Causal Representation Learning from Multimodal Biomedical Observations

Yuewen Sun[1,2*], Lingjing Kong[2*], Guangyi Chen[1,2], Loka Li[1], Gongxu Luo[1], Zijian Li[1], Yixuan Zhang[1], Yujia Zheng[2], Mengyue Yang[3], Petar Stojanov[4], Eran Segal[1], Eric P. Xing[1,2], Kun Zhang[1,2]

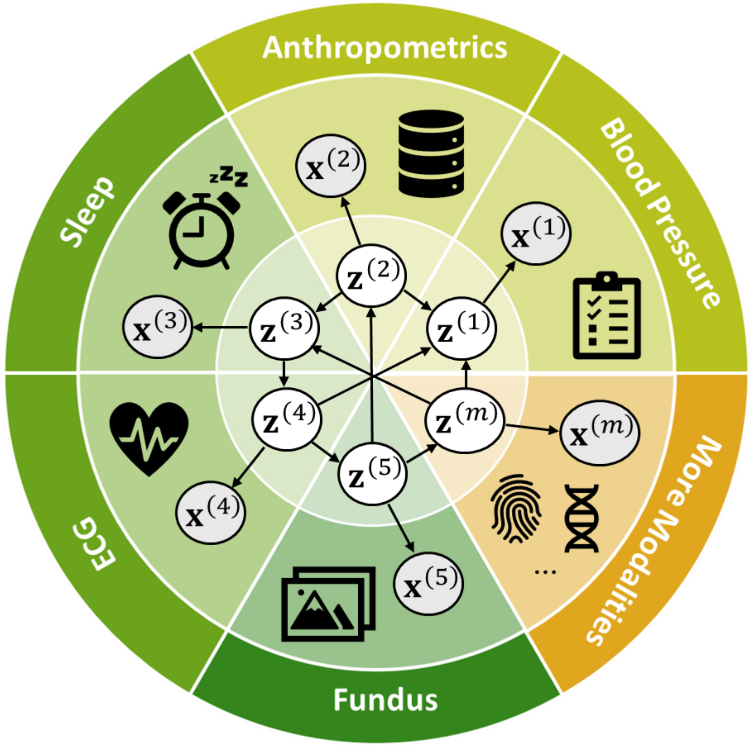
[1]Mohamed bin Zayed University of Artificial Intelligence, [2]Carnegie Mellon University,

[3]University of Bristol, [4]Broad Institute of MIT and Harvard

* Equal contribution

Background

- Biomedical dataset involves unique and related modalities



Tabular data: demographic, blood pressure, ⋯
Image data: fundus image, bone density, ⋯
Sequential data: sleep monitoring, ECG, ⋯

Discovery of novel molecular markers

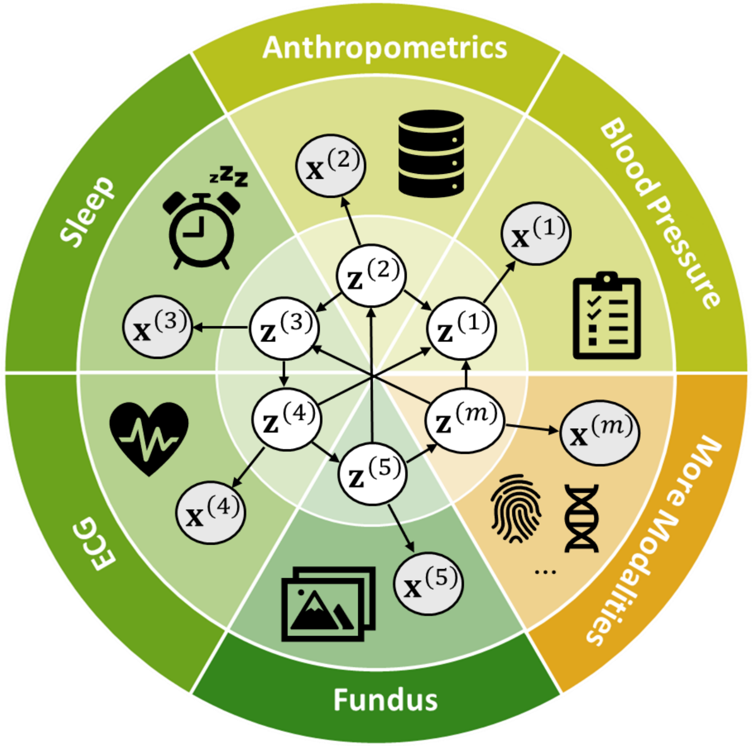
- (Gen, protein, RNA) → Lung cancer

Development of predictive models for disease

- (Health records, medical images, wearable data) → Type 2 diabetes

Background

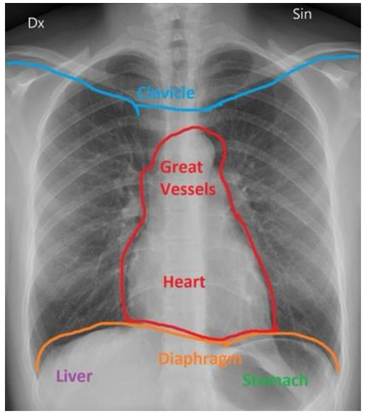
- Biomedical dataset involves unique and related modalities
- Biomedical interactions may be governed by some causally-related unobserved latent variables



