A Causal AI-Based Framework For Reliable And Optimal Decision Making In Healthcare Context

Uzma Hasan, Muhammad Hasan Ferdous, Sancharee Hom Chowdhury

December 2021

Abstract

With the increasing application of AI in healthcare, there is an increased concern regarding the predictive behaviour of these algorithms. No matter how sophisticated the predictive algorithms are, their users can mistakenly consider correlation as causation. Particularly for the healthcare sector, where decisions can highly impact human lives, relying solely on predictive AI models can be cause devastating consequences when correlations are mistaken for causation. On the other hand, AI systems that can explain the true causes behind a phenomenon are reliable for making informed decisions and achieving outcomes that can be beneficial for researchers, physicians, and patients The main objective of this research is to develop a Causal AI based healthcare framework that can overcome the existing limitations of the traditional predictive AI approaches by facilitating causal discovery from observational data (e.g EHRs) and perform experiments to identify the causal effects of any interventions.
Keywords: Causal AI, Causality in Healthcare, Correlation vs. Causation, Observational Data, Black-box systems.

1 Introduction

The application of Artificial Intelligence (AI) has been outstanding in reshaping modern healthcare by developing systems that can benefit both patients and health care professionals in decision making, medical monitoring and continuous health maintenance. Though AI has its widespread applications in digital healthcare, unfortunately, most of these predictive AI systems prone to biases are black box systems lacking explainability which breaches the moral responsibilities of clinicians. Therefore, the necessity of developing AI systems that can go beyond mere prediction and identify the causal relationships has become a major concern in the fields such as machine learning, clinical research, medical diagnostics, and precision medicine. If we do not know the root causes of a certain predicted behavior, there is a high chance of making poor decisions and supporting ineffective policies. No matter how sophisticated the predictive algorithms are, their users can mistakenly consider correlation as causation [1]. Relying solely on predictive AI models in the health care sector can risk devastating consequences when correlations are mistaken for causation.

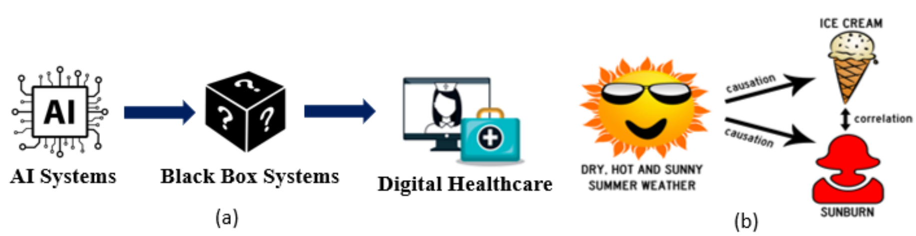


Figure 1: Application of black-box AI systems in healthcare (a) & Dilemma of correlation & causation (b) [1]

On the contrary, AI systems that can explain the true causes behind a phenomenon are reliable for making informed decisions and achieving outcomes that can be beneficial for researchers, physicians, and patients. These systems can provide a complete picture by modeling the true causal relationships which increases reliability in the decision making process. A closer look at causal AI will show how it can open up the black box within which purely predictive models of AI operate. It can answer critical medical questions like "How will this drug fare in a trial – and why?", "What combination of treatments does this patient need now?", "Which individuals would most benefit from this drug?", "What's the right care to administer to specific patients to reduce hospital stays?, etc. If an AI based healthcare system addresses these concerns, it can not only accelerate the decision making process, but also overcome the hurdles and risks faced while using generic predictive models for healthcare purposes.

The gold standard for studying causal relationships is to perform randomized control trials (RCTs) since randomization eliminates much of the bias present in other study designs [13]. But RCTs are often infeasible due to their cost ineffectiveness, ethical concerns, time consuming nature or impracticality [3]. So far we have performed experiments for our first scenario. It is a clinical dataset depicting the scenario of the factors causing heart disease with a total of 12 variables. In this case, RCT is not a feasible option as it would require observing a group of heart disease patients over a long period of time and see what factors cause CVDs and how controlling any factor could affect a patient. Such a randomized experiment is quite risky and time consuming. Hence we used benchmark causal discovery approaches for finding the causal structure for the heart disease dataset.

The aim of our research is to develop a Causal AI based healthcare framework that describes the causal story behind a medical problem by performing causal discovery from observational data and discovering the findings of what-if questions using suitable causal effect estimation techniques. To evaluate our proposed approach, we aim to consider to two real-life scenarios: 1) to identify which factors are responsible for causing cardiovascular diseases (CVDs) commonly known as heart disease and 2) to detect the causal factors that contribute to the risk of having a lung cancer, which is considered as a fatal disease.

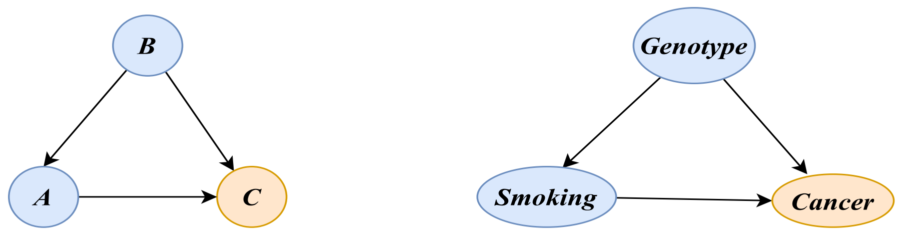
2 Background and Related Works

In the late 1950s and 1960s, there were a debate regarding the causing of the lung cancer. It was assumed that smoking was the vital reason of different types of cancer. [15] Several findings characterize the tobacco smoke increases the mortality rate. It was high time to take action to reduce the smoking. However, tobacco industry had created doubt in the evidence even though there were mounted epidemiological evidences. A tension was arised due to the doubt and in the 1960s casual inference was introduced to solve the debate. [11] This framework found that smoking was the potent cause of cancers.

