

CDANs: Temporal Causal Discovery from Autocorrelated and Non-Stationary Time Series Data

Muhammad Hasan Ferdous

*Causal AI Lab, Department of Information Systems
University of Maryland, Baltimore County
Baltimore, Maryland, USA*

H.FERDOUS@UMBC.EDU

Uzma Hasan

*Causal AI Lab, Department of Information Systems
University of Maryland, Baltimore County
Baltimore, Maryland, USA*

UZMAHASAN@UMBC.EDU

Md Osman Gani

*Causal AI Lab, Department of Information Systems
University of Maryland, Baltimore County
Baltimore, Maryland, USA*

MOGANI@UMBC.EDU

Editor: Editor's name

Abstract

Time series data are found in many areas of healthcare such as medical time series, electronic health records (EHR), measurements of vitals, and wearable devices. Causal discovery, which involves estimating causal relationships from observational data, holds the potential to play a significant role in extracting actionable insights about human health. In this study, we present a novel constraint-based causal discovery approach for autocorrelated and non-stationary time series data (CDANs). Our proposed method addresses several limitations of existing causal discovery methods for autocorrelated and non-stationary time series data, such as high dimensionality, the inability to identify lagged causal relationships and overlooking changing modules. Our approach identifies lagged and instantaneous/contemporaneous causal relationships along with changing modules that vary over time. The method optimizes the conditioning sets in a constraint-based search by considering lagged parents instead of conditioning on the entire past that addresses high dimensionality. The changing modules are detected by considering both contemporaneous and lagged parents. The approach first detects the lagged adjacencies, then identifies the changing modules and contemporaneous adjacencies, and finally determines the causal direction. We extensively evaluated our proposed method on synthetic and real-world clinical datasets, and compared its performance with several baseline approaches. The experimental results demonstrate the effectiveness of the proposed method in detecting causal relationships and changing modules for autocorrelated and non-stationary time series data.

1. Introduction

The ever-increasing adoption of electronic health records (EHR) in modern healthcare has facilitated the collection of a large amount of observational data that can be used in diagnostics, disease identification, treatment effect estimation, etc. (Cowie et al., 2017; Nordo

et al., 2019; Casey et al., 2016). Causal inference techniques can leverage this vast amount of observational clinical data to derive new therapies or valuable insights (Cowie et al., 2017). However, such an inference requires developing a graphical representation, commonly in the form of a directed acyclic graph (DAG), that captures the causal relationships between the variables (Glymour et al., 2016). Causal discovery (CD)/ causal structure learning is the process of identifying the causal graph (which represents the causal relations) from observational data (Spirtes et al., 2000). Sometimes data alone may not fully capture the actual underlying causal mechanism, making it necessary to utilize additional sources of causal information to gain a complete understanding (Pearl et al., 2016). Numerous efforts have been made to integrate causal information from various sources in the discovery of causal relationships (Meek, 2013; Adib et al., 2022; Hasan and Gani, 2022).

Over the years, substantial methods have been developed to estimate the underlying causal graph from observational time-series data (Hasan et al., 2023). The analysis of time series data has become increasingly important in various fields including healthcare, and understanding the causal relationships between variables can provide valuable insights into the dynamics of complex systems. Often, we may encounter multivariate time series data which is non-stationary and autocorrelated (i.e. past influences the present, and future) (Lawton et al., 2001). However, most of the existing temporal CD approaches perform poorly when the time-series data are both non-stationary and autocorrelated. The presence of these components makes causal structure discovery from time-series data a challenging task. Especially, it is more challenging in a multivariate distribution where two or more variables are time-dependent, and both autocorrelation and lagged causal relationships exist (Hamman, 1967). Moreover, the seasonal and cyclical nature of variables has a time influence that can cause a change in their distributions. This time influence is known as *changing modules* and can be represented using a *surrogate* variable C to represent the hidden factors that cause the distribution shift of the variables (Zhang et al., 2017). To find causal relationships in autocorrelated data, some approaches use conventional conditional independence (CI) tests between variables that may include the whole past in the conditioning set (Spirtes et al., 2000; Colombo et al., 2014). This might result in significantly increasing the number of conditional variables. Further, the conditioning set may contain some uncorrelated variables as well (Entner and Hoyer, 2010; Malinsky and Spirtes, 2018). The inclusion of such variables in the conditioning set increases the dimensionality, lowers the detection power, and also, can yield misleading results (Bellman, 1966; Runge et al., 2019a). Although the PCMCI+ (Runge, 2020) method tries to address the problem of high dimensionality by optimizing the conditioning set while conducting CI tests, it does not consider time dependency among the variables that can result in false causal edges. Huang et al. (2020) proposed an approach called "extended CD-NOD" to identify time dependency. But, it uses a conventional PC algorithm (Spirtes et al., 2000) initially developed for non-temporal data to identify the temporal causal structure. As a result, it inherits the limitations of the PC algorithm when applied to time series data. Specifically, when applied to high-dimensional temporal datasets, the PC algorithm has two main limitations. Firstly, its runtime is exponential in relation to the number of variables, rendering it inefficient for high-dimensional settings. Secondly, its results are dependent on the order of variables in the input dataset, meaning that changing the order of variables may alter the results (Le

et al., 2016). For these reasons, the identified causal structure using extended CD-NOD is order-dependent, suffers from high dimensionality, and is unable to handle autocorrelation.

Therefore, to address these challenges, we propose an algorithm (CDANs) for causal discovery from autocorrelated and non-stationary time series data which works as follows. *First*, it finds the *lagged parents* to *avoid conditioning on irrelevant variables* and thereby, reduces the conditioning set size that addresses high dimensionality. This enables CDANs to systematically prevent conditioning on the entire past. *Second*, it develops a partially completed undirected graph using lagged parents, contemporaneous variables, and the surrogate variable. *Third*, it estimates the *causal skeleton* by identifying the *changing modules* and the *contemporaneous relations* using marginal and optimized CI tests. *Fourth*, it determines the *causal directions* using the *time order of causation*, *generalization of invariance*, and *independent changes in causal modules* (Runge, 2020; Huang et al., 2020). We evaluate CDANs using synthetic datasets with 4, 6, and 8 variables with different lags up to period 8, and a real-world clinical dataset of 12 variables. We describe the synthetic data generation process and clinical data in Section 6.1, and discuss the clinical application’s cohort in the Section 5. Our *contributions* are summarized below:

- We propose a novel temporal causal discovery approach that considers both autocorrelation and non-stationarity properties of time series data. Our method can detect both contemporaneous and lagged relations between the variables as well as the variables whose distribution changes over time.
- We evaluate the performance of CDANs on real-world clinical and synthetic datasets and compare it with multiple baselines for temporal CD. The empirical results show that our method outperformed baselines (Runge, 2020; Ogarrío et al., 2016; Ramsey et al., 2017; Lam et al., 2022; Malinsky and Spirtes, 2018; Huang et al., 2020) in multiple metrics across different experimental settings.
- The consider clinical application entails a timely research problem about oxygen therapy intervention in ICU. It is related to recovering the causal structure of 12 time series variables in the ICU which is useful in a variety of disease conditions, including severe acute respiratory syndrome COVID-19.

In Figure 1, we show a causal graph of autocorrelated nonstationary time-series data with lagged, contemporaneous variables and changing modules. Time dependency, lagged dependencies, and contemporaneous dependencies are denoted by red arrows, blue arrows, and green arrows, respectively.

Generalizable Insights about Machine Learning in the Context of Healthcare

Time series data are prevalent across various healthcare domains, encompassing medical time series, electronic health records (EHR), vital sign measurements, and wearable device data. Identification of causal relations from observational data is a growing area of research in machine learning. The knowledge of causal graphs represents the underlying data-generating mechanism that can support crucial decision-making in several areas of healthcare. The methods that support causal discovery from time series data are of great importance as these types of data is common in healthcare sectors. Investigating these data and causal graphs using appropriate causal inference techniques can lead to actionable

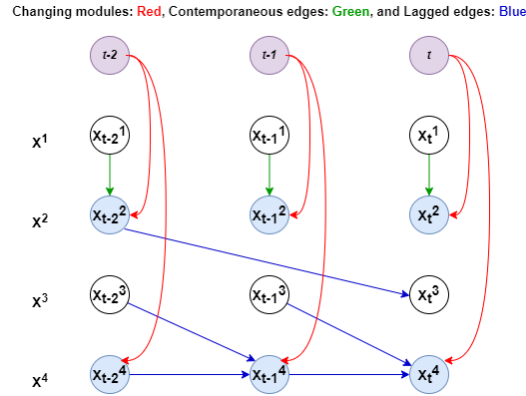


Figure 1: A causal graph of autocorrelated nonstationary time-series data with lagged, contemporaneous variables and changing modules. Time dependency, lagged dependencies, and contemporaneous dependencies are denoted by red, blue, and green arrows respectively.

insights. Time series data often have properties such as non-stationarity, autocorrelation, and time dependency that need to be addressed when performing causal discovery. Existing temporal CD approaches lack the ability to handle all of these properties efficiently, especially time dependency. Identifying time dependencies among variables in multivariate time series data has become increasingly important in healthcare due to its potential to improve patient outcomes and advance our understanding of disease progression (Batal et al., 2016). The analysis of temporal relationships among variables can reveal critical insights into various aspects of healthcare, such as patient monitoring, early warning systems, disease progression, personalized medicine, and treatment effectiveness evaluation (Clifton et al., 2012; Churpek et al., 2016; Jameson and Longo, 2015; Nemati et al., 2016). By investigating these dependencies, researchers and practitioners can gain valuable insights into the complex dynamics of patient health, improve disease management, and ultimately enhance patient outcomes (Rajkomar et al., 2018). Thus, in this study, we propose an approach called CDANs that can effectively handle the aforementioned crucial properties of time series data and detect both contemporaneous and lagged relations between the variables, as well as the variables whose distribution changes over time. Our approach will improve the discovery of temporal causal graphs with the mentioned properties, which, in turn, will help healthcare researchers to estimate treatment effects better and make informed clinical decisions.

