Causal Discovery in Alzheimer's Data

A comparative analysis of causal discovery algorithms.

COLUMBIA | COLUMBIA UNIVERSITY IRVING MEDICAL CENTER

Outline

Data Source

- ADNI Database
- ADNI1, ADNI2/GO
- Dataset summary

Existing Literature

- AD Causal Paper
  - Discrepancies in dataset
  - Addition of prior knowledge (time ordering)
  - Recreating results
    - Precision and recall metrics (as defined in the paper and more general definition)
    - FCI, FCES, SEM

Performance Analysis of other Algorithms

- gCastle Package
- PC Algorithm, LinGAM, No Tears, ANM Non-Linear, GranDAG,

New Algorithm

- GAE Paper
  - Algorithm Details
- Performance analysis
- Addition of background knowledge

Discussion

- Justification for algorithm, and causal discovery more generally

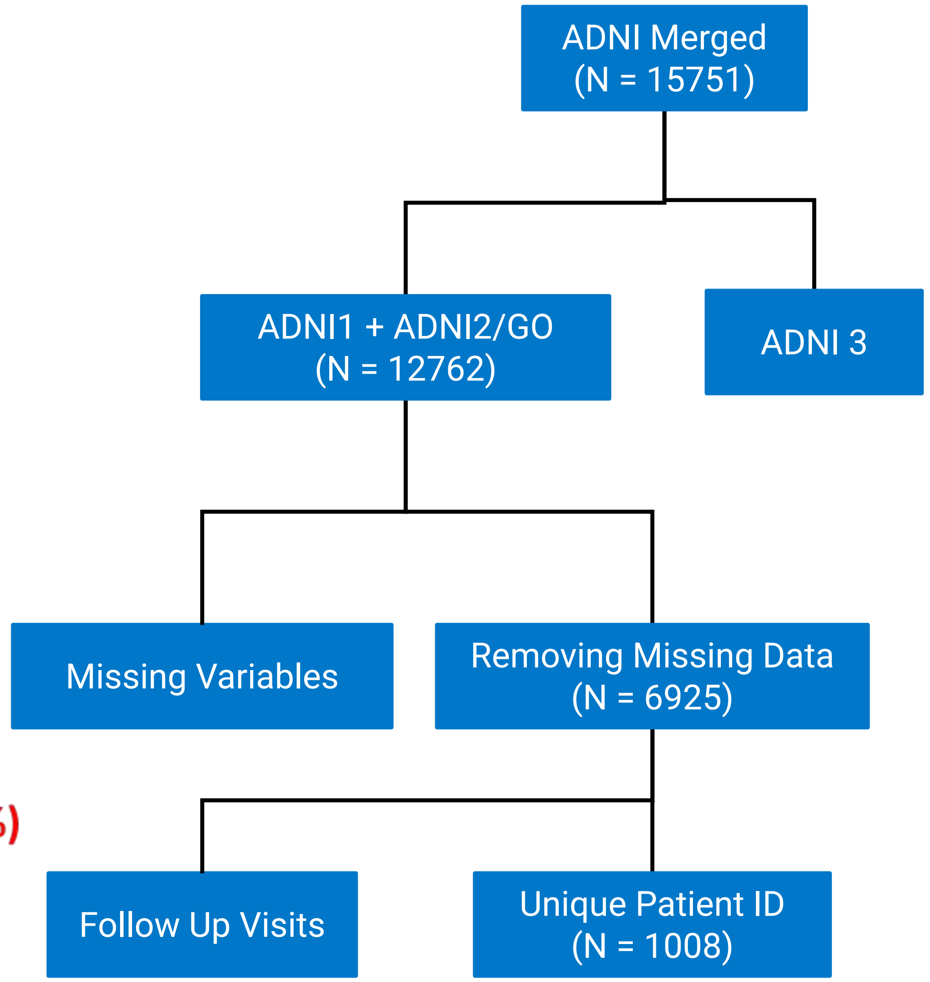
Data Source

Cohort Selection

- Deviance from AD Causal Paper
- Split APOE into 2 variables

	Label	Mean (SD)	
Demographic variables			
AGE	AGE	74.09 (7.46)	72.98 (7.25)
SEX	SEX	0.55 (0.50)	0.56 (0.50)
Education Level	EDU	16.15 (2.71)	16.13 (2.73)
Biomarkers			
Fludeoxyglucose PET	FDG	1.22 (0.17)	1.24 (0.15)
Amyloid Beta	ABETA	986.29 (459.94)	1000.53 (456.32)
Phosphorylated tau	PTAU	27.67 (14.76)	27.40 (14.61)
	Label	**subtype (%)**	
Genetics			
APOE epsilon 4 allele	APOE4	0 (54%)/ 1 (36%)/ 2 (10%)	0 (54%)/ 1 (36%)/ 2(10%)
Diagnosis			
Diagnosis of Alzheimer's Dementia	DX	CN (31%)/ MCI (46%)/ AD (23%)	CN (31%)/ 1 (51%)/ 2(18%)

Table 1. Characteristics for Continuous and Categorical Variables. N = 1008.