2.1 Causality

Causality refers to the influence that one event, process or state (a cause) has on the production of another event, process, or state (an effect), where the cause is partly responsible for the effect and the effect is partly dependent on the cause. [2] Causes can be divided into necessary and sufficient causes.[7] If x is a necessary cause of y, then the presence of y presupposes the presence of x. The presence of x, on the other hand, does not guarantee the occurrence of y.[32] If x is a sufficient cause of y, then the presence of x presupposes the occurrence of y. However, another reason z might also be the cause of y. As a result, the presence of y does not presuppose the presence of x.[32]

The data-generating process of the distribution is shown by a causal graph or causal structure in the form of a Directed Acyclic Graph (DAG).[30] A DAG is a graph that flows in one direction, where no element can be a child of itself. Figure 1 (a) represents a simple DAG with three variables A, B, and C whare cause



(a) (b)

Figure 2: Typical DAG (a) and a practical DAG of smoking and cancer (b)

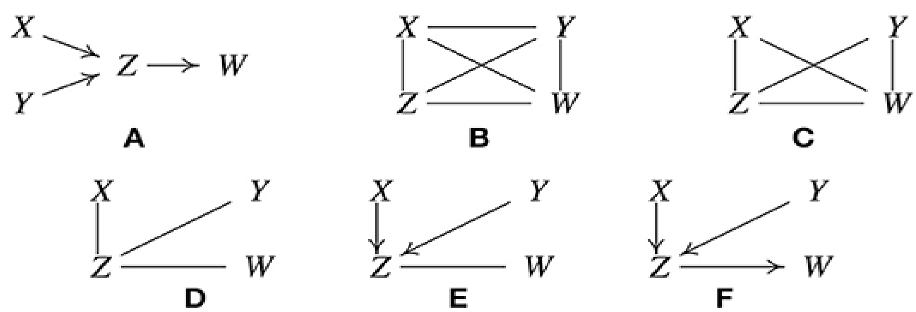


Figure 3: Mechanism of PC algorithm

to effect relationships are represented by an arrow. In figure 1 (a), A is a cause of C, B is the cause of both A and C.

2.2 Casual Discovery

Causal discovery is a prerequisite for estimating causal effect because it provides the data generating process under the study. Once we have the causal structure of the study variables, we can apply Pearl's do calculus [24] to estimate the causal effect. Causal discovery algorithms divided into three groups based on the differences between the methodologies: constraint-based methods, score-based approaches, and hybrid techniques.

To estimate the causal dependency between the variables, constraint-based techniques apply the Causal Markov condition and a set of conditional independence (CI) tests. One of the oldest and most prominent of such causal discovery methods is Peter-Clerk algorithm [29] PC algorithm starts with complete undirected graph (Figure 2A). Then it eliminates edges that are unconditionally independent. In figure 2C, the edge between X and Y is removed because of this. Then for each pair of connected variables with a common connected edge between then, PC algorithm tests conditional independence between the variables given the common variable and removed the edge if the test indicates conditional independence. In Figure 2D, the edge X and W is removed because they are independent given common variable Z.It continues to increase the number of conditional variables until all of the variables in the subset are adjacent to the connected variables. In the edge orientation phase, it also incorporates conditional independence tests among the triples. Figure 2 depicts the PC algorithm for four variables.One of the limitations of PC algorithm is that it can not perform in the presence of unobserved confounders. But inclusion of all variables is unpractical and difficult in case of observed variables. Fast Causal Inference (FCI) [29] overcomes this limitation by finding unshielded colliders and performing conditional tests on "possible D-separation set". However, this method is computationally less effective due to inclusion of conditional Independence test on all possible subsets of the separation set. Some other methods based on FCI, such as 'Anytime FCI'[28], RFCI [6], FCI+ [5] are proposed to overcome the computational limitations of the FCI algorithm.

Constraint-based methods require a large sample size to draw a valid conclusion about the causal structure because of their CI tests. Score-based approaches overcome this limitation by defining a score function S(D,G) from the given data $D = (X^1, X^2, \ldots, X^n)$ to validate the candidate graph G. The causal structure with the best score is selected as the outcome. Mathematically, the goal is defined as:

$$\widehat{G} := argmax_{Goverx} S(D, G) \tag{1}$$

Here, score function S is can be defined using different criteria such as BIC [10], Minimum Description Length [18], Bayesian Gaussian equivalent score [10] etc. Some widely used score-based methods for identifying causal structure are Greedy Equivalence Search (GES) algorithm [4], Adaptive Lasso [26], GIES [14]

Hybrid methods combines constraint and score-based methods to identify the underlying causal structure. For example, Max-Min Hill-Climbing (MMHC) algorithm [31] uses constraint-based conditional independence tests to get the Parents and Children sets of each node.Then in the edge orientation phase, it uses score-based greedy Bayesian-scoring hill-climbing search. Greedy Fast Causal Inference (GFCI) [12] is another hybrid method where it uses GES algorithm to find the skeleton and FCI algorithm to orient the edges.

Functional causal models express relations among the variables in a functional form. Linear Non-Gaussian Acyclic Model (LINGAM) algorithm [25] can detect the complete causal structure of continuous data of the data generating process is linear, no unobserved confounders in the model and the error variables follow non-Gaussian distributions with non-zero variances. A similar approach named TEARS [33] used optimization of loss function for learning the causal structures. VAR-LiNGAM [17], DYNOTEARS [23], SVAR-FCI [21] are the time series implementation of LiNGAM, TEARS, and FCI respectively.

