CS 520 Causal Inference and Learning

# Benchmarking Causal Discovery Algorithms on Medical Data

Colliders

Giuseppe Concialdi, Judhajit Roy, Vaibhav Jolly, Aditya R

# Motivation

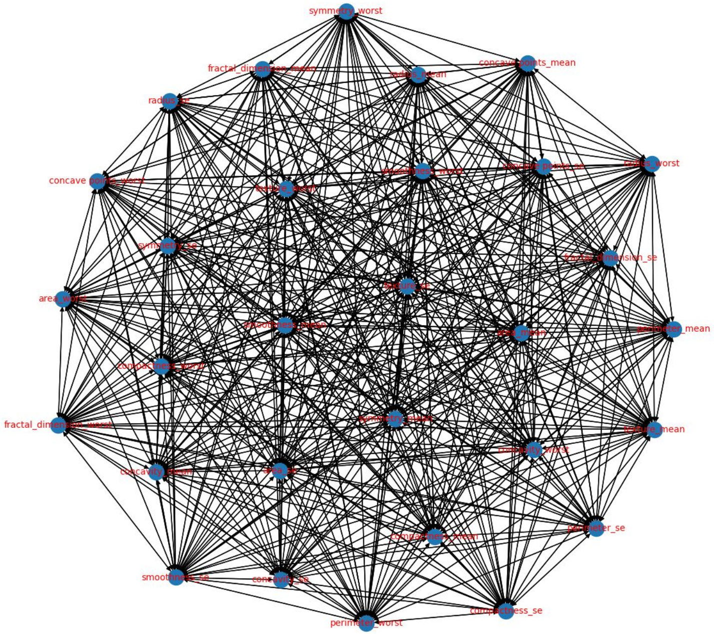
- Identify and validate the causal relationships between variables in medical datasets
- Benchmarking and assess the performance of several causal discovery algorithms
- Explore and compare classic causal discovery algorithms with recent deep-learning based approaches

## What is causal discovery?

Causal discovery is the task of identifying the causal relationships among a set of variables

# Proposed approach

- We explored several causal discovery algorithms:
    - **PC**
    - **GES**
    - **LiNGAM**
    - **NOTEARS**
    - **SAM**
    - **CGNN**
- Each algorithm perform some assumptions on the nature of the underlying data



# Proposed approach

Classic discovery algorithms

PC

GES

LiNGAM

- Conditional independence testing
- Greedy search of DAG space
- Exploiting asymmetry and linearity

- Hybrid techniques
- Few assumption about the data
- Requires large dataset

NOTEARS

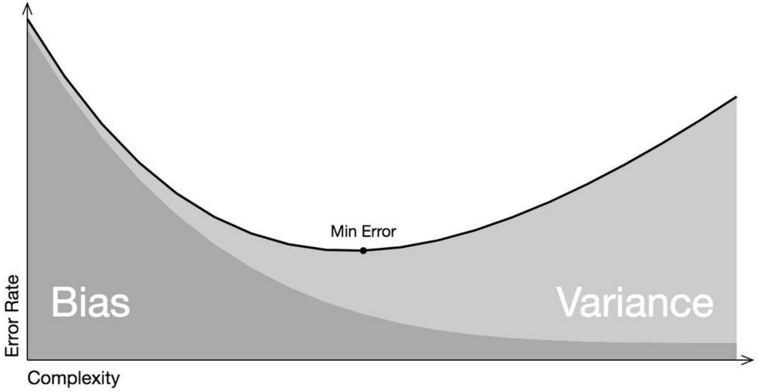
SAM

CGNN

Recent discovery algorithms

# Proposed approach

- **Unconstrained models**
  - The algorithm is applied directly on the data
  - Better predictions
  - Risk of overfitting the dataset
  - Slower evaluation
- **Constrained models**
  - The prediction is performed on top of the regularized data
  - Lower performance
  - Faster evaluation
  - Greater generalization



# Datasets

## Wisconsin Breast Cancer Dataset

- From UCI repository and contains 569 datapoints and 32 attributes.
- Features: Area, Texture and categorical columns like Diagnosis, which contains 357 Benign and 212 Malignant instances.