Data Source

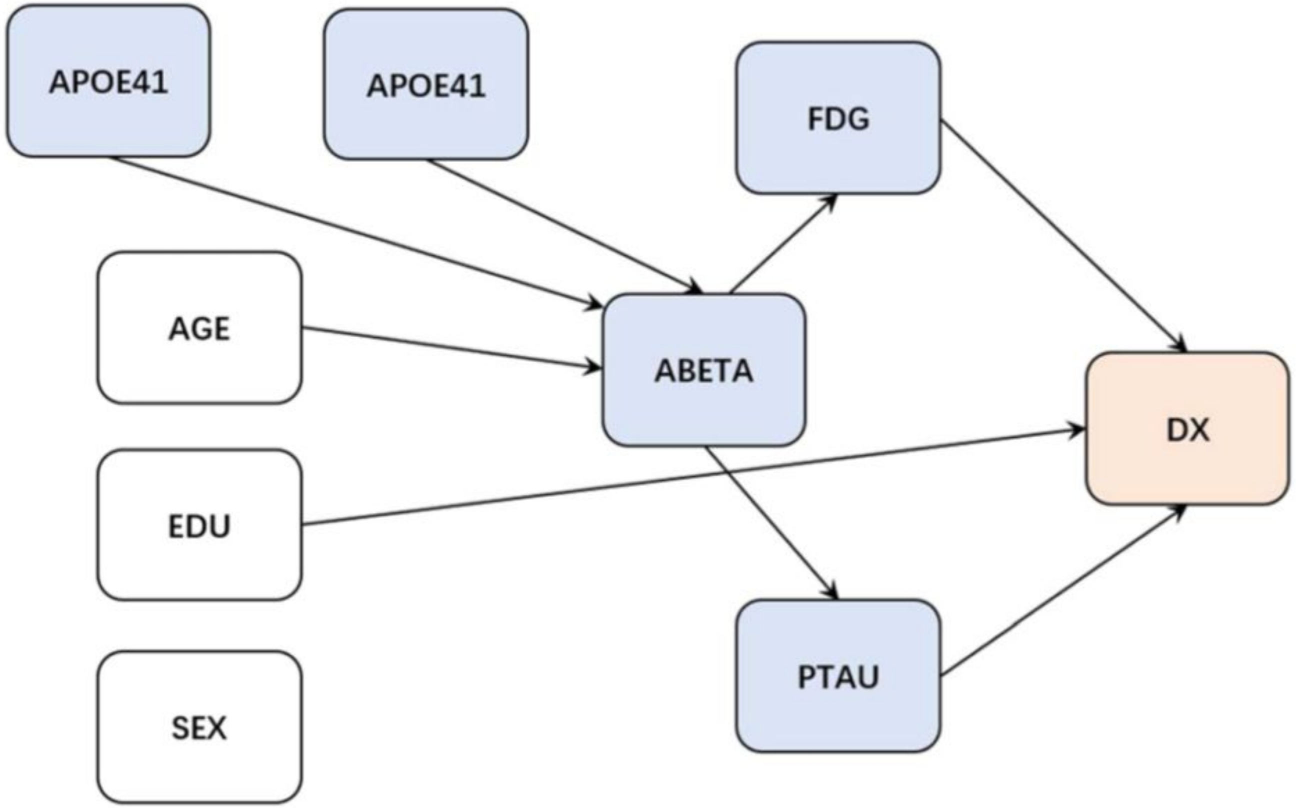