2.3 Identification of Causal Effect

Identifiability of causal effect refers to the ability to reliably estimate the causal effect of a treatment on outcomes using a sample drawn from a known sample distribution. "The causal effect of X on Y is said to be identifiable if the quantity $P(y \mid \hat{x})$ can e computed uniquely from any positive distribution of the observed variables, that is, if for every pair of theories T_1 and T_2 such that $P_{T_1}(v) = P_{T_2}(v) > 0$, we have $P_{T_1}(y \mid \hat{x}) = P_{T_2}(y \mid \hat{x})''$ [8]

DEFINITION (Back-door criterion): Given an ordered pair of treatment and outcome variables (X, Y), a set of variables Z is said to satisfy the back-door criterion in a DAG G if-

1. All nodes of Z are ascendants of X and
2. Z blocks all paths between X and Y which contains an edge into X

THEOREM (Back-door Criterion). [8] For estimating the causal effect of X on Y, if there exists a set of variables Z that satisfies the back-door criterion for (X, Y) and p(x,z)¿0 in the DAG G, then the causal effect is identifiable by the following formula:

$$P\left(y \middle| \hat{x}\right) = \sum_z P\left(y \middle| x, z\right) P\left(z\right) \tag{2}$$

Though widely used, back-door adjustment depends on some strong assumptions known as the Markovian assumptions. In the presence of unobserved confounders, using back-door adjustment will produce a biased causal estimate containing confounding bias.

DEFINITION (Front-Door). A set of variables M is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if:

1. M intercepts all directed paths from X to Y;
2. all back-door paths from X to M are blocked (by empty set); and
3. all back-door paths from M to Y are blocked by X.

THEOREM (Front-Door Criterion). [8] If M satisfies the front-door criterion relative to an ordered pair of variables (X, Y), then the causal effect of X on Y is identifiable and is given by the formula

$$P(y \mid \hat{x}) = \sum_{m} P(m \mid x) \sum_{x'} P\left(y \mid x', m\right) P\left(x'\right) \tag{3}$$

Pearl also proposed a set of inference rules that can be used to transform the causal relations into a probabilistic form enabling us to estimate the causal effect using observational data. These rules are known as do-Calculus. [8] Pearl initially showed that the do-Calculus is sufficient for identifying causal effects. Later other researchers showed that the do-calculus is complete to identify causal effects of $P_x(y \mid z)$ [16] [27]

2.4 Causality in Healthcare

In the American Journal of Public Health, Herna´n and other colleagues strongly urge to the importance of using the causal thinking. Using the concept of counterfactual, causality actually try to identify the effect of the intervention which is the key component of this framework. [22]

Randomized control trials are typically used to ascertain the causal relationship. This trial is suitable for certain circumstances, animal studies for instance. Due to some constraints, for example ethical, logical and cost issues experimental studies may not always the appropriate choice. This practical circumscriptions help to shed light on the importance of identifying causal relationship of the observational data like healthcare data.

State of the art methods were traditionally used for prediction and classification rather than causality. However, this situation is still evolving. Now a days, practitioners are not only interested with mere prediction but also want to know the cause-effect relation between input features and clinical outcome. This understanding will help doctors to treat patient more effectively and also helps us to apply better management. [34]

3 Methodology

This section describes the steps followed in our research. The Figure 4 depicts the step by step process involved to develop our proposed approach. All the steps are discussed below in details.

Step 1: In this phase, the data for the task is collected at first. Some of the sources of medical data include Electronic Health Records (EHRs), healthcare research, etc. Then the collected data is cleaned before feeding it to the algorithms. Data cleaning refers to replacing qualitative data with numerical values, filling the missing values with the mean value, normalizing the data, etc.

Step 2: A set of causal discovery algorithms are utilized to discover a causal graph of the problem. Any convenient causal discovery methods can be used as per the domain requirements. Hence, this part of our framework is modular. To find the underlying causal graph, we used three algorithms. First, is a constraint based method, the PC algorithm [29] that performs conditional independence tests to check dependency between the variables. Second, is a score-based method, the Greedy Equivalence Search (GES) [4] algorithm which finds causal structures by greedily adding or deleting edges until reaching a local maximum. Third, the LINGAM [25] method which is based on the non-Gaussianity of the data and considers a linear data-generating process with no unobserved confounders.

Step 3: After analysing the data using a set of causal discovery algorithms, it is now the turn to select the best scored or most accurate causal graph. For which some benchmark evaluation metrics can be used. We used the following metrics: 1) Accuracy, 2) False Discovery Rate (FDR) and 3) Structural Hamming Distance (SHD). In many cases, the ground truth graph might not be available. Then the prior knowledge and expert's opinion can be useful to formulate the final graph. Often some prior knowledge such as literature evidence, domain knowledge, tiers of evidence are available in every domain. This knowledge alongside the opinion of the experts' in the field can be used to make a decision. Thus, based on the performance metrics or combination of prior knowledge with experts' opinion, the final causal graph is chosen.

Step 4: A number of queries can be performed upon the causal graph which depicts the data generating process of the problem. This process is called causal effect estimation. It tries to estimate how much effect one variable has upon the other.

Step 5: The obtained results after causal discovery and causal effect estimation are useful in informed decision making improve policy effectiveness. Particularly such results are useful in the healthcare sector which can guide in taking prompt medical decisions with greater clarity than those involving solely predictive models.