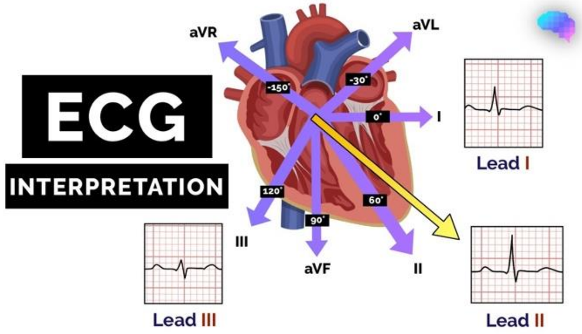
Heart health

Heart size

Heart rate variability



X-ray Image



ECG

Recent Advance

- Large-scale model
  - Exploit biomedical datasets for various tasks
  - Lack interpretability could be a big issue

- Causal representation learning
  - Identify latent causal structures directly from raw data
  - Prior work:
    - Identify latent subspaces shared by multiple modalities
    - Rely on specific assumptions about the latent variables

Motivation

- Large-scale model
  - Exploit biomedical datasets for various tasks
  - Lack interpretability could be a big issue

- Causal representation learning
  - Identify latent causal structures directly from raw data
  - Prior work:
    - Identify latent subspaces shared by multiple modalities
    - Rely on specific assumptions about the latent variables

- **Theoretically**, we provide identifiability guarantees for each latent component
- **Empirically**, we develop a theoretically grounded estimation framework to recover the latent components in each modality

Problem Formulation

- A set of observations/measuremnets from M modalities: $\mathbf{x} \coloneqq [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}]$
- A set of causally related latent variables underlying M modalities: $\mathbf{z} \coloneqq [\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(M)}]$

- Latent causal relations: $z_i^{(m)} \coloneqq g_{z_i^{(m)}}\left(\mathrm{Pa}\left(z_i^{(m)}\right), \epsilon_i^{(m)}\right)$
- Data generating functions: $\mathbf{x}^{(m)} \coloneqq g_{\mathbf{x}^{(m)}}(\mathbf{z}^{(m)}, \boldsymbol{\eta}^{(m)})$

Level 1 -- Subspace Identifiability

- The estimated latent subspace $\hat{\mathbf{z}}^{(m)}$ for any modality m and its true counterpart $\mathbf{z}^{(m)}$ are equivalent up to an invertible map $h^{(m)}$: i.e., $\hat{\mathbf{z}}^{(m)} = h^{(m)}(\mathbf{z}^{(m)})$

A1: Smoothness & Invertibility

Theorem 4.2 (Subspace Identifiability). *Let* $\boldsymbol{\theta} := \{g_{\mathbf{x}^{(m)}}, \tilde{g}_{\mathbf{z}^{(-m)}}, p(\boldsymbol{\epsilon}^{(m)}), p(\tilde{\boldsymbol{\epsilon}}^{(-m)})\}_{m=1}^{M}$ *and* $\hat{\boldsymbol{\theta}} := \{\hat{g}_{\mathbf{x}^{(m)}}, \hat{\tilde{g}}_{\mathbf{z}^{(-m)}}, p(\hat{\boldsymbol{\epsilon}}^{(m)}), p(\hat{\tilde{\boldsymbol{\epsilon}}}^{(-m)})\}_{m=1}^{M}$ *be two specifications of the data-generating process in Eq. (3). Suppose that they generate identical observational distributions (i.e.,* $p(\mathbf{x}) = \hat{p}(\mathbf{x})$*),* $\boldsymbol{\theta}$ *satisfies Condition 4.1, and* $\hat{\boldsymbol{\theta}}$ *satisfies Condition 4.1-A1. The latent subspace* $\hat{\mathbf{z}}^{(m)}$ *for any group* m *and its counterpart* $\mathbf{z}^{(m)}$ *are equivalent up to an invertible map* $h^{(m)}(\cdot)$*, i.e.,* $\hat{\mathbf{z}}^{(m)} = h^{(m)}(\mathbf{z}^{(m)})$.