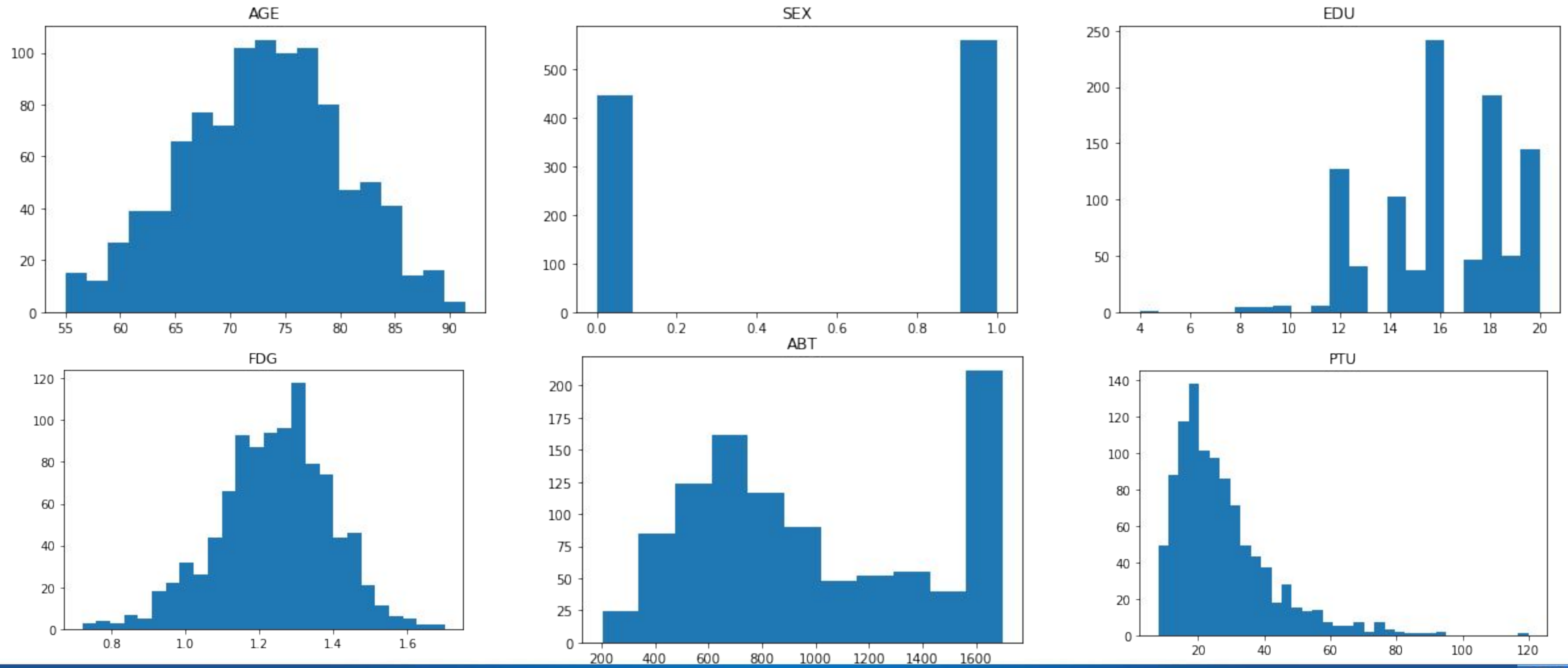


Figure 2. The "gold standard" graph.