## Thyroid Dataset

- Also from UCI repository and contains 3500+ datapoints
- Features: hyperthyroid and hypothyroid conditions,  age, sex, sick, pregnant, tumor, TSH, T3, T4, etc

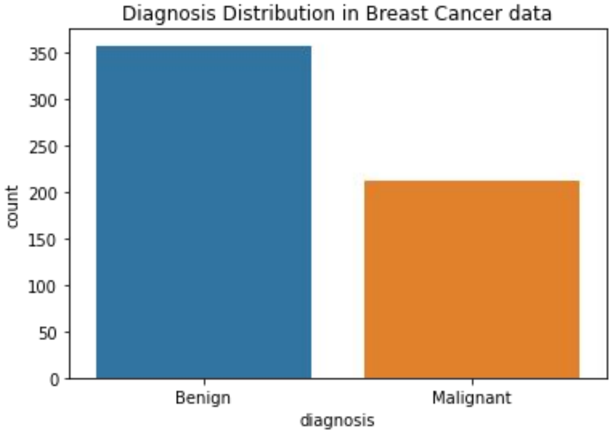
## Cardiovascular Disease Dataset

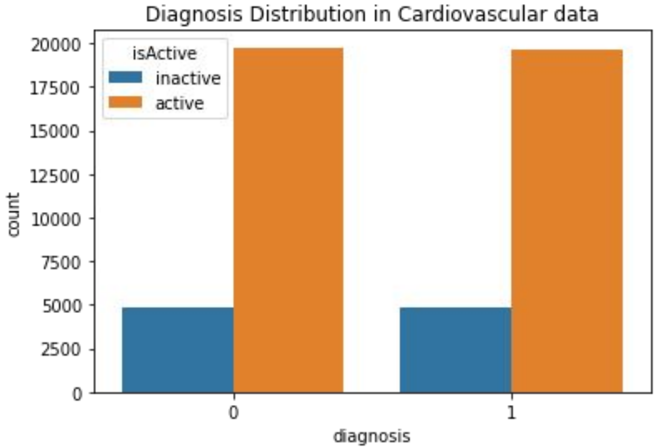
- Open-source large dataset on Kaggle. Contains 70,000 patient records and 11 features
- 34,979 patients are diagnosed with cardiovascular disease, and 35,021 are not.

# Data Exploration

Data pre-processed for all datasets:

- Irrelevant columns dropped, and missing data either replaced or removed
- Data was scaled in all datasets to ensure better graphical representation
- Labels encoded for categorical attributes like '*diagnosis*', '*gender*' and '*referral source*'