2. Background

Causal discovery is the process of identifying the underlying causal mechanisms from data, typically represented in the form of a directed acyclic graph (DAG). Our proposed method CDANs aim to uncover the underlying causal structure in autocorrelated and non-stationary

time series data with changing modules. In this section, we discuss the common notations and provide a brief overview of autocorrelation and changing modules.

For a set of time series variables $X = \{X^1, X^2, \dots, X^N\}$, the value for each variable at timestamp t can be represented as $X_t = \{X_t^1, X_t^2, \dots, X_t^N\}$. Here the j^{th} variable at time point t can be represented as an arbitrary measurable function f_j such that $X_t^j = f_j(P(X_t^j), \varepsilon_t^j)$, where $P(X_t^j)$ are the parents of X_t^j , ε_t^j is the mutually and serially independent dynamic noise. We denote X_t^- be the past observations up to time $(t-1)$. Thus parents of X_t^j can be defined as $P(X_t^j) \subset X_{t+1}^- = (X_t, X_{t-1}, \dots) \setminus \{X_t^j\}$, and the lagged parents can be defined as $\text{LPA}(X_t^j) = P(X_t^j) \cap X_t^-$. We briefly discuss autocorrelation and changing modules below.

Autocorrelation: Autocorrelation represents the degree of similarity between a given time series and a lagged version of itself over successive time intervals. It measures the relationship between a variable’s current value and its past values (Bence, 1995). Due to the dynamic nature of the autocorrelated data, a series of conditional independence (CI) tests need to be performed to find out the causal skeleton. For a particular CI test, sample size and significance level are fixed, thus the detection power of a CI test can be improved by lowering the dimensionality and increasing the effect size. The effect size of a conditional independence test is typically reported as a measure of the degree of dependence or independence between the two variables, given the third variable(s). However, *including uncorrelated variables in the conditioning set increases the dimensionality* resulting in lower detection power of the CI test, also known as the “*curse of dimensionality*” (Bellman, 1966).

Changing modules: Considering the non-stationary nature of the data, some variables will inevitably change their distribution over time. These are called changing modules, as described by Zhang et al. (2017), which are the functions of time or domain index (Figure 1). These influences can often act as confounders or latent common causes (Zhang et al., 2017) which can be divided into three types: i) as a function of domain index or smooth function of time index, ii) fixed distribution with no functional relationships, and iii) non-stationary variables with no functional relationship (Huang et al., 2020). In this work, we limit our focus to the first type of confounders to detect the changing modules. Detecting the edges due to the changing modules (i.e. the edges between time and variables in Figure 1) while learning causal structure from temporal data is important because ignoring such confounders may lead to the estimation of false or incorrect causal links between the variables.

3. Related Work

The discovery of causal structures from observational time series data presents significant challenges due to time order, data distribution, and autocorrelation (Runge et al., 2019a). The Granger causality (Granger, 1969) predicts one-time series based on another time series but is limited in its ability to detect true causal links when variables are generated from a common third variable, relationships are non-linear, or data is non-stationary (Maziarz, 2015). One of the earliest and most popular constraint-based approaches is the PC algorithm (Spirtes et al., 2000). It tests the conditional independence relationships between variables to construct a causal skeleton and then orients the remaining edges based on a set of orientation rules. The FCI (Spirtes et al., 2000) algorithm extended PC by incorporating

additional tests for conditional independence and handling latent variables (Spirtes et al., 2000). The PC algorithm serves as the foundation for numerous other algorithms such as RFCI (Colombo et al., 2012), PC-stable (Colombo et al., 2014), and Parallel-PC (Le et al., 2016). Modifications to the PC and FCI algorithms have enabled the identification of causal structures in time series data that account for unobserved confounders using time order and stationarity assumptions (Chu et al., 2008; Entner and Hoyer, 2010; Malinsky and Spirtes, 2018). However, these approaches are hindered by high dimensionality and autocorrelation present in time series data (Runge et al., 2019b).

The GES algorithm uses a greedy search strategy to explore the space of possible causal structure (Chickering, 2002). It initializes the search space with an empty graph, evaluates all possible additions and deletions of edges to the current graph, and selects the best one based on a score metric. FGES improves upon GES with a more efficient scoring algorithm (Ramsey et al., 2017). GFCI combines the strengths of several algorithms, including PC, FCI, and GES, to discover causal relationships in both linear and nonlinear models (Ogarrio et al., 2016). Moreover, in the field of economics, Structural Vector Autoregression (SVAR) (Sims, 1980) is a widely used approach, which has been extended with the GFCI algorithm to discover causal relationships in time series data (Malinsky and Spirtes, 2018). Recent developments include the Greedy Sparse Permutation (GSP) algorithm (Solus et al., 2021), and the Greedy Relations of Sparsest Permutation (GRaSP) algorithm (Lam et al., 2022), which combines multiple algorithms to improve performance.

Runge et al. (2019a) proposed the PCMCI algorithm to address high dimensionality by optimizing the conditional set of the CI tests. At first, the algorithm performs marginal independence tests for every pair of variables and removes independent causal edges. In the subsequent iterations, the algorithm adds additional variables on the conditional set according to the largest effect size derived from the earlier step and keeps removing the independent variables from the parent set. The algorithm stops after performing a predefined number of iterations or after including all variables in the conditioning set. Later, Runge (2020) proposed an extension of the PCMCI algorithm known as PCMCI+, that identifies both lagged and contemporaneous edges. The lagged edges are identified using PCMCI, and contemporaneous edges are identified by constructing contemporaneous adjacencies, and then performing MCI tests between those variables. Although, some additional spurious edges can still be detected because *none of the algorithms consider time influence*. Thus, both algorithms are unable to detect changing modules under the causal sufficiency assumption. If several variables of the underlying model are influenced by a time factor, it can act as a confounder and thereby, yield false edges between the variables.

Zhang et al. (2017) introduced CD-NOD, an algorithm for discovering causal structures from heterogeneous data where observed data are independent but not identically distributed, though it does not account for autocorrelation. Later, Huang et al. (2020) proposed an extension to the CD-NOD algorithm, named "extended CD-NOD", to address autocorrelation in time series data. However, the approach closely follows the PC algorithm (Spirtes et al., 2000) and inherits its limitations. The extended CD-NOD first identifies changing modules by performing conditional independence tests between contemporaneous and surrogate variables and then applies the PC algorithm to detect lagged and contemporaneous causal edges. Finally, the orientation rules used in CD-NOD are applied to obtain the final causal graph. Nevertheless, changing modules are identified only using

contemporaneous variables, which can lead to false positives as lagged parents in the conditioning set may be omitted. Furthermore, the algorithm is not order-independent, implying that changing the variable order may result in a different causal graph. Moreover, the practical applicability of the extended CD-NOD is limited due to the lack of experimental results and implementation code, which restricts its adoption in real-life scenarios.

Apart from the mentioned methods, approaches based on the technique of continuous optimization have been proposed for causal discovery (Zheng et al., 2018), and for the analysis of high-dimensional autocorrelated time series data (Pamfil et al., 2020; Sun et al., 2021). Despite using non-combinatorial optimization to identify causal structures, these methods may result in multiple minima, and the returned DAGs may not necessarily represent causal relationships (Reisach et al., 2021; Kaiser and Sipos, 2022). Additionally, these approaches cannot handle data re-scaling and may produce different DAGs when dealing with different scales (Kaiser and Sipos, 2022). Among the other approaches, DYNOTEARS (Pamfil et al., 2020) is a score-based method for dynamic Bayesian networks that simultaneously estimates contemporaneous and time-lagged relationships. More recently, Bussmann et al. (2021) introduced a neural approach, NAVAR, capable of discovering nonlinear relationships through the training of a deep neural network that extracts Granger causal influences from the time evolution in a multivariate time series.

Despite the advancements in temporal causal discovery methods, current approaches still struggle to comprehensively address both high-dimensionality and changing modules when learning causal structures from observational time series data. In our proposed approach, we address the limitations of the existing approaches by identifying lagged causal edges using MCI tests at the first step and then leveraging the lagged parents to identify changing modules and contemporaneous edges. Our proposed approach is order independent, can handle autocorrelation, and detect changing modules.