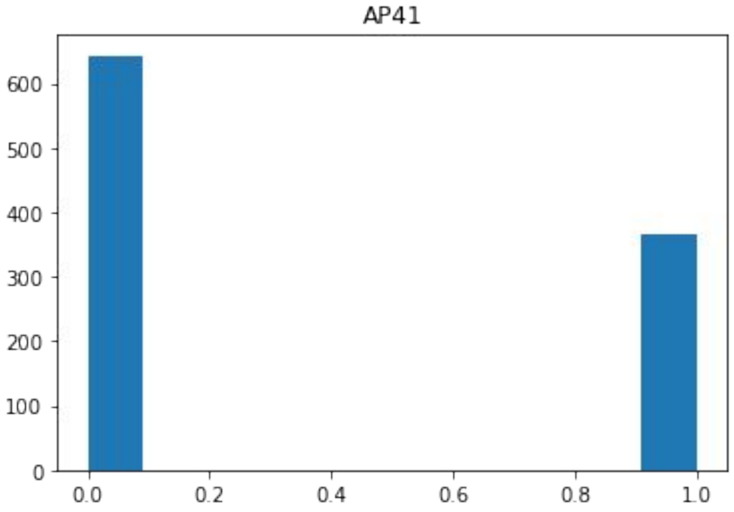
Data Source

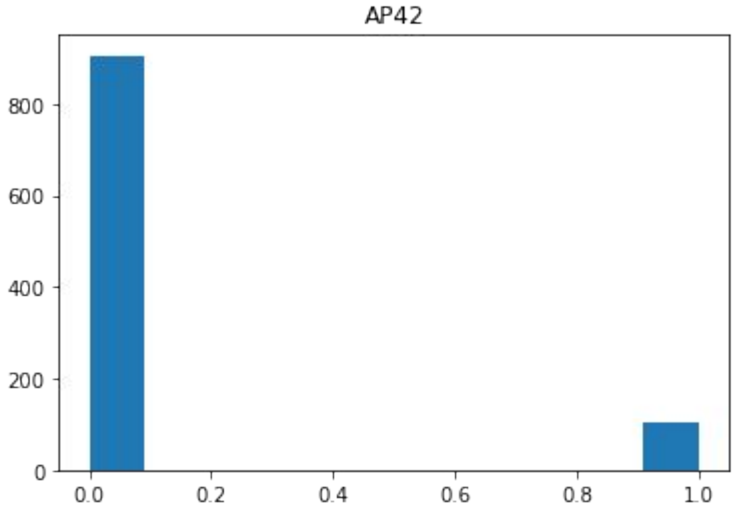
Variable Analysis

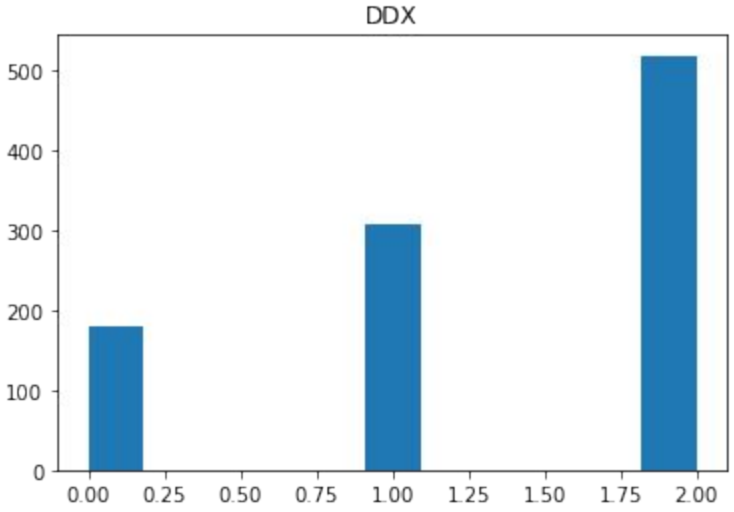


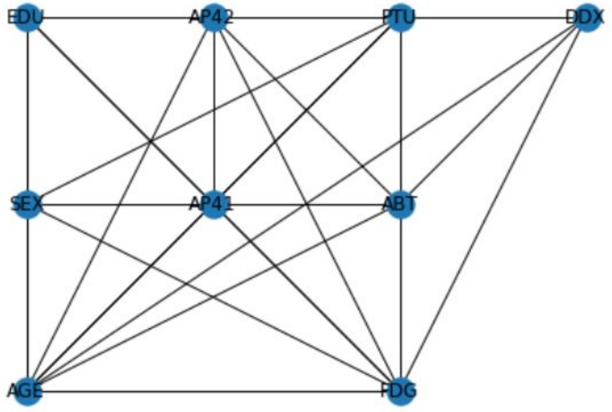
Data Source

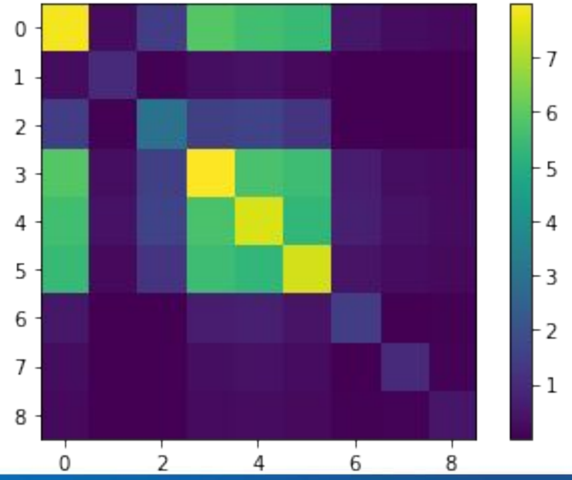
Variable Analysis