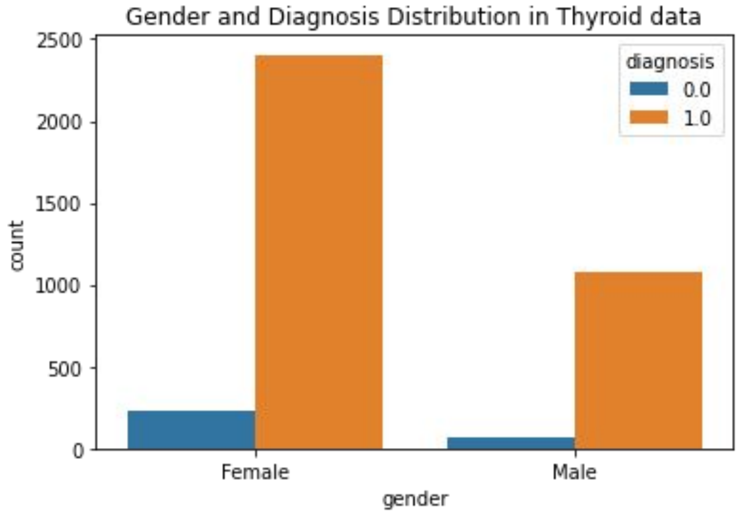


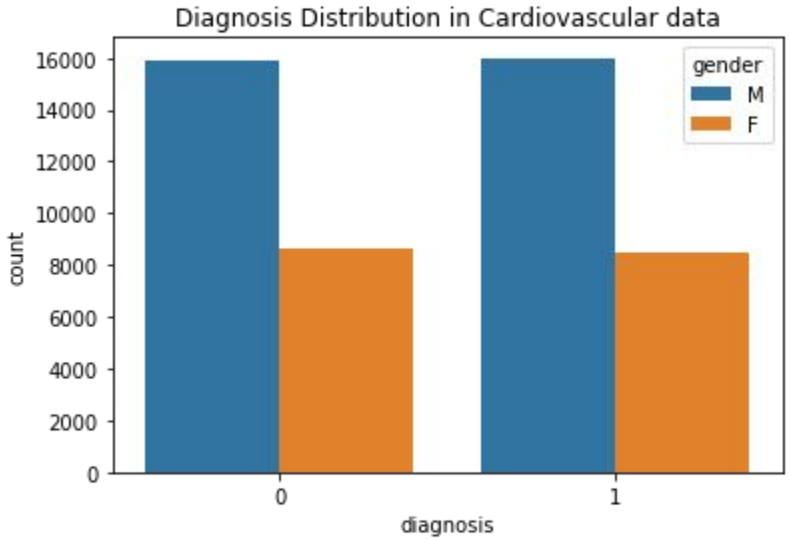



# Data Exploration

