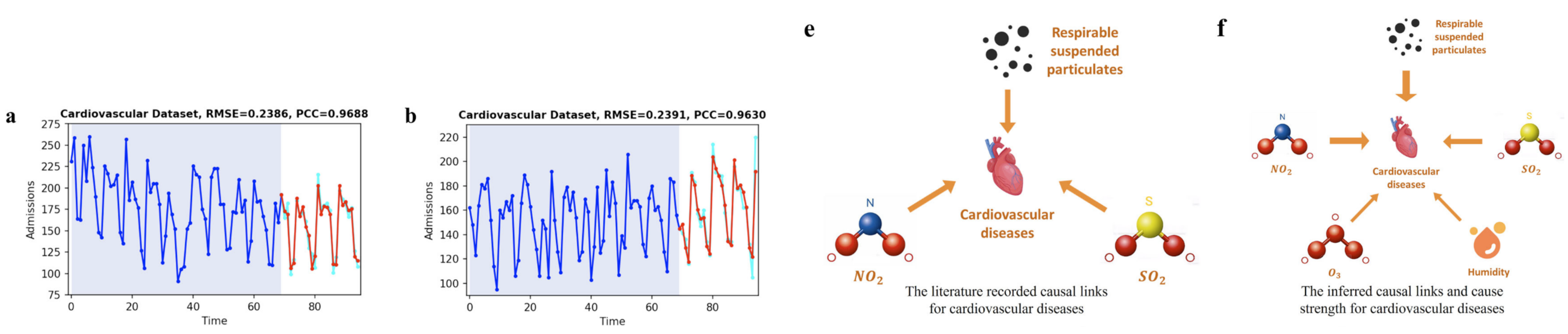

Figure 5. CReP on Hong Kong cardiovascular inpatients (70-day history, 25-day-ahead forecasting, 14 variables).

Conclusion

Key Points

- Causes leave **detectable traces** in the dynamics of what they influence.
- CReP learns to **extract these traces** from high-dimensional time series.
- These representations are used to **predict the future** and **identify causes and effects**.

Critical discussion

- Presence of **false positives** and **missed true causes** in Fig. 4.
- Table 1 is limited to relatively **basic baselines** and does not include SOTA methods.
- Use of different hyperparameter choices across datasets.

Future Work

- Integrating different causal learning methods to enhance applicability across diverse domains.
- Exploring causal detection methods through active intervention rather than passive observations.

References

[1] Sihua Cai, Hao Peng, Rui Liu, and Pei Chen. Causal-oriented representation learning for time-series forecasting based on the spatiotemporal information transformation. *Communications Physics*, 8(1):242, 2025.