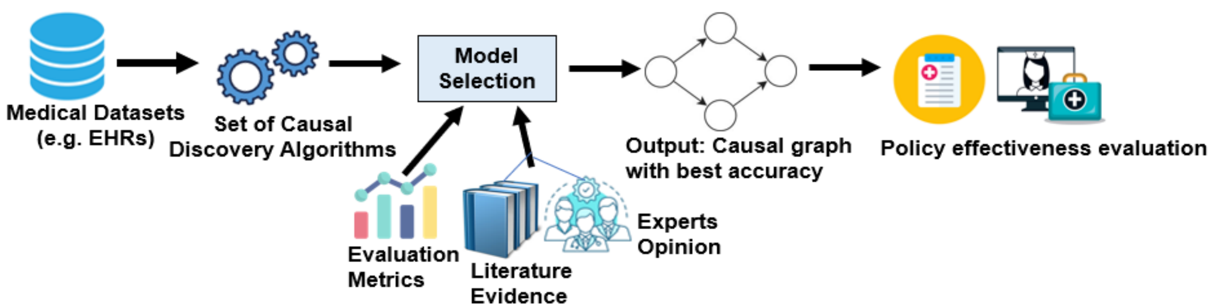


Figure 4: Conceptual framework of the proposed approach

4 Experimental Results

In this section, we have reported the generated causal graphs by all the 3 algorithms. We have also discussed the selection of the final causal graph and the evaluation process of the discovered graphs. Finally we reported the causal effects of the variables upon the target variable and reported some valuable conclusions from the obtained results.

4.1 Synthetic Dataset

For experimentation, we used the LUCAS (Lung Cancer Simple) [19] dataset which contains toy data generated artificially by causal Bayesian networks with binary variables. This dataset contains a total of 12 vertices and 12 edges. It has been developed for modeling a medical application for the diagnosis, prevention, and cure of lung cancer. Here the data generative model is a markov process. Figure 5 shows the ground truth graph of this dataset with the target variable *Lung Cancer* shaded in purple and the nodes in dark green are members of the Markov blanket of the target variable. The Figure 6 (left), Figure 6 (right) and Figure 7 are the graphs discovered by the GES, LINGAM and PC algorithms respectively for the LUCAS dataset.

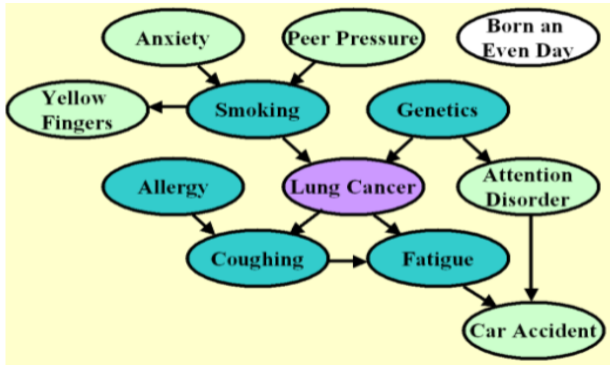


Figure 5: Ground Truth Graph Of Lucas Dataset [20]

From the Table 1, we can see that GES has the highest accuracy, the lowest false discovery rate and the lowest structural hamming distance. This proves that GES performs the best upon this dataset. While PC also has a good values of all the metrics. But, certainly LINGAM performs very poorly compared to the other two methods. This may be due to the fact that GES and PC have many assumptions in common whereas LINGAM follows a very different approach to searching the graph. Considering all the metrics, as GES outperforms the other two approaches, we select the causal graph produced by GES as the final causal graph or the output of our framework.

Table 1: Result analysis of the applied causal discovery methods on LUCAS dataset

Method	**Accuracy**	**FDR**	**SHD**
GES	0.92	0	1
PC	0.8	0.09	3
LINGAM	0.29	0.53	10

Table 2 represents a comparison of the correlation vs the causal effect of the variables upon the target variable *Lung Cancer*. From the ground truth graph we can see that *Smoking* and *Genetics* have a causal edge towards *Lung Cancer*. That is both of these are the causes of Lung Cancer. But, if we look at the correlation of the variables, it can be seen that *Coughing* has a very high correlation and *Fatigue* has a medium correlation with *Lung Cancer*. Although none of these variables have any causal effect upon *Lung Cancer*. These are actually the effects of *Lung Cancer*. So, if one tries to formulate a conclusion based upon the correlation of the variables solely, then the conclusion made can be erroneous. Before making any decision, hence it is required to observe the causal effects between the variables.

Table 2: Result analysis of the applied causal discovery methods on LUCAS dataset

Variable Pair	**Correlation**	**Causal Effect**
Smoking - Lung Cancer	0.49	0.51
Yellow Fingers - Lung Cancer	0.38	0.42
Genetics - Lung Cancer	0.23	0.29
Fatigue - Lung Cancer	0.37	N/A
Coughing - Lung Cancer	0.52	N/A