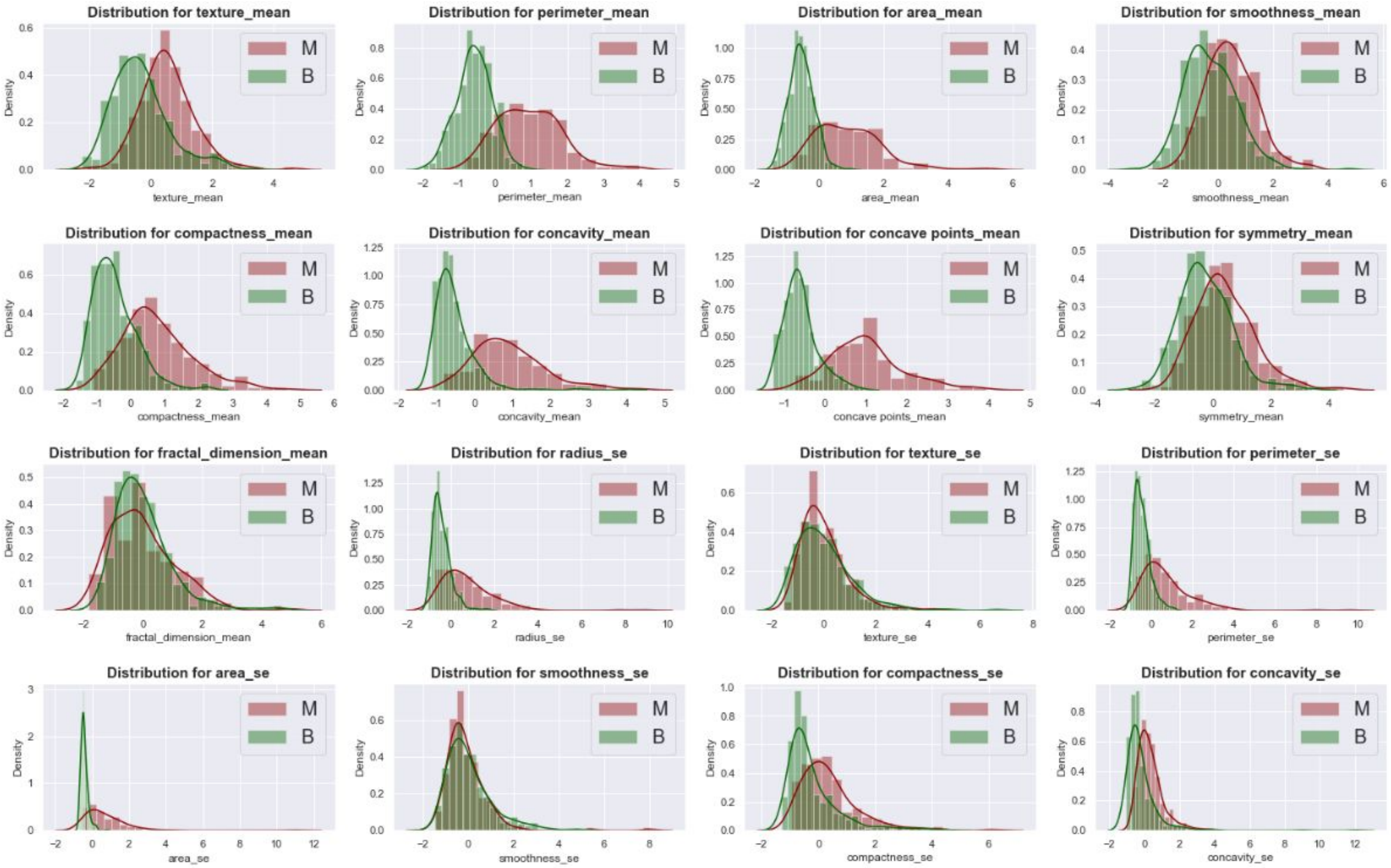
Interesting insights observed in each of the data sets:

- More positive thyroid outcomes observed among females than males
- The attributes in the cardiovascular data were mostly categorical/binary in nature
- There is no class imbalance in that dataset unlike the others

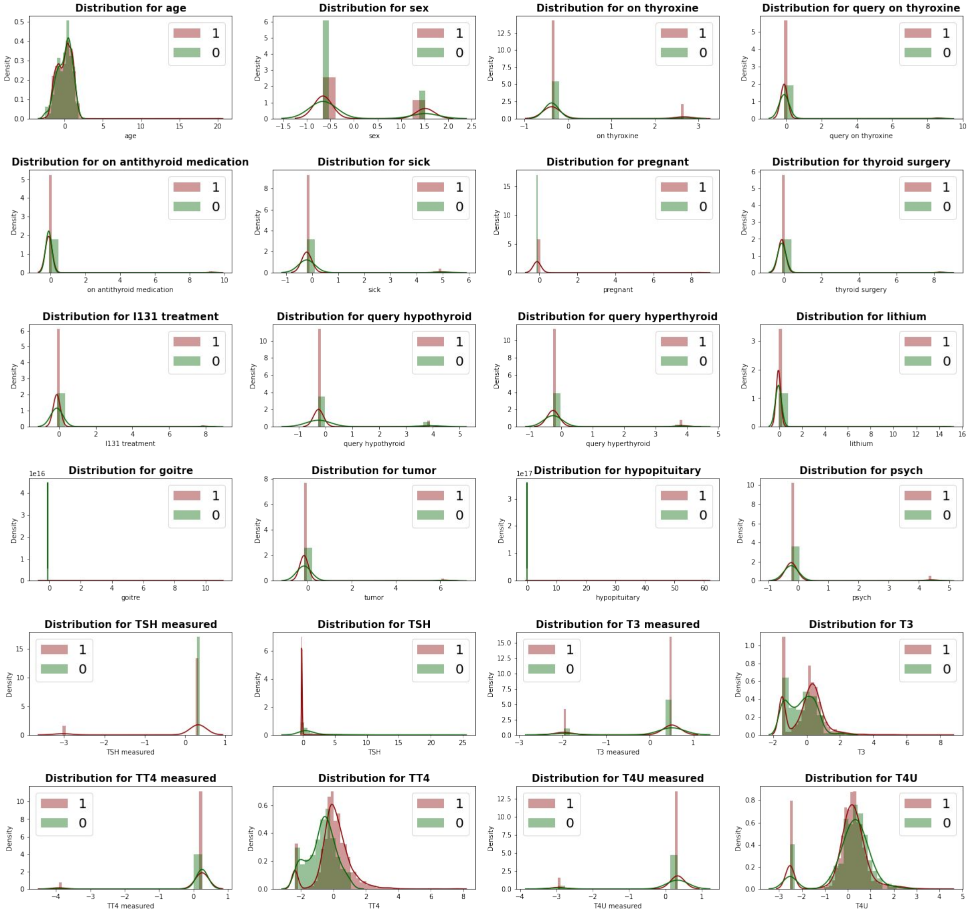


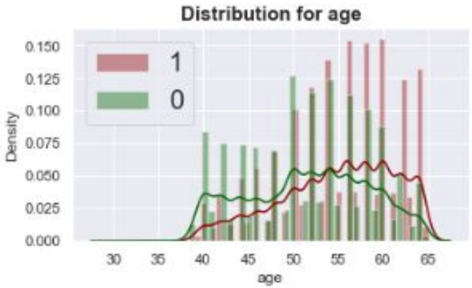
# Histogram of Attributes in Breast Cancer Dataset

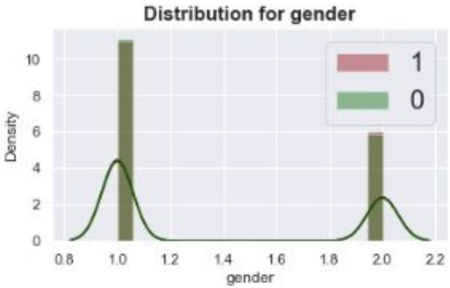


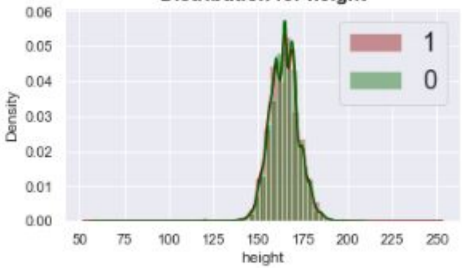
# Histogram of Attributes in Thyroid Dataset

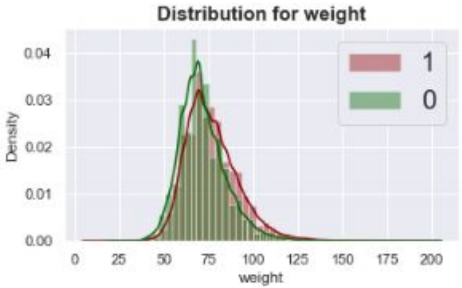


# Histogram of Attributes in Cardiovascular Disease Dataset



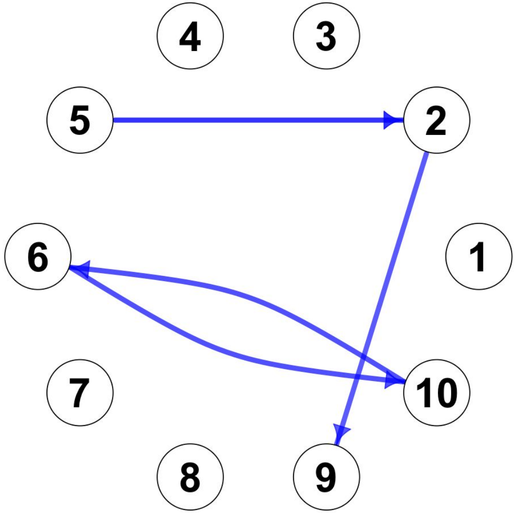



# Experiments

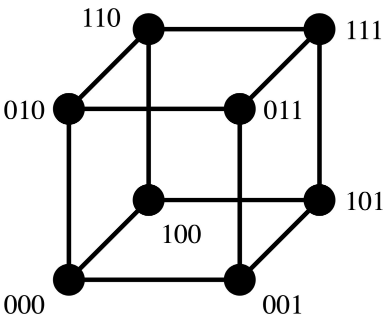
- Discovery of causal relations in medical-field datasets
- Validating causal structural models generated from various algorithms on medical data
- Understanding the relationships among the features and outcome
- Perform benchmarking using synthetic data on the various algorithms