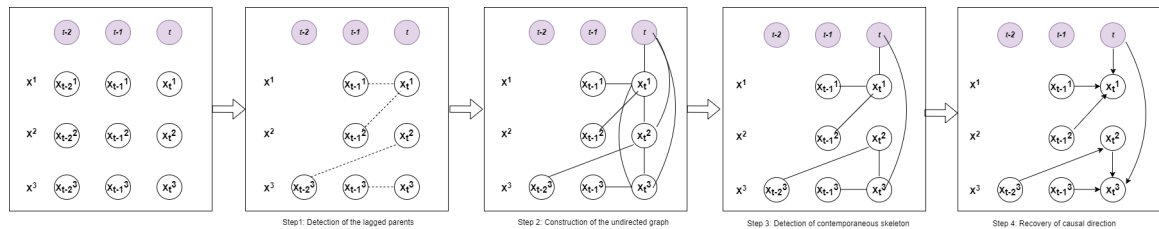


Figure 2: Schematic representation of the proposed CDANs methodology. The approach consists of four steps: (1) identification of lagged adjacencies using MCI tests, (2) construction of a partial undirected graph incorporating lagged adjacencies, contemporaneous variables, and surrogate variables, (3) execution of optimized CI tests to obtain the causal skeleton, and (4) application of orientation rules to determine the final causal structure.

4. Methodology

In this section, we discuss our proposed algorithm CDANs¹ with a brief introduction to the assumptions considered.

4.1. Assumptions

We assume that all or at least some variables will change according to some unobserved confounders that can be represented as a smooth function of time. Thus, we assume that causal sufficiency does not hold for the given data. However, we represent all unobserved confounders using a surrogate variable, and thus, consider the entire model with observed variables and a surrogate variable to achieve causal sufficiency also known as *pseudo causal sufficiency* (Huang et al., 2020).

4.2. Proposed Algorithm

We present our proposed approach CDANs in Algorithm 1, and discuss the details of the steps it has in the following paragraphs.

Step 1 (*Detection of lagged parents*): Let X_t^j be the j^{th} observation at time t , $X_{t-\tau}^i$ be the i^{th} observation at lag τ , and X_t^- be the past observations. Here, variables at time t are the contemporaneous variables, and variables that occurred before time t are lagged variables. We first find the *lagged parents* to *avoid conditioning on irrelevant variables*. In this step, we use PCMCI for all X_t^j and $X_{t-\tau}^i$ where $i = 1, 2, \dots, m$ and derive the lagged parent set $LPA(X_t^j)$ for every X_t^j . Here, m is the total number of variables. Derivation of the lagged parents is done using the following steps: first, unconditional tests are conducted between X_t^j and all lagged variables, with p-values and effect sizes recorded. Then, a new lagged parent set is constructed from only the variables with significant unconditional tests, sorted by effect size. Next, conditional independence tests are performed between X_t^j and the variables in the new lagged parent set, with non-significant links removed to construct a new lagged parent set. This process is iterated, with variables added to the conditioning set in descending order of effect size until all variables are included in the conditional set. After the detection of the lagged parents, this step produces a causal skeleton between the contemporaneous variables and their lagged parents. We illustrate this step in Figure 2 (Step 1), e.g., the lagged parent of X_t^1 are X_{t-1}^1 and X_{t-1}^2 , and lagged parent of X_t^3 is X_{t-1}^3 . This reduces the size of the conditioning set and thus, prevents conditioning on the entire past to address high dimensionality. Thus, CDANs eliminates the inclusion of uncorrelated variables in the conditioning set, resulting in fewer variables compared to existing approaches. This also helps to improve the detection power and reduce run time.

Step 2 (*Construction of the undirected graph*): After detecting the lagged parents, the algorithm creates a partially complete undirected graph G between the lagged parents $LPA(X_t)$, contemporaneous variables X_t , and the surrogate variable C (used to represent time). This helps the subsequent steps in the algorithm where we condition only on the respective lagged parent sets of the variables in CI tests. The resulting smaller

¹<https://github.com/hferdous/CDANs>

conditioning set size improves detection power. In the end, we get a complete undirected graph over the variables $(X_t \cup LPA(X_t) \cup C)$ (Figure 2).

Algorithm 1: Causal Discovery from Autocorrelated and Non-stationary data (CDANs)

1. Conduct CI tests between X_t^j and $X_{t-\tau}^i$ for all i ($i = 1, 2, \dots, m$) utilizing PC1 algorithm with lagged conditions and derive the parent lagged set $LPA(X_t^j)$ for every X_t^j .
 2. Build a partially complete undirected graph G over the variable set $(X_t \cup LPA(X_t) \cup C)$
 3. For every $j \in (1, 2, \dots, m)$, conduct marginal and CI test between X_t^j and C . Remove the edge between X_t^j and C if $X_t^j \perp\!\!\!\perp C$ conditional on a subset of $(LPA(X_t^j) \cup X_t) \setminus (X_t^j)$. At the same time, for all ($i \neq j = 1, 2, \dots, m$), test for marginal and CI between X_t^i and X_t^j . Remove the edge between X_t^i and X_t^j if they are independent conditional on a subset of $(X_t \cup LPA(X_t^i) \cup LPA(X_t^j) \cup C) \setminus (X_t^i, X_t^j)$.
 4. For $\tau = 1, 2, \dots, \tau_{\max}$, orient (i, j) as $i \rightarrow j$ according to the flow of time. Orient (C, X_t^j) as $C \rightarrow X_t^j$ if X_t^j is adjacent to C . For triple of the form $(C - X_t^i - X_t^j)$, recall the conditional set of the CI test between C and X_t^j . If the conditioning set does not include X_t^i , orient the triple as $C \rightarrow X_t^i \leftarrow X_t^j$. Otherwise, orient as $C \rightarrow X_t^i \rightarrow X_t^j$. When both X_t^i and X_t^j are adjacent to C , use extended HSIC to orient the edge between X_t^i and X_t^j .
-

Step 3 (Detection of changing modules and contemporaneous causal skeleton): Changing modules are assumed to be a smooth function of time and the time dependency is represented by a surrogate variable C . CDANs performs a series of kernel-based conditional independence (KCI) tests (Zhang et al., 2012) between the contemporaneous variables and the surrogate variable to identify the complete causal skeleton. To detect changing modules, CDANs starts with unconditional independence tests between the contemporaneous variables X_t^j and the surrogate variable C ; and keeps adding other variables in the conditioning set from the parent set $(LPA(X_t^j) \cup X_t) \setminus (X_t^j)$. It removes the edge between X_t^j and C if they are independent. At the end of this step, it produces a causal skeleton that has all of the components— contemporaneous edges, lagged edges, and the edges between contemporaneous variables and C .

Step 4 (Recovery of causal direction): The goal in this step is to recover the causal directions from the skeleton. We assume that the cause-effect relationships follow the flow of time i.e., the past always causes the future. Using this assumption, we will orient $(X_{t-\tau}^i, X_t^j)$ as $(X_{t-\tau}^i \rightarrow X_t^j)$ for all $\tau = 1, 2, \dots, \tau_{\max}$. As C is a surrogate variable for confounders which is one of the causes of the changing modules, we can then

orient (C, X_t^j) as $C \rightarrow X_t^j$ if X_t^j is adjacent to C . We then consider the triples of the form $(C - X_t^i - X_t^j)$ and use the conditional sets of the CI test, from step 3, between C and X_t^j to determine the direction. If the conditioning set does not include X_t^i , it orients the triple as $C \rightarrow X_t^i \leftarrow X_t^j$. Otherwise, orients as $C \rightarrow X_t^i \rightarrow X_t^j$. If both X_t^i and X_t^j are adjacent to C , then the causal direction between X_t^i and X_t^j is determined based on the causal effect from X_t^i to X_t^j and vice versa. We can calculate the causal effect for a given pair of variables (X_t^i, X_t^j) ; here $(X_t^i$ and X_t^j are independent if one of $P(X_t^i)$ and $P(X_t^i|X_t^j)$ changes while the other remains invariant. We determine the causal direction as $X_t^i \rightarrow X_t^j$ if $P(X_t^i)$ and $P(X_t^j|X_t^i)$ are independent but $P(X_t^j)$ and $P(X_t^i|X_t^j)$ are dependent. We use an extended version of Hilbert Schmidt Independence Criterion (HSIC) (Huang et al., 2020) to measure the dependence between $P(X_t^j|X_t^i)$ and $P(X_t^i)$, denoted by $\partial_{X_t^i \rightarrow X_t^j}$, and the dependence between $P(X_t^i|X_t^j)$ and $P(X_t^j)$, denoted by $\partial_{X_t^j \rightarrow X_t^i}$. Based on the dependencies between X_t^i and X_t^j , we can orient (X_t^i, X_t^j) as $(X_t^i \rightarrow X_t^j)$ if $\partial_{X_t^i \rightarrow X_t^j} < \partial_{X_t^j \rightarrow X_t^i}$. Otherwise, orient as $(X_t^i \leftarrow X_t^j)$.

The algorithmic performance of CDANs depends on the sparsity of the causal relationships within the network. When these relationships are sparse, the algorithm converges more quickly due to fewer lagged edges to orient and fewer contemporaneous conditioning sets to iterate through. In comparison to the original PC algorithm, which has a worst-case exponential complexity, CDANs have significantly lower complexity. The orientation of lagged edges has polynomial complexity (Runge et al., 2019b), while the detection of changing modules and contemporaneous edges only requires iteration through contemporaneous conditioning sets. As a result, the worst-case exponential complexity of CDANs is only applicable to the number of nodes in the network, and the surrogate variable, rather than the maximum number of lagged edges, i.e., for a time series dataset comprising N variables with a maximum lag of τ_{\max} , CDANs has a worst-case polynomial complexity applies to $(N + 1)$, as opposed to the $N\tau_{\max}$ complexity characterizing the PC algorithm.