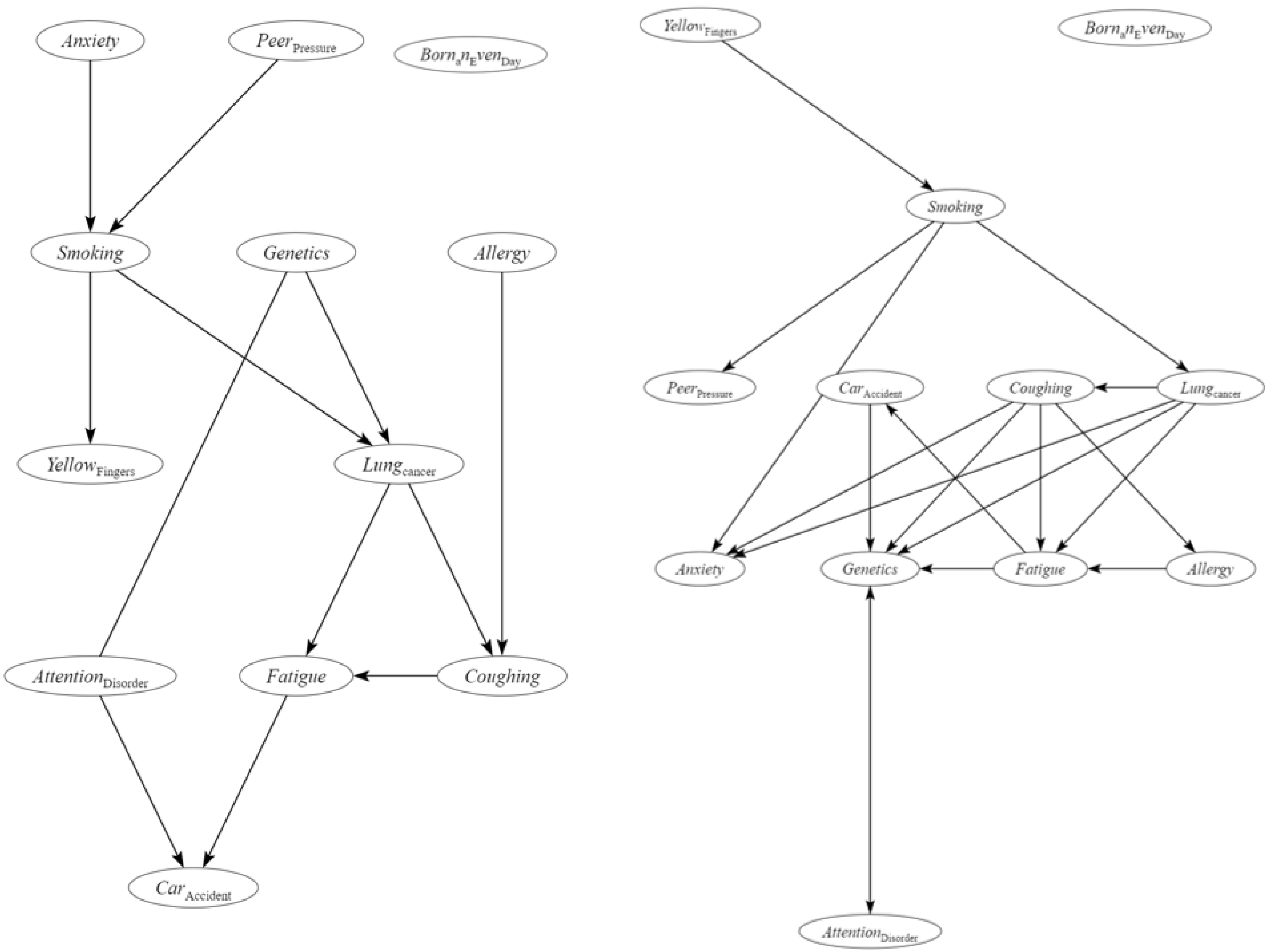


Figure 6: Causal Graph discovered by GES (left) and LINGAM (right) algorithm for LUCAS Dataset

4.2 Real Dataset

As a Real dataset, we experimented on the Heart Disease Dataset from Kaggle (link:`https://www.kaggle.com/sid321axn/heart-statlog-cleveland-hungary-final`) which is the largest heart disease dataset available for research purposes. It contains a total of 11 features with a total of 12 nodes including the target variable. In case of this dataset, we have reported the causal graphs found by the three algorithms GES (Figure), PC and LINGAM (Figure). For this dataset, we did not have any ground truth graph. In cases where the ground truth graph is not available, prior knowledge and expert's opinion can be useful to formulate the final graph. It is often seen that some evidence is always available for any problem domain which can be utilized for causal discovery. But, due to lack of available prior knowledge or any expert's opinion, we would like to formulate the final causal graph combining the results from the GES and PC algorithm. Because, from the discovered graphs by all the three methods, it seems that GES and PC agrees with each