# Synthetic data

- Data is generated in a controllable and flexible manner
- Easy to perform the evaluation of the causal algorithms
- Can misrepresent the complexity of the real-world data
- Inexpensive and easy to generate

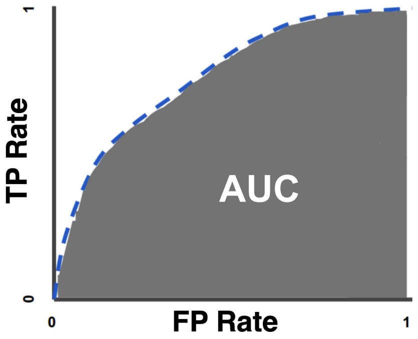


# Evaluation metrics



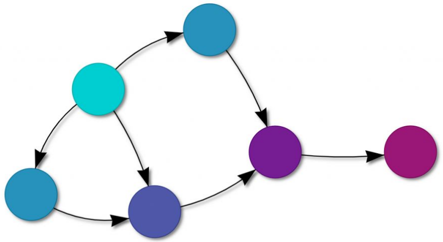

**Structural Hamming Distance**

This metric considers two directed acyclic graphs DAGs and counts how many edges do not coincide.




**Area Under ROC Curve**

Area under the curve of recall versus false positive rate (FPR) at different thresholds



**Bayesian Network Predictive Accuracy**

Models learned via causal discovery can be further used for multiple practical applications involving predictions

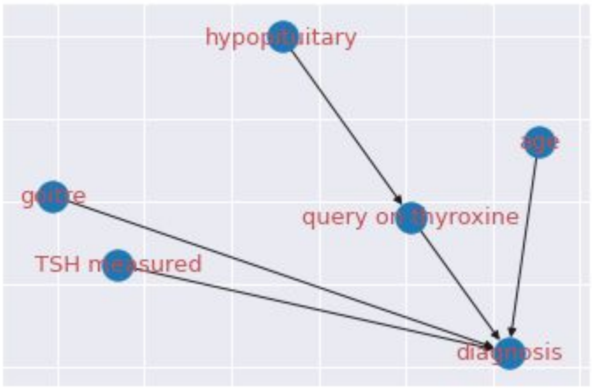
# Libraries and tools

- **Causal discovery toolbox** provides
  - A diverse pool of algorithms for causal discovery
  - Support for synthetic data generation
  - Several metrics for the evaluation of causal discovery
  - The API is entirely in Python
- **Causalnex** is a recent library that provides
  - Easy-to-use interface for causal discovery
  - Employs only the novel NOTEARS discovery algorithm

# Experiments on Medical Data

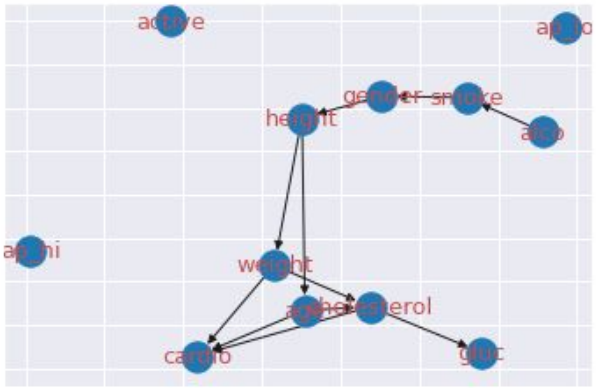
## Thyroid Dataset

- The most probable factors for **Thyroid** include: *Goitre, hypopituitary, thyroid surgery, TSH levels, pregnancy, age*
- The NOTEARS Causal Structure obtained show relationships between these factors and diagnosis