A1: Smoothness & Invertibility +
A2: Linear Independence

- Information of the subspace $\mathbf{z}^{(m)}$ is preserved in its corresponding observation $\mathbf{x}^{(m)}$ and exerts sufficient influence on other modalities' observations $\mathbf{x}^{(-m)}$

Level 2 -- Component-wise Identifiability

- Further disentangle each subspace into individual components

A1: Smoothness & Invertibility +
A2: Linear Independence

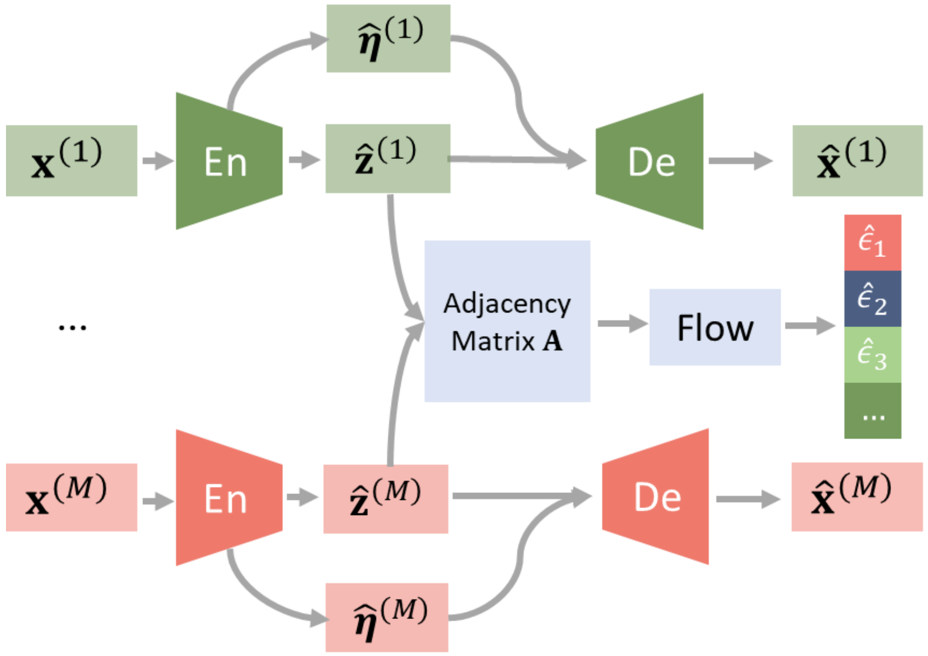
A3: Component Identifiability Conditions

Theorem 4.4 (Component-wise Identifiability). *Let $\boldsymbol{\theta} := (\{g_{\mathbf{x}^{(m)}}, g_{\mathbf{z}^{(m)}}, p(\boldsymbol{\epsilon}^{(m)})\}_{m=1}^{M})$ and $\hat{\boldsymbol{\theta}} := (\{\hat{g}_{\mathbf{x}^{(m)}}, \hat{g}_{\mathbf{z}^{(m)}}, \hat{p}(\boldsymbol{\epsilon}^{(m)})\}_{m=1}^{M})$ be two specifications of the data-generating process in Eq. (1) and Eq. (2). Suppose that they generate identical observational distributions (i.e., $p(\mathbf{x}) = \hat{p}(\mathbf{x})$) and $\boldsymbol{\theta}$ satisfies Condition 4.1 and Condition 4.3. If $\hat{\boldsymbol{\theta}}$ satisfies the following sparse regularization condition:*

$$\sum_{m \neq n \in [M]} \left\| [\hat{\mathbf{G}}]_{(m),(n)} \right\|_0 \leq \sum_{m \neq n \in [M]} \left\| [\mathbf{G}]_{(m),(n)} \right\|_0, \tag{5}$$

each component $z_i^{(m)}$ and its counterpart $\hat{z}_{\pi(i)}^{(m)}$ are equivalent up to an invertible map $h(\cdot)$, i.e., $\hat{z}_{\pi(i)}^{(m)} = h(z_i^{(m)})$ under a permutation π over $[d(\mathbf{z}^{(m)})]$.

Estimation Framework

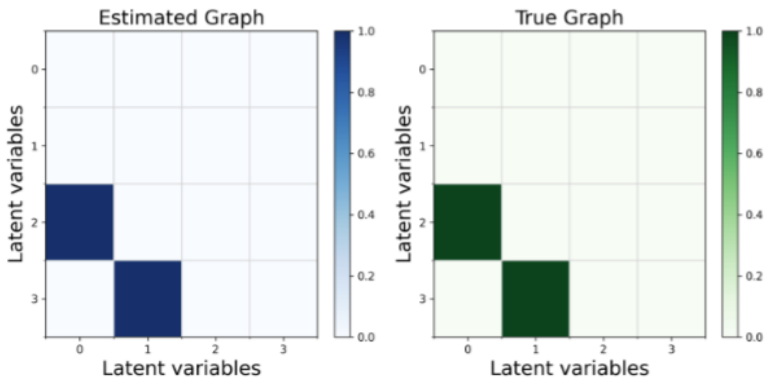


- Encoder-decoder framework
- Corresponding encoder-decoder for each modality
- Learnable adjacency matrix to enforce sparse causal relations
- Combination objective