5. Cohort

In this section, we briefly discuss the cohort of our real-world clinical application. There are approximately 3 million patients per year in the US that receive invasive mechanical ventilation (IMV) in intensive care units (ICU) (Wunsch et al., 2010; Adhikari et al., 2010). Most of the patients receiving IMV also receive supplemental oxygen therapy (OT) to maintain safe levels of tissue oxygenation estimated through peripheral oxygen saturation, SpO_2 (Vincent and De Backer, 2013). Current clinical practice and recommendations related to OT are based on physiological values in healthy adults and lack systematic results from large-scale clinical trials (Meade et al., 2008; Panwar et al., 2016; Shari et al., 2000). We collected a clinical observational dataset on OT from MIMIC-III (Johnson et al., 2016) database following the study by Gani et al. (2023); Bikak et al. (2020). The study emulates a pilot randomized control trial (RCT) on OT by closely following the study protocol as described in (Panwar et al., 2016). We extracted 12 time-series variables (recorded every 4 hours) related to oxygenation parameters and ventilator settings. We described these

variables in Section 6.1.2, and the selection of these variables is based on the parameters used in the pilot RCT (Panwar et al., 2016). Our goal is to discover the causal structure underlying these variables and leverage it later (not part of this study) to perform virtual experiments using observational data.

6. Experiments

We evaluated our proposed approach CDANs against the following baselines: (1) PCMCI+ (Runge, 2020), (2) Fast Greedy Equivalence Search (FGES) (Ramsey et al., 2017) which is an optimized and parallelized version of GES (Chickering, 2002), (3) Greedy Fast Causal Inference (GFCI) (Ogarrio et al., 2016) which is a combination of FCI (Spirtes et al., 2000) and FGES (Ramsey et al., 2017) algorithms, (4) Greedy Relations of Sparsest Permutation (GRaSP) (Lam et al., 2022) which is a generalization and extension of GSP (Greedy Sparsest Permutation) algorithm, and (5) SVAR-GFCI (Structural Vector Autoregression with Greedy Fast Causal Inference) is an algorithm that combines the use of Structural Vector Autoregression (SVAR) and Greedy Fast Causal Inference (GFCI) to infer the causal structure of a system from time series data (Malinsky and Spirtes, 2018). We utilize the `tetrad` package for their implementations which is available at <https://github.com/cmu-phil/tetrad>. Apart from these approaches, CD-NOD is the only approach that can detect changing modules. However, it cannot identify lagged causal edges (Huang et al., 2020). Hence, we compare the performance of CD-NOD with CDANs only on contemporaneous causal edges. The performance of all the approaches has been evaluated based on three evaluation metrics– the true positive rate (TPR), the false discovery rate (FDR), and the structural hamming distance (SHD) (Norouzi et al., 2012). FDR and SHD are better when lower, whereas a higher TPR indicates better performance. Details about TPR, FDR, and SHD are given in Appendix A. Code and datasets are available at <https://github.com/hferdous/CDANs>.

6.1. Datasets

6.1.1. SYNTHETIC DATASET

We demonstrate the performance of our proposed approach on a variety of synthetic datasets consisting of 4, 6, and 8 variables with lag periods of 2, 4, 6, and 8. The data generation process is described in Appendix C. For ease of understanding, we present the associated causal graph for 4 variables with lag 2 in Figure 3 (True Causal Graph). The 4-variable models with lag 2 have 1 changing module, 1 contemporaneous edge, 2 autocorrelated edges, and 2 lagged edges. The data generating process of i^{th} variable at time t with maximum lag τ_{\max} can be mathematically described as,

$$X_t^i = a_i f_i(X_{t-1}^j) + b_i f_i(X_{t-2}^k) + \dots + c_i f_i(X_{t-\tau_{\max}}^l) + d_i f(t) + \varepsilon_t^i,$$

for $i \in \{1, \dots, N\}$ where $f_i(x)$ is the non-linear functional dependency; a_i , b_i , c_i and d_i are coefficient parameters, for autocorrelated variables $i = j = k = l$, for other lagged relations $i \neq j \neq k \neq l$. Only changing modules have time dependency as an additional component $d_i \cdot f(t)$ where $f(t)$ is a sine or cosine function of time. We define changing modules as a sine or cosine function of time. Doing this ensures both non-linearity and time dependency of

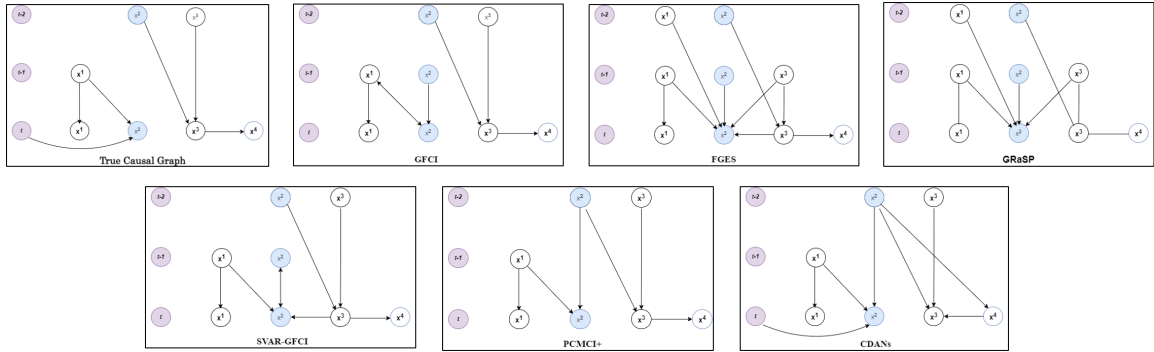


Figure 3: Estimated causal graphs for a 4-variable synthetic model with a 2-lag period, comparing CDANs with other algorithms. CDANs exhibits remarkable accuracy by correctly identifying all six causal edges, including the changing module. However, one of the identified edges (from x^4 to x^3) has the opposite direction compared to the true causal graph and CDANs falsely identifies 2 causal edges. In contrast, GFCI accurately identifies five edges but suggests one incorrect hidden confounder. FGES provides four correct edges but includes five incorrect ones. GRaSP only detects two causal edges and includes six erroneous or un-oriented edges. SVAR-GFCI correctly detects five edges but includes two false edges. PCMCI+ accurately identifies five edges but has one incorrect edge.

the changing modules. The data generation process, including the criteria used for creating the multivariate time series models, is available in Appendix C.1.

6.1.2. REAL-WORLD CLINICAL DATASET

We evaluate our method and compare the results with other approaches on a clinical dataset based on oxygen therapy for ICU patients collected from the MIMIC-III (Johnson et al., 2016) database. We collected time series data for ICU patients who received either conservative or liberal oxygenation. We extracted 12 variables by following the study protocol described in Panwar et al. (2016), Gani et al. (2023), and Bikak et al. (2020). Data were recorded every 4 hours for the 12 variables which are as follows: fraction of inspired oxygen (FiO_2), hemoglobin, lactate, partial pressure of carbon dioxide ($PaCO_2$), partial pressure of oxygen (PaO_2), arterial oxygen saturation (SaO_2), peripheral oxygen saturation (SpO_2), minute ventilation volume ($vent$), peak air pressure ($airpr$), positive end-expiratory pressure ($PEEP$), potential of hydrogen (pH), and tidal volume (VT). We considered the values of these variables for up to 2 weeks and estimated the causal structures in the case of both conservative and liberal oxygen therapies. The cohort of this study is described in Section 5.

Table 1: Performance metrics of different algorithms for 4, 6, and 8 variables with lag periods 2, 4, 6, and 8.

	Lag 2			Lag 4			Lag 6			Lag 8			
	TPR ¹	FDR ²	SHD ³	TPR	FDR	SHD	TPR	FDR	SHD	TPR	FDR	SHD	
4 variables	PCMCI+	0.83	0.17	2	0.67	0.43	5	0.83	0.29	3	0.83	0.29	3
	GFCI	0.67	0.43	5	0.67	0.33	4	0.67	0.43	5	0.83	0.38	4
	FGES	0.67	0.56	7	0.67	0.50	6	0.83	0.44	5	0.83	0.44	5
	GraSP	0.33	0.75	10	0.50	0.57	7	0.67	0.50	6	0.67	0.56	7
	SVAR-GFCI	0.83	0.29	3	0.67	0.20	3	0.83	0.38	4	0.50	0.57	7
	CDANs	0.83	0.38	4	1.00	0.54	7	1.00	0.25	2	0.80	0.60	7
6 variables	PCMCI+	0.78	0.30	5	0.67	0.40	7	0.56	0.44	8	0.67	0.40	7
	GFCI	0.67	0.45	8	0.67	0.40	4	0.67	0.57	11	0.67	0.54	10
	FGES	0.67	0.60	12	0.67	0.57	5	0.67	0.63	13	0.67	0.63	13
	GraSP	0.56	0.64	13	0.56	0.64	7	0.56	0.64	13	0.56	0.69	15
	SVAR-GFCI	0.67	0.45	8	0.56	0.38	7	0.56	0.55	10	0.44	0.64	12
	CDANs	0.89	0.47	8	0.89	0.43	7	0.73	0.38	7	0.89	0.43	7
8 variables	PCMCI+	0.73	0.33	7	0.55	0.14	6	0.55	0.40	11	0.73	0.43	9
	GFCI	0.73	0.47	10	0.55	0.57	13	0.73	0.43	9	0.73	0.47	10
	FGES	0.73	0.58	14	0.55	0.67	17	0.73	0.50	11	0.73	0.56	13
	GraSP	0.64	0.61	15	0.55	0.65	16	0.64	0.56	13	0.64	0.61	15
	SVAR-GFCI	0.73	0.43	9	0.55	0.45	10	0.64	0.42	9	0.55	0.54	12
	CDANs	0.82	0.53	12	0.73	0.53	12	0.82	0.53	12	0.73	0.58	14