## Cardiovascular Dataset

- The most probable factors for **Cardiovascular** include: *Smoking, Alcohol Consumption, Inactivity, raised glucose levels*
- The Causal Structure is able to depict relationships for these factors

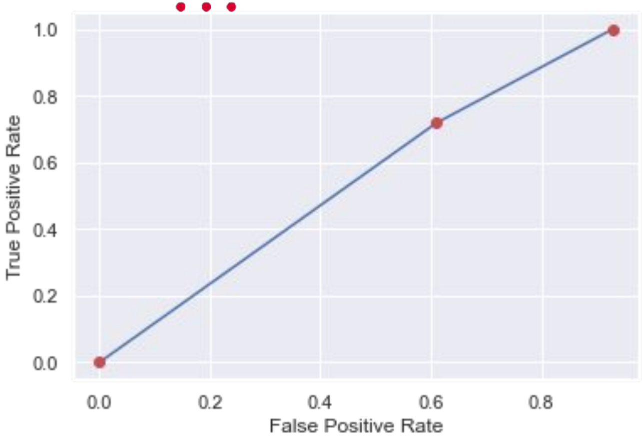


## Breast Cancer Dataset

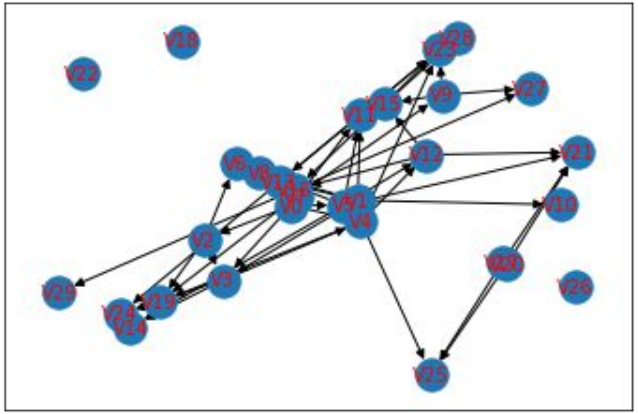
- Smaller dataset size affected causal inference
- Not many realistic relationships discovered

# Benchmark on Synthetic Data

- Causal Algorithms run on synthetic data and obtained Structure Hamming Distance and AUC
- SAM and CGNN, the deep learning based models have high hardware requirements
- This limited number of epochs for training