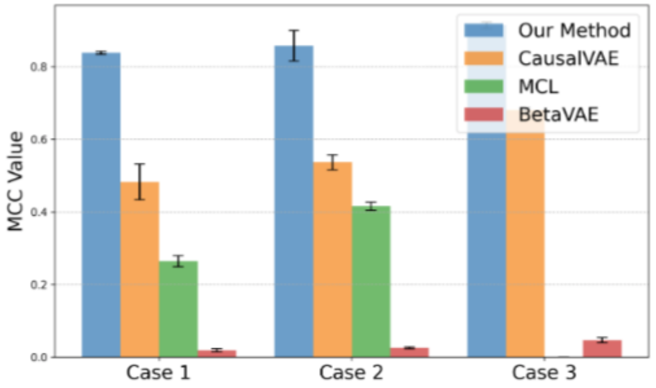
$$\mathcal{L} = \alpha_{\text{Recon}}\mathcal{L}_{\text{Recon}} + \alpha_{\text{Ind}}\mathcal{L}_{\text{Ind}} + \alpha_{\text{Sp}}\mathcal{L}_{\text{Sp}}.$$

Experiments: Synthetic Dataset

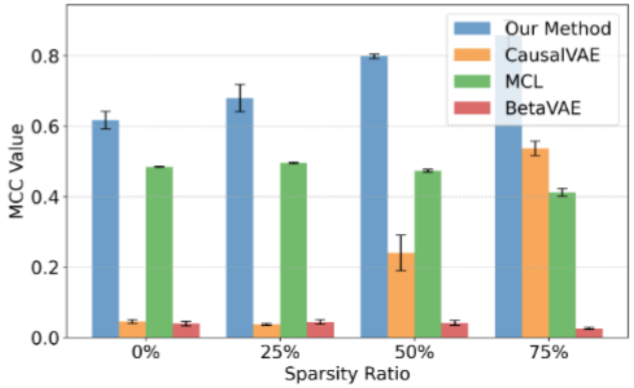
- **Settings**: 15~30 dims observations, with causally-related latent variables in each modality
- **Conclusion**:
  - High MCC show that our method successfully recovers latent variables across all cases
  - Inter-modal causal relations are accurately identified
  - The identifiability improves when the sparsity increases



(a) Causal comparison between estimated and true graphs (SHD=0).



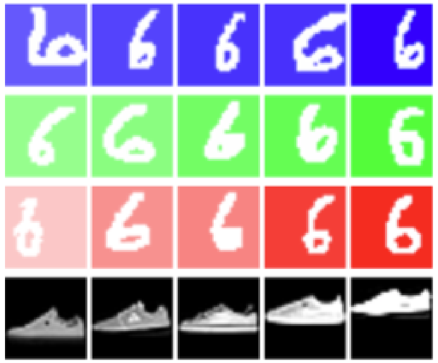
(b) Comparison of the identifiability result in different cases.



(c) Identifiability result under different sparsity ratios.

Experiments: Variant MNIST

- We manually created a variant of the MNIST dataset to encode causal relationships between modalities
- Two different modalities: colored MNIST + fashion MNIST



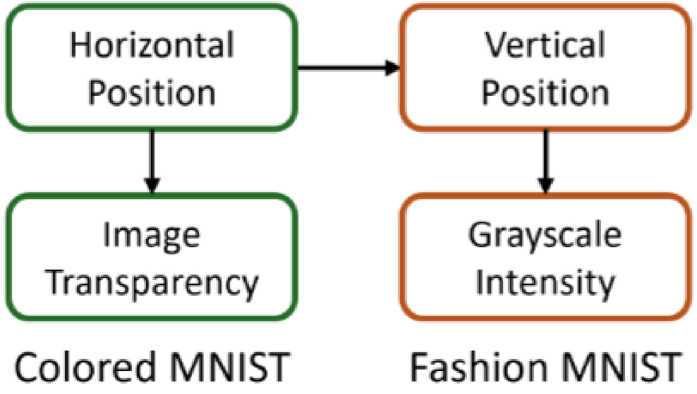


Table 2: The results of MNIST dataset.

	MCL	BetaVAE	CausalVAE	Ours
R2	0.79 ± 6e-5	0.68 ± 2e-3	0.50 ± 4e-3	**0.90** ± 9e-5
MCC	0.63 ± 2e-6	0.53 ± 1e-3	0.74 ± 2e-3	**0.85** ± 3e-5

Experiments: Human Phenotype