6.2. Evaluation

6.2.1. PERFORMANCE ON SYNTHETIC DATA

We compare the performance of our approach, CDANs, with several baseline methods, including PCMCI+, SVAR-GFCI, GRaSP, FGES, and CD-NOD. We use synthetic datasets consisting of 4, 6, and 8 variables with different lags (2, 4, 6, and 8). Table 1 reports the performance metrics of all the algorithms on the synthetic datasets. The experimental findings demonstrate the superiority of CDANs in various settings, particularly in terms of the true positive rate (TPR). For the 4-variable model with lag 2, CDANs, PCMCI+, and SVAR-GFCI, correctly identifies 5 out of 6 edges ground truth edges (Figure 3). However, both PCMCI+ and SVAR-GFCI detect some spurious edges whereas CDANs do not produce any false edges. In some of the other cases, PCMCI+ performs well in terms of FDR and SHD. GFCI, FGES, GraSP, and SVAR-GFCI have a moderate performance in the case of the different settings. Particularly, GRaSP struggles in performance across all settings. On the contrary, CDANs consistently performs well even when the number of lags increases, identifying the maximum number of correct edges and having the highest TPR compared to all approaches. In the case of the 4 and 6 variable models with lag 6, CDANs outperforms others with respect to all the metrics. As the mentioned baseline approaches are not capable of detecting changing modules, we further compare CDANs with CD-NOD. Performance metrics of CDANs and CD-NOD are presented in Table 2, considering only contemporaneous variables and changing modules instead of the full causal graph as CD-NOD cannot detect

¹True Positive Rate, higher is better

²False Discovery Rate, lower is better

³Structural Hamming Distance, lower is better

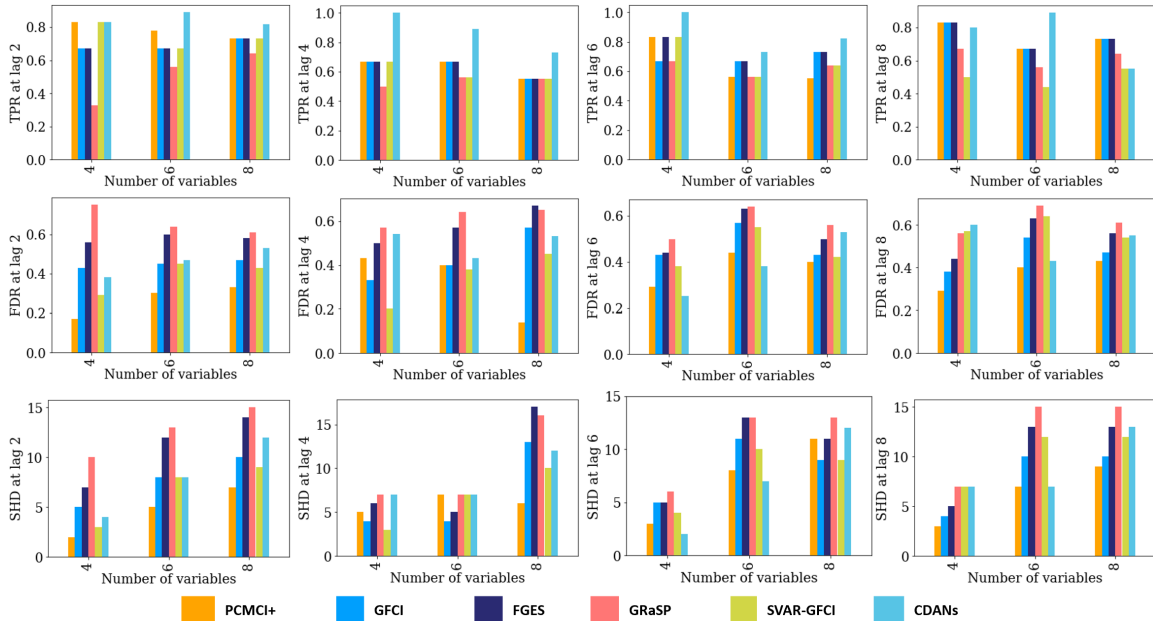


Figure 4: Performance comparison of PCMCi+, GFCI, FGES, GRaSP, SVAR-GFCI, and CDANs on synthetic datasets with 4, 6, and 8 variables and lag values of 2, 4, 6, and 8. CDANs consistently outperforms other approaches in terms of TPR across all scenarios, showcasing its excellence in identifying causal edges. SVAR-GFCI performs notably well for lag 2 in terms of FDR, while CDANs performs best for lag 6. GRaSP exhibits the lowest performance among all methods. FGES and GRaSP have the highest SHD, while CDANs and the other approaches exhibit comparable performance in different settings. Overall, the results highlight the effectiveness of CDANs in various contexts.

lagged edges. CDANs outperforms CD-NOD in all cases, having a lower FDR in two cases and a lower SHD in all cases. This highlights the importance of considering lagged confounders during temporal causal discovery. Since their ignorance by CD-NOD leads to higher FDR and SHD due to the emergence of false causal edges among the contemporaneous variables.

Table 2: Performance of CDANs and CD-NOD for the synthetic datasets. The results highlight CDANs’ consistent effectiveness in identifying true causal relationships and maintaining a lower FDR and SHD compared to CD-NOD, particularly for datasets with 6 and 8 variables. This comparison shows the efficiency of CDANs for a better causal discovery across datasets of variable sizes.

	4 variables		6 variables		8 variables	
	CDANs	CD-NOD	CDANs	CD-NOD	CDANs	CD-NOD
TPR	1	1	0.75	0.75	0.43	0.43
FDR	0.50	0.50	0.50	0.57	0.57	0.67
SHD	1	2	4	5	8	10

6.2.2. PERFORMANCE ON REAL-WORLD DATA

We report here the performance of CDANs and the baseline methods on the real-world clinical dataset having 12 variables. The estimated causal graph for 12 variables with a lag period of 2 is provided in Appendix C. We use the non-temporal causal graph of these variables developed by Gani et al. (2023) as a reference for evaluation since a ground truth temporal causal graph is unavailable. Remarkably, CDANs identifies FiO_2 and vt as changing modules, while CD-NOD detects FiO_2 and $PaCO_2$ as changing modules. The other approaches fail to identify any changing modules. Furthermore, CDANs recognizes vt as an autocorrelated variable and identifies two lagged causal edges: $hemo \rightarrow airpressure$ and $peep \rightarrow fio2$ with lag 1, and $pH \rightarrow lactate$ with lag 2. In comparison, PCMCI+ detects only two lagged causal variables: FiO_2 with lag 2 and vt with lag 1. In the case of the contemporaneous edges, CD-NOD estimates six of them and two undirected edges (Figure 7). However, none of the causal edges estimated by CD-NOD match the non-temporal graph. Other methods generate much denser causal graphs with less explainability. GFCE, FGES, and GRASP struggle to differentiate between true and false causal edges, identifying causal graphs with 37, 43, and 40 edges, respectively (Figure 8), which are not consistent with the non-temporal graph. PCMCI+ identifies three contemporaneous edges with one undirected edge. In contrast, CDANs outperforms all as it discovers four contemporaneous causal edges and one undirected causal edge, offering a more accurate and interpretable representation of the causal relationships based on the existing non-temporal causal graph. In fact, CDANs estimated an undirected edge between pH and $PaCO_2$, which is present in the non-temporal ground truth graph ($pH \rightarrow PaCO_2$). This highlights the better performance of CDANs over other methods. Moreover, CDANs identifies a lagged causal edge of lag 2 from pH to $lactate$, whereas the non-temporal graph has a causal edge from pH to $lactate$ through $PaCO_2$ and $APSII$. The non-temporal true causal graph and estimated causal graphs of these methods are given in Appendix C.

Table 3: Comparison of causal skeletons identified using PCMCI+, CD-NOD, and CDANs on the real clinical dataset. CDANs successfully detects 3 causal edges, outperforming PCMCI+ and CD-NOD, which each identifies only a single causal edge.

Causal edges	PCMCI+	CD-NOD	CDANs
PEEP to FiO2	-	-	Identified
PEEP to paO2	-	-	-
PEEP to SaO2	-	-	-
pH to paCO2	-	-	Undirected
pH to paO2	Undirected	Undirected	Undirected
SpO2 to FiO2	-	-	-
SpO2 to paO2	-	-	-
SpO2 to Hemoglobin	-	-	-
Air Pressure to paO2	-	-	-
Air Pressure to SaO2	-	-	-
Air Pressure to Lactate	-	-	-

Table 4: Performance comparison of PCMCI+, CD-NOD, and CDANs on the clinical dataset showing the TPR, FDR, and SHD for each method. CDANs achieves the highest TPR and lowest FDR and SHD values, outperforming others.