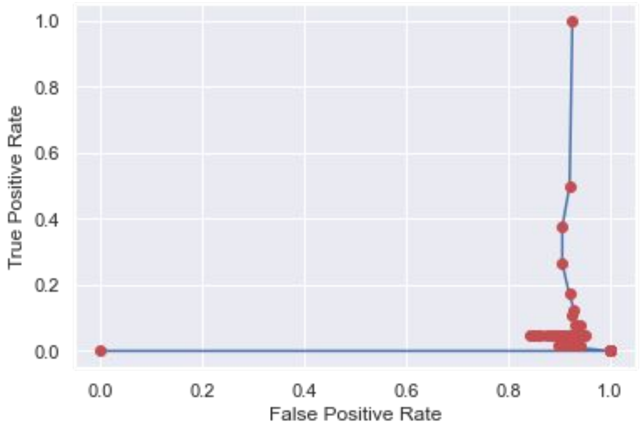


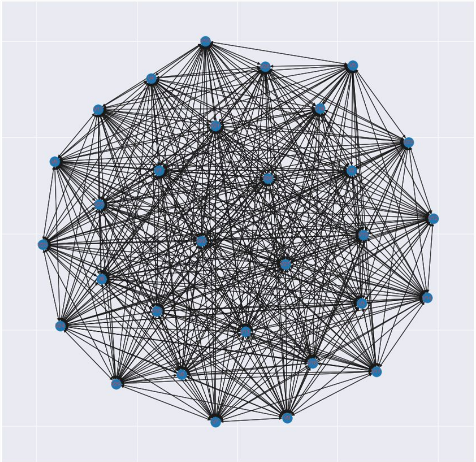

| Algorithm | SHD | AUC |
|---|---|---|
| **PC** | 75 | 0.367 |
| **GES** | 90 | 0.564 |
| **LiNGAM** | 127 | 0.289 |
| **SAM** | 806 | 0.080 |
| **NOTEARS** | 113 | 0.162 |
| **CGNN** | In progress | In progress |



# Benchmark on Synthetic Data

- Implemented the unconstrained versions of the algorithms to ensure fair experimental setup
- SAM shows the weakest performance due to model training for lesser epochs
- CGNN model was unable to converge due to stated hardware restrictions
- Edge threshold parameters needs to be explored in NOTEARS for better score



# Conclusions

From experiments using medical data, we infer and validate that:

- The Hybrid Causal Models perform better than the classical models in depicting representations

Through the benchmarking using synthetic data, we are able to show that

- GES gives the best model performance
- The model is able to draw a balance between the SHD and the AUC
- The SAM model shows potential in terms of performance with more training enough resources to compute the structure

# Challenges and Next Steps

- Limitations in hardware resources
- Difficulty in understanding causal structures due to large number of nodes and relations
- NN based models require more computation and training epochs to produce better results.
- Train/test CGNN and SAM for more epochs
- Validate results of medical causal structures, and implement constraints to reduce graph size and edges

# Thank you for your attention!

UNIVERSITY OF ILLINOIS CHICAGO

# References


- Christina Heinze-Deml, Marloes H Maathuis, and Nicolai Meinshausen. 2018. Causal structure learning. Annual Review of Statistics and Its Application 5 (2018), 371–391.
- Karamjit Singh, Garima Gupta, Vartika Tewari, and Gautam Shroff. 2018. Comparative benchmarking of causal discovery algorithms. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data. 46–56
- Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. 2000. Causation, prediction, and search. MIT press.
- Matthew J Vowels, Necati Cihan Camgoz, and Richard Bowden. 2021. D'ya like DAGs? A survey on structure learning and causal discovery. ACM Computing Surveys (CSUR) (2021).
- Yue Yu, Jie Chen, Tian Gao, and Mo Yu. 2019. DAG-GNN: DAG structure learning with graph neural networks. In International Conference on Machine Learning. PMLR, 7154–7163.
- Xun Zheng, Bryon Aragam, Pradeep Ravikumar, and Eric P. Xing. 2018. DAGs with NO TEARS: Continuous Optimization for Structure Learning.

# References


- Daniel C Castro, Ian Walker, and Ben Glocker. 2020. Causality matters in medical Imaging. Nature Communications 11, 1 (2020), 1–10.
- David Maxwell Chickering. 2002. Optimal structure identification with greedy search. Journal of machine learning research 3, Nov (2002), 507–554.
- Olivier Goudet, Diviyan Kalainathan, Philippe Caillou, Isabelle Guyon, David Lopez-Paz, and Michèle Sebag. 2017. Causal generative neural networks. arXiv preprint arXiv:1711.08936 (2017).
- Joris M Mooij, Jonas Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. 2016. Distinguishing cause from effect using observational data: methods and benchmarks. The Journal of Machine Learning Research 17, 1 (2016), 1103–1204.
- Chandramouli Rathnam, Sanghoon Lee, and Xia Jiang. 2017. An algorithm for direct causal learning of influences on patient outcomes. Artificial intelligence in medicine 75 (2017), 1–15.
- Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, Antti Kerminen, and Michael Jordan. 2006. A linear non-Gaussian acyclic model for causal discovery. Journal of Machine Learning Research 7, 10 (2006).