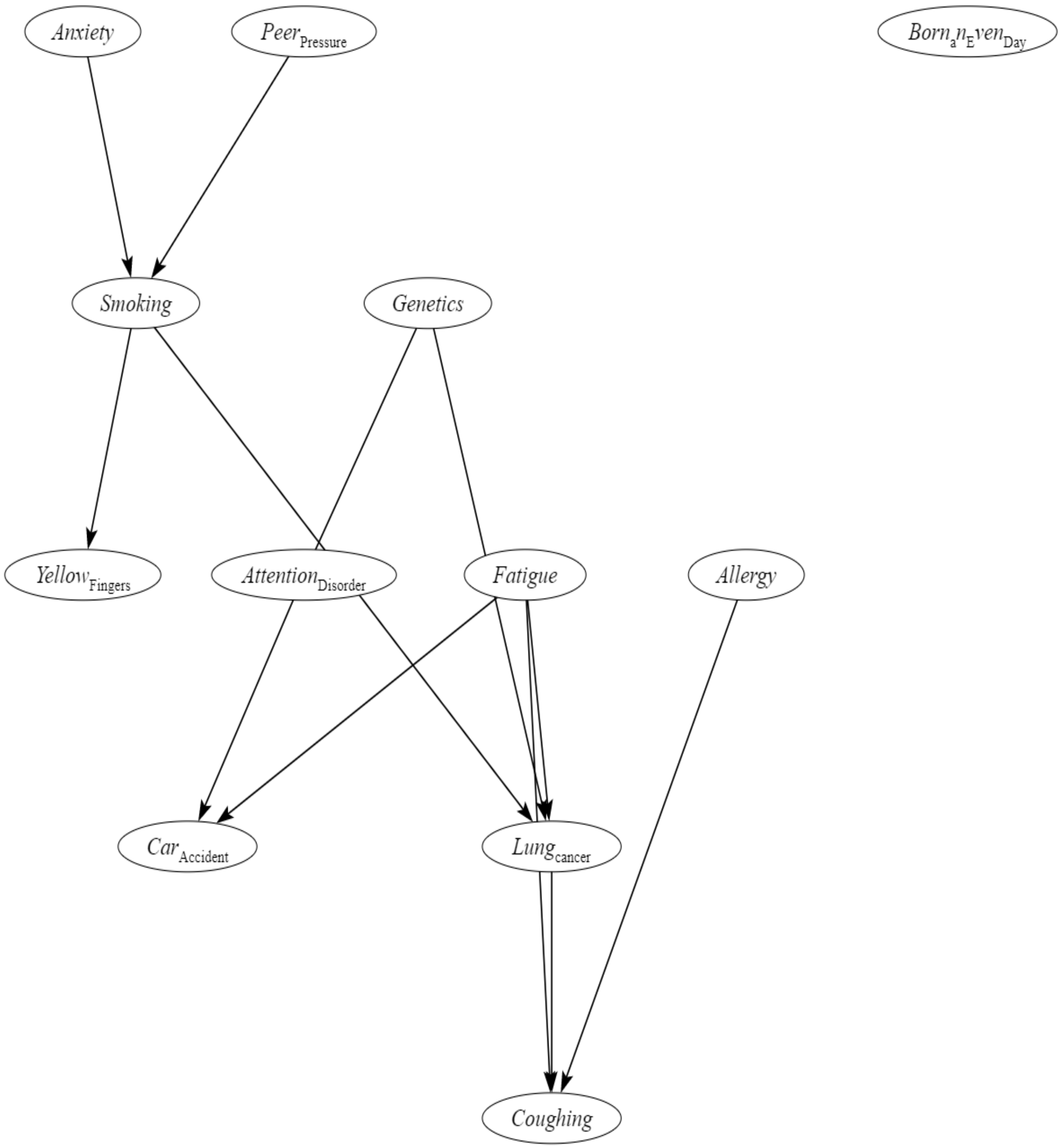


Figure 7: Causal Graph discovered by PC algorithm for LUCAS Dataset

other in case of most of the edges. Moreover, PC and GES algorithms have similar type of assumptions. While, LINGAM follows a very different approach to discover the graph and the edge orientation discovered by LINGAM is also very different or frequently reverse than those by other two algorithms.