	PCMCI+	CD-NOD	CDANs
TPR	0.09	0.09	0.27
FDR	0.8	0.83	0.63
SHD	14	15	13

We also compare the ability of PCMCI+, CD-NOD, and CDANs to detect the causal skeleton without considering the time lag (Table 4). To obtain the causal skeleton from each approach, we transform the identified causal graphs into non-temporal versions without causal directions. Here, CDANs identifies three causal edges with the highest TPR compared to the other approaches. Both PCMCI+, and CD-NOD identify only one undirected causal edge (Table 4). CDANs achieves the best TPR of 0.27, compared to PCMCI+ and CD-NOD with poor TPR values of 0.09 each. CDANs also has the lowest FDR of 0.63 and lowest SHD of 13, compared to PCMCI+ and CD-NOD. The higher TPR and lower FDR and SHD values of CDANs suggests its efficiency in inferring causal relationships from complex real-world time series data.

7. Discussion

In this study, we present a novel temporal causal discovery approach, CDANs for non-stationary and autocorrelated time series data. The method utilizes the momentary conditional independence (MCI) test to detect lagged causal relationships, and enables the efficient identification of changing modules and contemporaneous causal edges by the inclusion of lagged parents in the conditioning set. The existing approaches lack the ability to detect all of the different types of temporal causal edges (contemporaneous and lagged) along with changing modules. Experimental results on synthetic datasets containing 4, 6, and 8 variables with different time lags of 2, 4, 6, and 8, as well as on an important clinical dataset related to oxygen therapy in ICU comprising of 12-time series variables demonstrate the effectiveness of our approach. We compare the performance of CDANs with six existing baselines where our approach excels in identifying contemporaneous, autocorrelated, lagged causal relationships, and changing modules while maintaining a higher true positive rate (TPR) and lower false discovery rate (FDR) in high-dimensional settings.

Limitations One limitation of our study is the usage of conventional independence tests. Future research may explore recent conditional independence tests such as the classifier-based Conditional Mutual Information (Mukherjee et al., 2020) or the Generative Conditional Independence Test (Bellot and van der Schaar, 2019). Additionally, estimating contemporaneous parents alongside lagged parents may further enhance performance (Ferdous et al., 2023). Also, evaluating CDANs on large-scale data to see how it performs in those settings can be explored.

Acknowledgments

We sincerely thank the anonymous reviewers and area chair for their valuable feedback, which greatly contributed to the enhancement of this study. This research received partial support from the National Science Foundation (NSF Award 2118285) and the UMBC Strategic Awards for Research Transitions (START). The views expressed in this work do not necessarily reflect the policies of the NSF, and endorsement by the Federal Government should not be inferred.

References

- Neill KJ Adhikari, Robert A Fowler, Satish Bhagwanjee, and Gordon D Rubenfeld. Critical care and the global burden of critical illness in adults. *The Lancet*, 376:1339–1346, 2010. doi: 10.1016/S0140-6736(10)60446-1. PMID: 24171518.
- Riddhiman Adib, Md Mobasshir Arshed Naved, Chih-Hao Fang, Md Osman Gani, Ananth Grama, Paul Griffin, Sheikh Iqbal Ahamed, and Mohammad Adibuzzaman. Ckh: Causal knowledge hierarchy for estimating structural causal models from data and priors. *arXiv preprint arXiv:2204.13775*, 2022.
- Iyad Batal, Hamed Valizadegan, Gregory F. Cooper, and Milos Hauskrecht. A temporal pattern mining approach for classifying electronic health record data. *ACM Transactions on Intelligent Systems and Technology*, 8(1):1–19, 2016.
- Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.
- Alexis Bellot and Mihaela van der Schaar. Conditional independence testing using generative adversarial networks. *Advances in Neural Information Processing Systems*, 32, 2019.
- James R Bence. Analysis of short time series: correcting for autocorrelation. *Ecology*, 76(2):628–639, 1995.
- Marvi Bikak, Shravan Kethireddy, Md Osman Gani, and Mohammad Adibuzzaman. Structural causal model with expert augmented knowledge to estimate the effect of oxygen therapy. *CHEST*, 158(4):A636, 2020.
- Bart Bussmann, Jannes Nys, and Steven Latré. Neural additive vector autoregression models for causal discovery in time series. In *International Conference on Discovery Science*, pages 446–460. Springer, 2021.
- Joan A Casey, Brian S Schwartz, Walter F Stewart, and Nancy E Adler. Using electronic health records for population health research: a review of methods and applications. *Annual review of public health*, 37:61–81, 2016.
- David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Nov):507–554, 2002.
- Tianjiao Chu, Clark Glymour, and Greg Ridgeway. Search for additive nonlinear time series causal models. *Journal of Machine Learning Research*, 9(5), 2008.

- Matthew M. Churpek, Trevor C. Yuen, Christopher Winslow, David O. Meltzer, Michael W. Kattan, and Dana P. Edelson. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Critical Care Medicine*, 44(2):368–374, 2016.
- David A. Clifton, Daniel Wong, Lei Clifton, Stephanie Wilson, Rebecca Way, Roy Pullinger, and Lionel Tarassenko. A large-scale clinical validation of an integrated monitoring system in the emergency department. *IEEE Journal of Biomedical and Health Informatics*, 16(3):471–477, 2012.
- Diego Colombo, Marloes H Maathuis, Markus Kalisch, and Thomas S Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pages 294–321, 2012.
- Diego Colombo, Marloes H Maathuis, et al. Order-independent constraint-based causal structure learning. *J. Mach. Learn. Res.*, 15(1):3741–3782, 2014.
- Martin R Cowie, Juuso I Blomster, Lesley H Curtis, Sylvie Duclaux, Ian Ford, Fleur Fritz, Samantha Goldman, Salim Janmohamed, Jörg Kreuzer, Mark Leenay, et al. Electronic health records to facilitate clinical research. *Clinical Research in Cardiology*, 106(1):1–9, 2017.
- Doris Entner and Patrik O Hoyer. On causal discovery from time series data using fci. *Probabilistic graphical models*, pages 121–128, 2010.
- Muhammad Hasan Ferdous, Uzma Hasan, and Md Osman Gani. ecdans: Efficient temporal causal discovery from autocorrelated and non-stationary data (student abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(13):16208–16209, Jun. 2023. doi: 10.1609/aaai.v37i13.26964. URL <https://ojs.aaai.org/index.php/AAAI/article/view/26964>.
- Md Osman Gani, Shravan Kethireddy, Riddhiman Adib, Uzma Hasan, Paul Griffin, and Mohammad Adibuzzaman. Structural causal model with expert augmented knowledge to estimate the effect of oxygen therapy on mortality in the icu. *Artificial Intelligence in Medicine*, 137:102493, 2023.
- Madelyn Glymour, Judea Pearl, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.
- Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.
- Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer, 2005.
- Edward J Hannan. The estimation of a lagged regression relation. *Biometrika*, 54(3-4):409–418, 1967.

- Uzma Hasan and Md Osman Gani. Kcrl: A prior knowledge based causal discovery framework with reinforcement learning. *Proceedings of Machine Learning Research*, 182(2022): 1–24, 2022.
- Uzma Hasan, Emam Hossain, and Md Osman Gani. A survey on causal discovery methods for temporal and non-temporal data. *arXiv preprint arXiv:2303.15027*, 2023. Under review in TMLR: <https://openreview.net/forum?id=YdMrdhGx9y>.
- Biwei Huang, Kun Zhang, Jiji Zhang, Joseph D Ramsey, Ruben Sanchez-Romero, Clark Glymour, and Bernhard Schölkopf. Causal discovery from heterogeneous nonstationary data. *J. Mach. Learn. Res.*, 21(89):1–53, 2020.
- J. Larry Jameson and Dan L. Longo. Precision medicine—personalized, problematic, and promising. *New England Journal of Medicine*, 372(23):2229–2234, 2015.
- Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(160035), 2016. doi: <https://doi.org/10.1038/sdata.2016.35>.
- Marcus Kaiser and Maksim Sipos. Unsuitability of notears for causal graph discovery when dealing with dimensional quantities. *Neural Processing Letters*, pages 1–9, 2022.
- Wai-Yin Lam, Bryan Andrews, and Joseph Ramsey. Greedy relaxations of the sparsest permutation algorithm. In *The 38th Conference on Uncertainty in Artificial Intelligence*, 2022.
- Richard Lawton et al. Time series analysis and its applications: Robert h. shumway and david s. stoffer; springer texts in statistics; 2000, springer-verlag. *International Journal of Forecasting*, 17(2):299–301, 2001.
- Thuc Duy Le, Tao Hoang, Jiuyong Li, Lin Liu, Huawen Liu, and Shu Hu. A fast pc algorithm for high dimensional causal discovery with multi-core pcs. *IEEE/ACM transactions on computational biology and bioinformatics*, 16(5):1483–1495, 2016.
- Daniel Malinsky and Peter Spirtes. Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of 2018 ACM SIGKDD Workshop on Causal Discovery*, pages 23–47. PMLR, 2018.
- Mariusz Maziarz. A review of the granger-causality fallacy. *The journal of philosophical economics: Reflections on economic and social issues*, 8(2):86–105, 2015.
- Maureen O. Meade, Deborah J. Cook, Gordon H. Guyatt, Arthur S. Slutsky, Yaseen M. Arabi, D. James Cooper, Andrew R. Davies, Lori E. Hand, Qi Zhou, Lehana Thabane, Peggy Austin, Stephen Lapinsky, Alan Baxter, James Russell, Yoanna Skrobik, Juan J. Ronco, Thomas E. Stewart, and for the Lung Open Ventilation Study Investigators. Ventilation Strategy Using Low Tidal Volumes, Recruitment Maneuvers, and High Positive End-Expiratory Pressure for Acute Lung Injury and Acute Respiratory Distress Syndrome: A Randomized Controlled Trial. *JAMA*, 299(6):637–645, 02 2008. ISSN 0098-7484. doi: 10.1001/jama.299.6.637. URL <https://doi.org/10.1001/jama.299.6.637>.