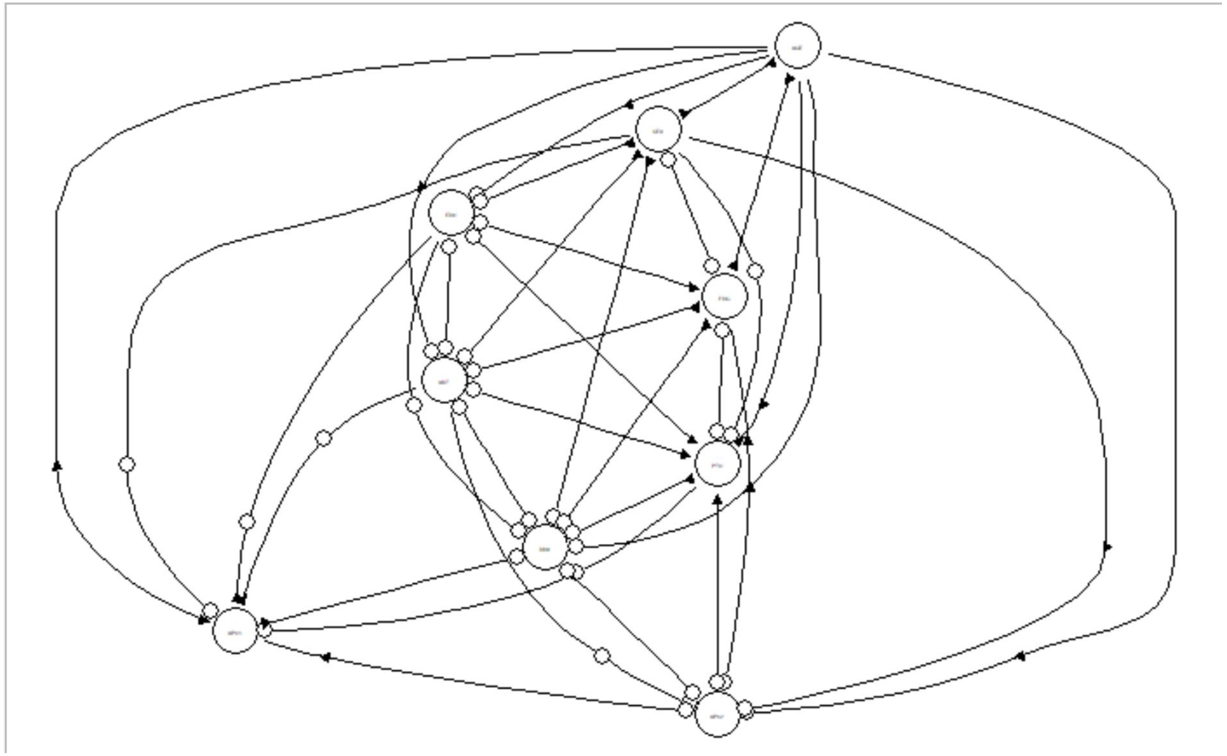




Background

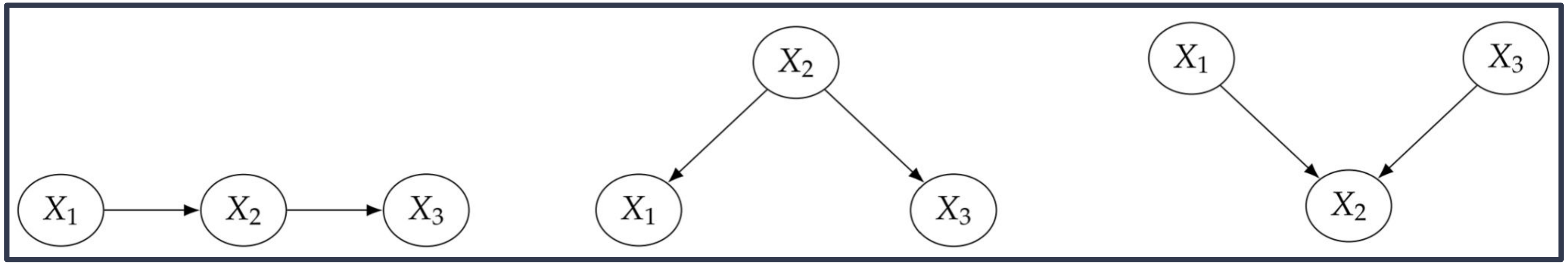
Existing Literature

- AD Causal Paper
  - Discrepancies in dataset
  - Addition of prior knowledge (time ordering)
  - Recreating results
    - Precision and recall metrics (as defined in the paper and more general definition)
    - FCI, FGES, SEM



Common Metrics

- Precision = # correct or semi-correct edges / number of edges in inferred graph
- Recall = # edges correct or semi-correct/ # number of edges in true graph