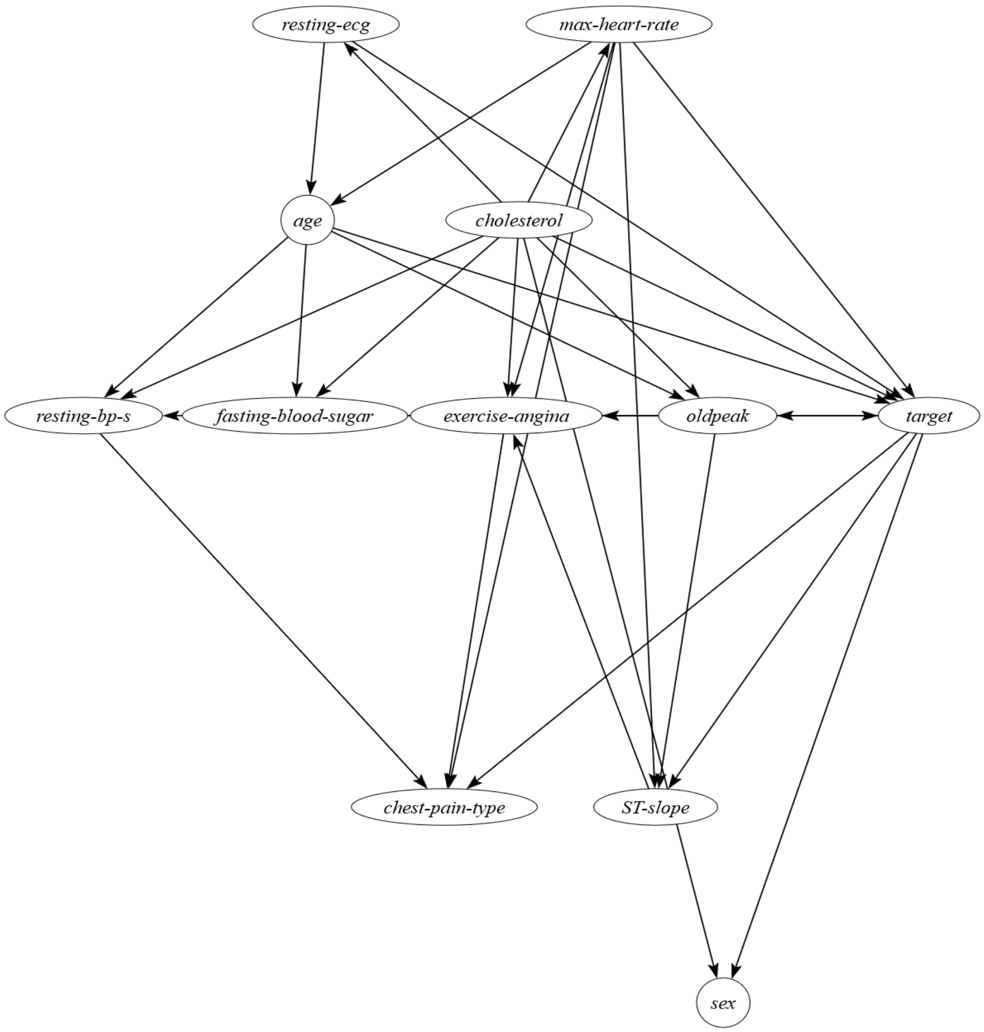


Figure 8: Causal Graph discovered by GES algorithm for the Real Dataset

Also, based on the estimation by the three causal discovery algorithms, the final casual graph could be computed by considering the majority voting for each pair of variables in the graph. That is an edge can be only considered to incorporate in the final causal graph only if the obtained vote count is greater than or equal to floor((a/2)+ 1), where a is the total 2 number of algorithms used [9]. In our case, we have used a total of 3 algorithms for causal structure search. Hence, we will consider those edges only which get a total of 2 or more votes [since, floor(3/2 +1) = 2]. But as this is a naive approach to select the true edges without any defined soundness, we did not use it.

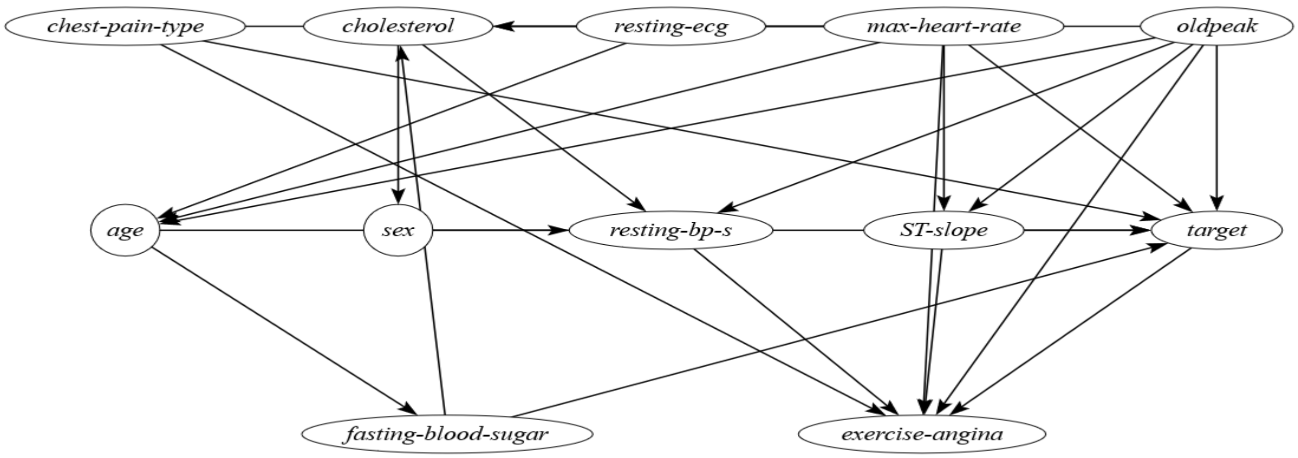


Figure 9: Causal Graph discovered by PC algorithm for the Real Dataset

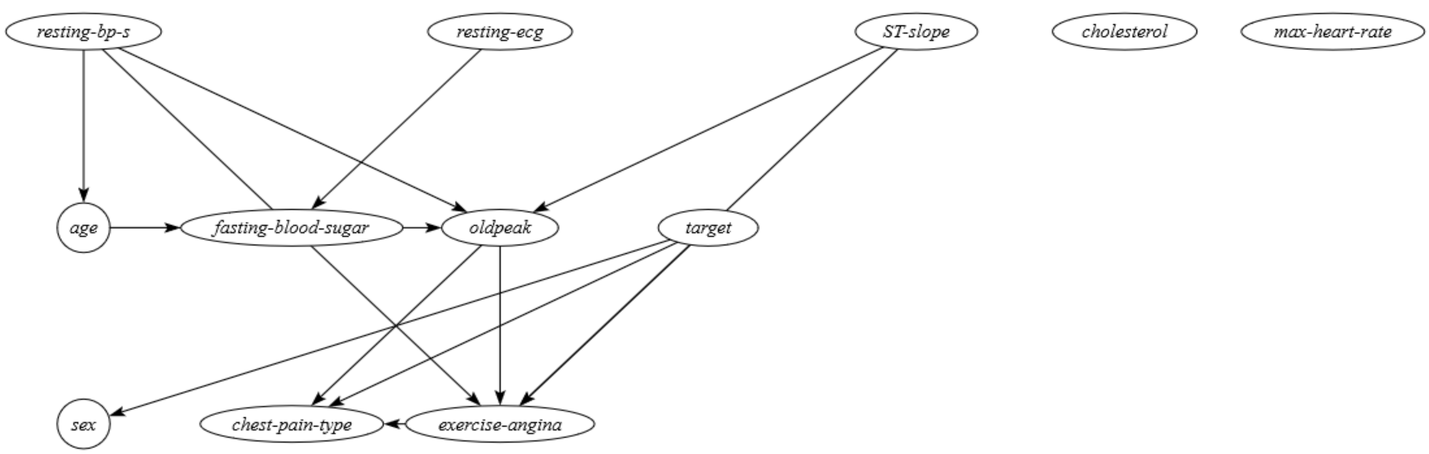


Figure 10: Causal Graph discovered by LINGAM algorithm for the Real Dataset

5 Conclusion

The analysis of the experimental results implies that correlation does not imply causation. For the LUCAS dataset, we found three variables highly correlated to the target variable *Lung Cancer* but none of these were a cause of it. Also, we see that *Fatigue* and *Coughing* have no causal effect upon *Lung Cancer* . Though they showed medium to high correlation with *Lung Cancer* which implies these are the effects of *Lung Cancer*, not a cause of it. This finding also implies that decision in a healthcare setting can not be made solely basing upon correlation. In fact, correlation dependent decisions can be risky and faulty. Thus, casual relationships need to be considered for making any informed decision and optimization of policy particularly in cases where the patient health is involved. We believe our approach will be useful for people in healthcare sector to take reliable decisions in a systematic way. This approach can be further improved by evaluating upon large medical datasets and incorporating professionals suggestions.