- Christopher Meek. Causal inference and causal explanation with background knowledge. *arXiv preprint arXiv:1302.4972*, 2013.
- Sudipto Mukherjee, Himanshu Asnani, and Sreeram Kannan. Ccmi: Classifier based conditional mutual information estimation. In *Uncertainty in artificial intelligence*, pages 1083–1093. PMLR, 2020.
- Shamim Nemati, Andre Holder, Fereshteh Razmi, Matthew D. Stanley, Gari D. Clifford, and Timothy G. Buchman. An interpretable machine learning model for accurate prediction of sepsis in the icu. *Critical Care Medicine*, 46(4):547–553, 2016.
- Amy Harris Nordo, Hugh P Levaux, Lauren B Becnel, Jose Galvez, Prasanna Rao, Komathi Stem, Era Prakash, and Rebecca Daniels Kush. Use of ehra data for clinical research: Historical progress and current applications. *Learning health systems*, 3(1):e10076, 2019.
- Mohammad Norouzi, David J Fleet, and Russ R Salakhutdinov. Hamming distance metric learning. In *Advances in neural information processing systems*, pages 1061–1069, 2012.
- Juan Miguel Ogarrio, Peter Spirtes, and Joe Ramsey. A hybrid causal search algorithm for latent variable models. In *Conference on probabilistic graphical models*, pages 368–379. PMLR, 2016.
- Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. Dynotears: Structure learning from time-series data. In *International Conference on Artificial Intelligence and Statistics*, pages 1595–1605. PMLR, 2020.
- Dr R. Panwar, G. Capellier, N. Schmutz, A. Davies, D. J. Cooper, M. Bailey, D. Baguley, D. V. Pilcher, and R. Bellomo. Current oxygenation practice in ventilated patients—an observational cohort study. *Anaesthesia and Intensive Care*, 41(4):505–514, 2013. doi: 10.1177/0310057X1304100412. URL <https://doi.org/10.1177/0310057X1304100412>. PMID: 23808511.
- Rakshit Panwar, Miranda Hardie, Rinaldo Bellomo, Loïc Barrot, Glenn M Eastwood, Paul J Young, Gilles Capellier, Peter WJ Harrigan, and Michael Bailey. Conservative versus liberal oxygenation targets for mechanically ventilated patients. a pilot multicenter randomized controlled trial. *American journal of respiratory and critical care medicine*, 193(1):43–51, 2016.
- Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.
- Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M Dai, Nissan Hajaj, Michaela Hardt, Peter J Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, et al. Scalable and accurate deep learning with electronic health records. *NPJ digital medicine*, 1(1):18, 2018.
- Joseph Ramsey, Madelyn Glymour, Ruben Sanchez-Romero, and Clark Glymour. A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *International journal of data science and analytics*, 3(2):121–129, 2017.

- Alexander Reisach, Christof Seiler, and Sebastian Weichwald. Beware of the simulated dag! causal discovery benchmarks may be easy to game. *Advances in Neural Information Processing Systems*, 34:27772–27784, 2021.
- Jakob Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on Uncertainty in Artificial Intelligence*, pages 1388–1397. PMLR, 2020.
- Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marí, et al. Inferring causation from time series in earth system sciences. *Nature communications*, 10(1):1–13, 2019a.
- Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11):eaau4996, 2019b.
- G Shari, M Kojicic, G Li, et al. The acute respiratory distress syndrome network. *N Engl J Med*, 342:1301–1308, 2000.
- Christopher A Sims. Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pages 1–48, 1980.
- Liam Solus, Yuhao Wang, and Caroline Uhler. Consistency guarantees for greedy permutation-based causal inference algorithms. *Biometrika*, 108(4):795–814, 2021.
- Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2000.
- Xiangyu Sun, Guiliang Liu, Pascal Poupart, and Oliver Schulte. Nts-notears: Learning nonparametric temporal dags with time-series data and prior knowledge. *arXiv preprint arXiv:2109.04286*, 2021.
- Jean-Louis Vincent and Daniel De Backer. Circulatory shock. *New England Journal of Medicine*, 369(18):1726–1734, 2013.
- Hannah Wunsch, Walter T. Linde-Zwirble, Derek C. Angus, Mary E. Hartman, Eric B. Milbrandt, and Jeremy M. Kahn. The epidemiology of mechanical ventilation use in the united states. *Critical Care Medicine — Society of Critical Care Medicine*, 38(10):1947–1953, 2010. doi: 10.1097/CCM.0b013e3181ef4460. PMID: 24171518.
- Kun Zhang, Jonas Peters, Dominik Janzing, and Bernhard Scholkopf. Kernel-based conditional independence test and application in causal discovery. *arXiv preprint arXiv:1202.3775*, 2012.
- Kun Zhang, Biwei Huang, Jiji Zhang, Clark Glymour, and Bernhard Scholkopf. Causal discovery from nonstationary/heterogeneous data: Skeleton estimation and orientation determination. In *IJCAI: Proceedings of the Conference*, volume 2017, page 1347. NIH Public Access, 2017.

Xun Zheng, Bryon Aragam, Pradeep Ravikumar, and Eric P Xing. Dags with no tears: Continuous optimization for structure learning. *arXiv preprint arXiv:1803.01422*, 2018.

Appendix A. Metrics Details

The performance of a causal discovery approach depends on how accurately it identifies the true causal edges, the proportion of correctly identified edges, and how closely it resembles the true causal graph.

True Positive Rate (TPR), also known as Sensitivity or Recall, measures the ability of the model to accurately identify the causal edges. In the case of causal discovery, TPR is defined as follows:

$$TPR = \frac{TP}{TP + FN}$$

Here, TP (true positive) represents the total number of correctly identified causal edges and FN (false negative) denotes the total number of unidentified causal edges.

TPR alone is not sufficient to measure the performance of a model because it primarily focuses on correctly identified positive cases, and can give an impressive result even with many false edges. For this reason, False Discovery Rate (FDR) is used in conjunction with TPR because FDR considers both the number of correctly identified causal edges and the number of incorrectly identified causal edges. FDR is defined as follows:

$$FDR = \frac{FP}{FP + TP}$$

Here, FP (false positive) represents the total number of wrongly identified directed edges.

Structural Hamming Distance (SHD) is another metric used to evaluate the difference between a true causal graph and an estimated causal graph. SHD measures the number of operations (edge addition, removal, or reverse) required to convert an estimated DAG into its ground-truth causal graph. That is, it counts the total number of edge insertions, deletions, or flips required to transform the generated causal graph into the true causal graph (Norouzi et al., 2012). In general, a low SHD score indicates high similarity between true and estimated causal graphs, while a high SHD score indicates low similarity.

Appendix B. Causal Direction

CDANs identifies causal direction in 2 steps- lagged causal direction and contemporaneous causal direction. According to the time flow, CDANs first orients lagged edges from past to present and then orients contemporaneous variables. There are two types of contemporaneous variables in CDANs – surrogate variable and other variables in the model. As C is a surrogate variable for the unobserved confounders which is one of the causes of the changing modules, we can then orient (C, X_t^j) as $C \rightarrow X_t^j$ if X_t^j is adjacent to C . We then consider the triples of the form $(C - X_t^i - X_t^j)$ and use the conditional sets of the CI test, from step 3, between C and X_t^j to determine the direction. If the conditioning set does not include X_t^i , it orients the triple as $C \rightarrow X_t^i \leftarrow X_t^j$. Otherwise, orients as $C \rightarrow X_t^i \rightarrow X_t^j$. If both X_t^i and X_t^j are adjacent to C , then the causal direction between X_t^i and X_t^j is determined based on the causal effect from X_t^i to X_t^j and vice versa. We can calculate the causal effect for a given pair of variables (X_t^i, X_t^j) ; here X_t^i and X_t^j are independent if one of $P(X_t^i)$ and

$P(X_t^i | X_t^j)$ changes while the other remains invariant. We determine the causal direction as $X_t^i \rightarrow X_t^j$ if $P(X_t^i)$ and $P(X_t^j | X_t^i)$ are independent but $P(X_t^j)$ and $P(X_t^i | X_t^j)$ are dependent. We use an extended version of Hilbert Schmidt Independence Criterion (HSIC) (Gretton et al., 2005) to measure the dependence between $P(X_t^j | X_t^i)$ and $P(X_t^i)$, denoted by $\partial_{X_t^i \rightarrow X_t^j}$, and the dependence between $P(X_t^i | X_t^j)$ and $P(X_t^j)$, denoted by $\partial_{X_t^j \rightarrow X_t^i}$.

For a pair of variables (X, Y) , the dependence between $P(Y | X)$ and $P(X)$ is measured by the following equation proposed by (Huang et al., 2020) where G_X and $G_{Y|X}$ are the Gram matrix of $\widehat{\mu}_{X|C}$ and $\widehat{\mu}_{Y|X|C}$ at $C = c_1, c_2, \dots, c_N$ and H is the center of the features.

$$\partial_{X \rightarrow Y} = \frac{\text{tr}(G_X H G_{Y|X} H)}{\text{tr}(G_X H) \text{tr}(G_{Y|X} H)}$$

Based on the dependencies between X_t^i and X_t^j , we can orient (X_t^i, X_t^j) as $(X_t^i \rightarrow X_t^j)$ if $\partial_{X_t^i \rightarrow X_t^j} < \partial_{X_t^j \rightarrow X_t^i}$. Otherwise, orient as $(X_t^j \rightarrow X_t^i)$.