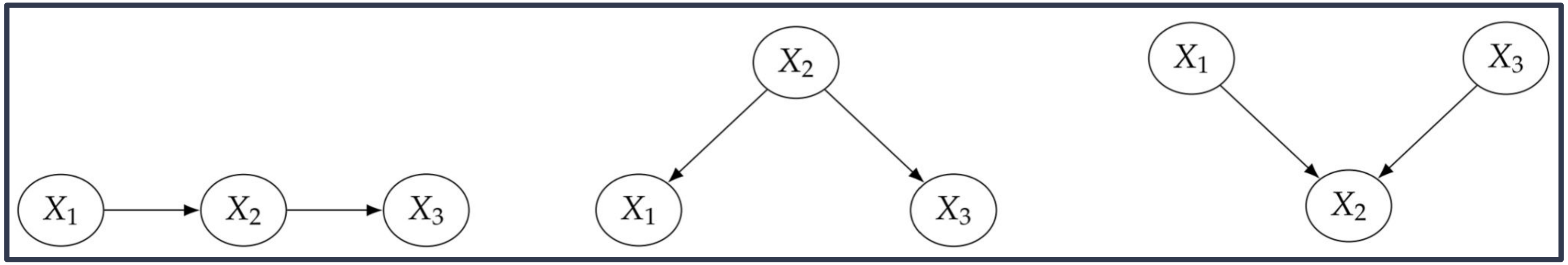
Background

Algorithm Types

- Constraint based
  - Different causal structures imply different independence and conditional independence relationships.
    - e.g. X1→X2→X3 implies X1 is independent of X3 when conditioned on X2, while X1→X2←X3 implies X1 and X3 are independent unconditionally.
  - Constraint based algorithms construct causal relationships based on repeated conditional independence testing i.e. condition over power set of all other variables
- Score based
  - General a number of possible causal structures and assign scores to each.
    - e.g. FGES uses Bayes Information Score (Penalizing extra parameters)

Properties & Assumptions

- Irreducibility
- Acyclicity
- Distinguishability



Comparison

Performance Analysis of other Algorithms

- gCastle Package
- PC Algorithm, LinGAM, No Tears, ANM Non-Linear, GranDAG

Algorithm	Precision	Recall
FCI	0.7500	0.1500
FGES		
SEM		
PC	0.3750	0.1667
LinGAM	0.3750	0.1765
No Tears	0.3750	0.1304
ANM Non-Linear	0	0
GranDAG	0.3750	0.1200
GAE		

New Algorithm

New Algorithm

- GAE Paper
  - Algorithm Details
- Performance analysis
- Addition of background knowledge
- Uses a smooth characterisation of the acyclicity constraint
- Reduces to a continuous optimisation of real matrices - “NO TEARS”
- Gradient based approach
- Can apply standard continuous optimisation under constraint algorithms!
- Apply Graph Autoencoder

$$\min_{W \in \mathbb{R}^{d \times d}} \; F(W) \quad \text{subject to} \quad \mathsf{G}(W) \in \mathsf{DAGs} \qquad \Longleftrightarrow \qquad \min_{W \in \mathbb{R}^{d \times d}} \; F(W) \quad \text{subject to} \quad h(W) = 0,$$

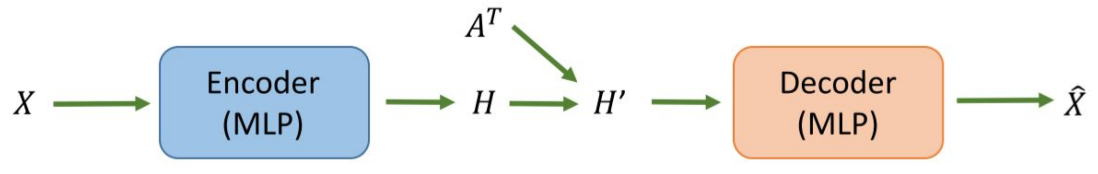


Figure 1: Model architecture for our proposed method.

Better Measures of Causality

Definition of Causality

- We quantify causality by using the notion of the causal relation introduced by Granger (Wiener 1956; Granger 1969), where a signal X is said to Granger-cause Y if the future realizations of Y can be better explained using the past information from X and Y rather than Y alone.
- Assumes strict time ordering.

Granger-Causality

Definition 4.2 X does not Granger-cause Y relative to side information Z if and only if $Y_{t+1} \perp\!\!\!\perp X^t \mid Y^t, Z^t$.

Transfer Entropy

Given a coupled system (X, Y), where $P_Y(y)$ is the pdf of the random variable Y and $P_{X,Y}$ is the joint pdf between X and Y, the joint entropy between X and Y is given by the following:

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} P_{X,Y}(x,y) \log P_{X,Y}(x,y).$$

The conditional entropy is defined by the following:

$$H(Y|X) = H(X,Y) - H(X).$$

We can interpret $H(Y|X)$ as the uncertainty of Y given a realization of X.