References

[1] Figueroa Anthony. *Correlation is not causation.* Aug. 2019. URL: `https://towardsdatascience.com/correlation-is-not-causation-ae05d03c1f53`.

[2] Mario Bunge. *Causality and modern science.* Routledge, 2017.

[3] Ruichu Cai et al. "On the Role of Entropy-based Loss for Learning Causal Structures with Continuous Optimization". In: *arXiv preprint arXiv:2106.02835* (2021).

[4] David Maxwell Chickering. "Learning equivalence classes of Bayesian-network structures". In: *The Journal of Machine Learning Research* 2 (2002), pp. 445–498.

[5] Tom Claassen, Joris Mooij, and Tom Heskes. "Learning sparse causal models is not NP-hard". In: *arXiv preprint arXiv:1309.6824* (2013).

[6] Diego Colombo et al. "Learning high-dimensional directed acyclic graphs with latent and selection variables". In: *The Annals of Statistics* (2012), pp. 294–321.

[7] Susanna S Epp. *Discrete mathematics with applications.* Cengage learning, 2010.

[8] David Galles and Judea Pearl. "Testing identifiability of causal effects". In: *arXiv preprint arXiv:1302.4948* (2013).

[9] Md Osman Gani et al. "Structural Causal Model with Expert Augmented Knowledge to Estimate the Effect of Oxygen Therapy on Mortality in the ICU". In: *arXiv preprint arXiv:2010.14774* (2020).

[10] Dan Geiger and David Heckerman. "Learning gaussian networks". In: *Uncertainty Proceedings 1994.* Elsevier, 1994, pp. 235–243.

[11] Thomas A Glass et al. *Causal inference in public health.* 2013.

[12] Clark Glymour, Kun Zhang, and Peter Spirtes. "Review of causal discovery methods based on graphical models". In: *Frontiers in genetics* 10 (2019), p. 524.

[13] Eduardo Hariton and Joseph J Locascio. "Randomised controlled trials—the gold standard for effectiveness research". In: *BJOG: an international journal of obstetrics and gynaecology* 125.13 (2018), p. 1716.

[14] Alain Hauser and Peter Bühlmann. "Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs". In: *The Journal of Machine Learning Research* 13.1 (2012), pp. 2409–2464.

[15] Austin Bradford Hill. *The environment and disease: association or causation?* 1965.

[16] Yimin Huang and Marco Valtorta. "Pearl's calculus of intervention is complete". In: *arXiv preprint arXiv:1206.6831* (2012).

[17] Aapo Hyvärinen et al. "Estimation of a structural vector autoregression model using non-gaussianity." In: *Journal of Machine Learning Research* 11.5 (2010).

[18] Dominik Janzing and Bernhard Schölkopf. "Causal inference using the algorithmic Markov condition". In: *IEEE Transactions on Information Theory* 56.10 (2010), pp. 5168–5194.

[19] Peter JF Lucas, Linda C Van der Gaag, and Ameen Abu-Hanna. "Bayesian networks in biomedicine and health-care". In: *Artificial intelligence in medicine* 30.3 (2004), pp. 201–214.

[20] *LUCAS and LUCAP are lung cancer toy datasets.* URL: http://www.causality.inf.ethz.ch/data/LUCAS.html.

[21] Daniel Malinsky and Peter Spirtes. "Learning the structure of a nonstationary vector autoregression". In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, pp. 2986–2994.

[22] André Moser, Milo A Puhan, and Marcel Zwahlen. *The role of causal inference in health services research I: tasks in health services research.* 2020.

[23] Roxana Pamfil et al. "Dynotears: Structure learning from time-series data". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 1595–1605.

[24] Judea Pearl. "The do-calculus revisited". In: *arXiv preprint arXiv:1210.4852* (2012).

[25] Shohei Shimizu et al. "A linear non-Gaussian acyclic model for causal discovery." In: *Journal of Machine Learning Research* 7.10 (2006).

[26] Ali Shojaie and George Michailidis. "Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs". In: *Biometrika* 97.3 (2010), pp. 519–538.

[27] Ilya Shpitser and Judea Pearl. "Complete identification methods for the causal hierarchy". In: *Journal of Machine Learning Research* 9 (2008), pp. 1941–1979.

[28] Peter Spirtes. "An anytime algorithm for causal inference". In: *International Workshop on Artificial Intelligence and Statistics*. PMLR. 2001, pp. 278–285.

[29] Peter Spirtes et al. *Causation, prediction, and search.* MIT press, 2000.

[30] Krishnaiyan Thulasiraman and Madisetti NS Swamy. *Graphs: theory and algorithms.* John Wiley & Sons, 2011.

[31] Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. "The max-min hill-climbing Bayesian network structure learning algorithm". In: *Machine learning* 65.1 (2006), pp. 31–78.

[32] Michael R Waldmann and York Hagmayer. "Causal reasoning." In: (2013).

[33] Jiji Zhang. "On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias". In: *Artificial Intelligence* 172.16-17 (2008), pp. 1873–1896.

[34] Wenhao Zhang, Ramin Ramezani, and Arash Naeim. *Causal Inference in medicine and in health policy, a summary.* 2021.