Better Measures of Causality

Transfer Entropy Continued

The Transfer Entropy can be defined as the difference between the conditional entropies:

$$TE(X \to Y|\, Z) = H\left(Y^F | Y^P, Z^P\right) - H\left(Y^F | X^P, Y^P, Z^P\right), \tag{4.3}$$

which can be rewritten as a sum of Shannon entropies:

$$TE(X \to Y) = H\left(Y^P, X^P\right) - H\left(Y^F, Y^P, X^P\right) + H\left(Y^F, Y^P\right) - H\left(Y^P\right),$$

where Y^F is a forward time-shifted version of Y at lag Δt relatively to the past time-series X^P, Y^P and Z^P. Within this framework we say that X does not G-cause Y relative to side information Z if and only if $H\left(Y^F | Y^P, Z^P\right) = H\left(Y^F | X^P, Y^P, Z^P\right)$, i.e., when $TE\left(X \to Y, Z^P\right) = 0$.

G-Causality and TE Relationship:

- If all variables are jointly Gaussian, TE is equivalent to G-Causality up to a multiplicative factor.

Better Measures of Causality

Jiheum's Simulated Data

- Gaussian synthetic data with same dimensions as ADNI Data and gold standard causal structure.
- N = 20000

Issues

- Differences between G-Causal and Transfer Entropy

G-Causality:

	AGE	SEX	EDU	APOE4	ABETA	FDG	PTAU	DX
AGE	0.000000	-0.000027	-0.000034	-0.000042	-0.000035	-4.908477e-05	-0.000041	-0.000017
SEX	0.000002	0.000000	-0.000049	0.000139	-0.000031	-4.959363e-05	0.000012	-0.000042
EDU	-0.000047	-0.000045	0.000000	0.000049	-0.000048	-4.833645e-05	-0.000046	-0.000023
APOE4	-0.000047	0.000030	-0.000022	0.000000	-0.000028	-5.000652e-05	-0.000049	-0.000050
ABETA	-0.000044	-0.000039	-0.000045	-0.000048	0.000000	-4.150573e-05	-0.000014	0.000025
FDG	-0.000012	-0.000049	-0.000039	-0.000020	0.000024	0.000000e+00	-0.000017	0.000075
PTAU	-0.000050	0.000028	-0.000012	-0.000033	-0.000012	-4.506872e-05	0.000000	-0.000049
DX	-0.000015	-0.000033	-0.000022	0.000007	0.000085	5.758210e-08	-0.000050	0.000000

Transfer Entropy:

	AGE	SEX	EDU	APOE4	ABETA	FDG	PTAU	DX
AGE	0.000000	0.000494	0.000474	0.000090	0.000242	0.000261	0.000181	0.000128
SEX	0.000622	0.000000	0.000908	0.000142	0.000227	0.000228	0.000523	0.000835
EDU	0.000253	0.000526	0.000000	0.000563	0.000211	0.000427	0.000740	0.000367
APOE4	0.000185	0.000135	0.000384	0.000000	0.000350	0.000303	0.000125	0.000292
ABETA	0.000342	0.000468	0.000952	0.000147	0.000000	0.000171	0.000117	0.000335
FDG	0.000210	0.000420	0.000284	0.000178	0.000075	0.000000	0.000115	0.000365
PTAU	0.000638	0.000422	0.000652	0.000202	0.000214	0.000288	0.000000	0.000513
DX	0.000250	0.000506	0.000622	0.000163	0.000414	0.000495	0.000307	0.000000

Next

Discussion

- Justification for algorithm, and causal discovery more generally?
- Compare over variety of synthetic data for performance?
- What to do with our ADNI Dataset?
- Explore limitations of Time Ordering in TE & G-Causality