Appendix C. Data generation, and Graphs

C.1. Synthetic Data Generation

To generate the dataset, we consider the first variable (X^1) to be an autocorrelated variable without any contemporaneous dependency. The second variable (X^2) has a lagged dependency on the first variable and has a time dependency. The third variable (X^3) has autocorrelation with period 2 and lagged dependency of period 2 on the second variable (X^2). Finally, the fourth variable (X^4) has contemporaneous dependency on the third variable (X^3). Data for 4 variable model with lag 2 is generated using the following equations:

$$\begin{aligned} X_t^1 &= 0.6 * X_{t-1}^1 + \varepsilon_t^1 \\ X_t^2 &= 0.8 * X_{t-1}^1 + 1.5 * \sin\left(\frac{t}{50}\right) + \varepsilon_t^2 \\ X_t^3 &= 0.7 * X_{t-2}^2 + 0.5 * X_{t-2}^3 + \varepsilon_t^3 \\ X_t^4 &= 0.6 * X_t^3 + \varepsilon_t^4 \end{aligned}$$

We then increased the number of variables to 6, and 8 with 10, and 12 causal links. The data-generating process of four variables remains the same and the rest variables are generated by the below equations:

$$\begin{aligned} X_t^5 &= 0.8 * X_{t-2}^4 + 0.8 * \sin\left(\frac{t}{20}\right) + \varepsilon_t^5 \\ X_t^6 &= 0.7 * X_t^5 + \varepsilon_t^6 \\ X_t^7 &= 0.4 * X_{t-1}^6 + \varepsilon_t^7 \\ X_t^8 &= 0.6 * X_t^7 + \varepsilon_t^8 \end{aligned}$$

For the four-variable models, we consider lag periods of 4, 6, and 8, and modify the lag period between X^2 and X^3 to 2, 4, 6, and 8, respectively. In the six and eight-variable models, we adjust the lag between X^4 and X^5 to maintain sparsity and accommodate the

maximum lag period. Consequently, the four-variable models contain one changing module, while the six and eight-variable models feature two changing modules.

To introduce non-linearity in the variables, we employ sine and cosine functions. The 8-variable model with a lag of 8 is depicted in Figure 5, demonstrating the complex interactions and relationships among the variables across different time steps.

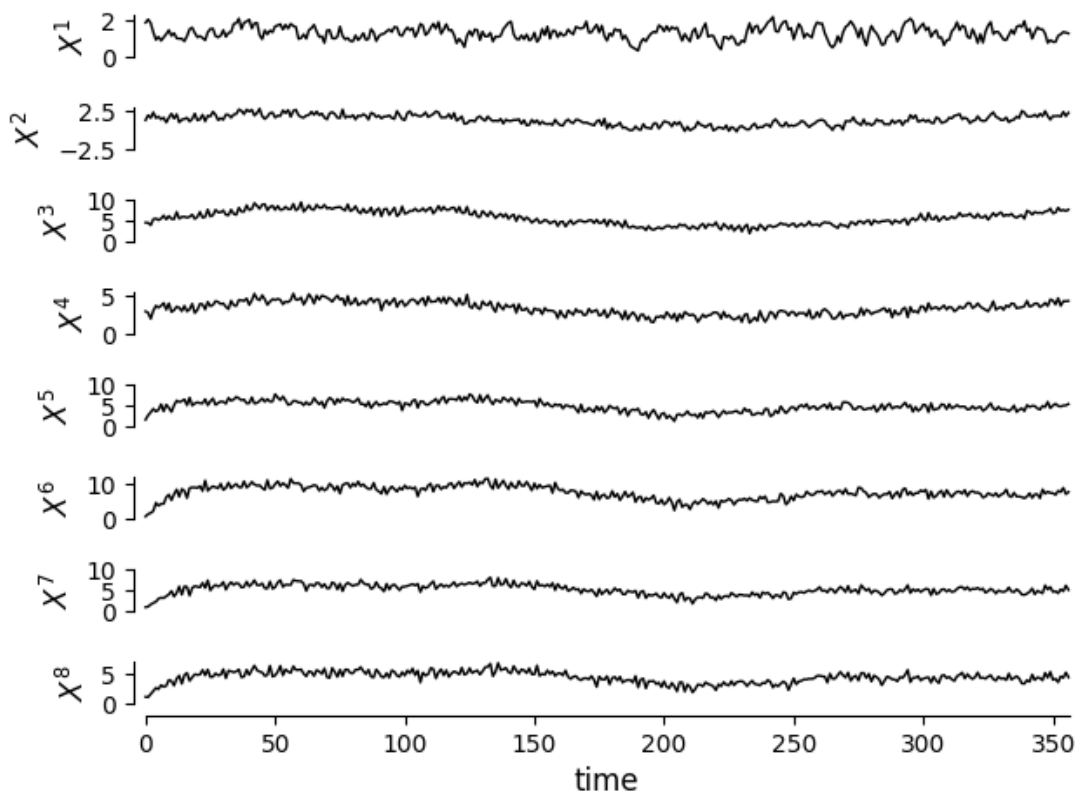


Figure 5: Visualization of a multivariate time series consisting of 8 variables, with each variable plotted over time (x-axis) and corresponding values (y-axis). The time-lagged relationships are highlighted up to a maximum lag of 8, illustrating the potential causal influences between the variables at different time steps.

C.2. Causal Graphs

This section explores the different causal graphs associated with a real-life oxygen therapy dataset. We follow a study protocol described in [Panwar et al. \(2013\)](#) to extract 12 time-series variables from the MIMIC-III database, which are recorded every 4 hours for up to 2 weeks. These variables include a fraction of inspired oxygen, hemoglobin, lactate, partial pressure of carbon dioxide, partial pressure of oxygen, and others. The true causal graph for these temporal settings is unknown, so we use the non-temporal causal graph of the

same variables proposed by Gani et al. (2023). The authors of this study estimate the causal graphs using 7 algorithms and perform majority voting by selecting edges with the highest number of votes. They then incorporate domain knowledge to identify the final causal graph. This study includes 26 variables, however for our study, only the 12 time-series variables are considered, and the remaining 14 non-temporal variables are omitted. The final causal graph with relevant variables is presented in Figure 6.

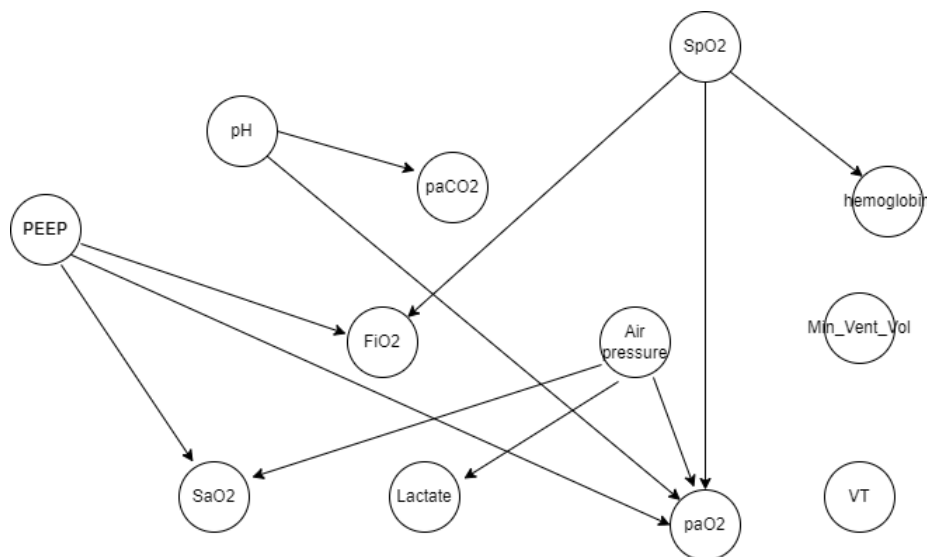


Figure 6: The non-temporal causal graph of an oxygen therapy dataset of ICU patients who received either conservative or liberal oxygenation. The authors estimate causal graphs from 7 algorithms, then perform majority voting by considering edges with the maximum votes. They finally incorporate domain knowledge to estimate the final causal graph. This study omits 14 non-temporal variables and includes 12 time-series variables relevant to the analysis.

As the true causal graph is unknown, we compare all approaches with lag 2. Among the baseline approaches, we only compare the outcomes of PCMCI+ and CD-NOD with CDANs. This is because the other approaches generate denser graphs, which goes against the non-temporal true causal graph. The estimated causal graphs with PCMCI+, CD-NOD, and CDANs are presented in Figure 7.

Recovered the causal graphs of the oxygen therapy dataset using GFCI, FGES, GRaSP, and SVAR-GFCI are presented in Figure 8. We generate all the graphs using the Tetrad package, which allows for the use of numeric variables instead of names. Therefore, we assign numeric values to represent the variables, which are following the previously used variable names. The lagged period is represented by a number followed by a colon sign, for example, 4:1 represents $PaCO_2$ with a lag of 1

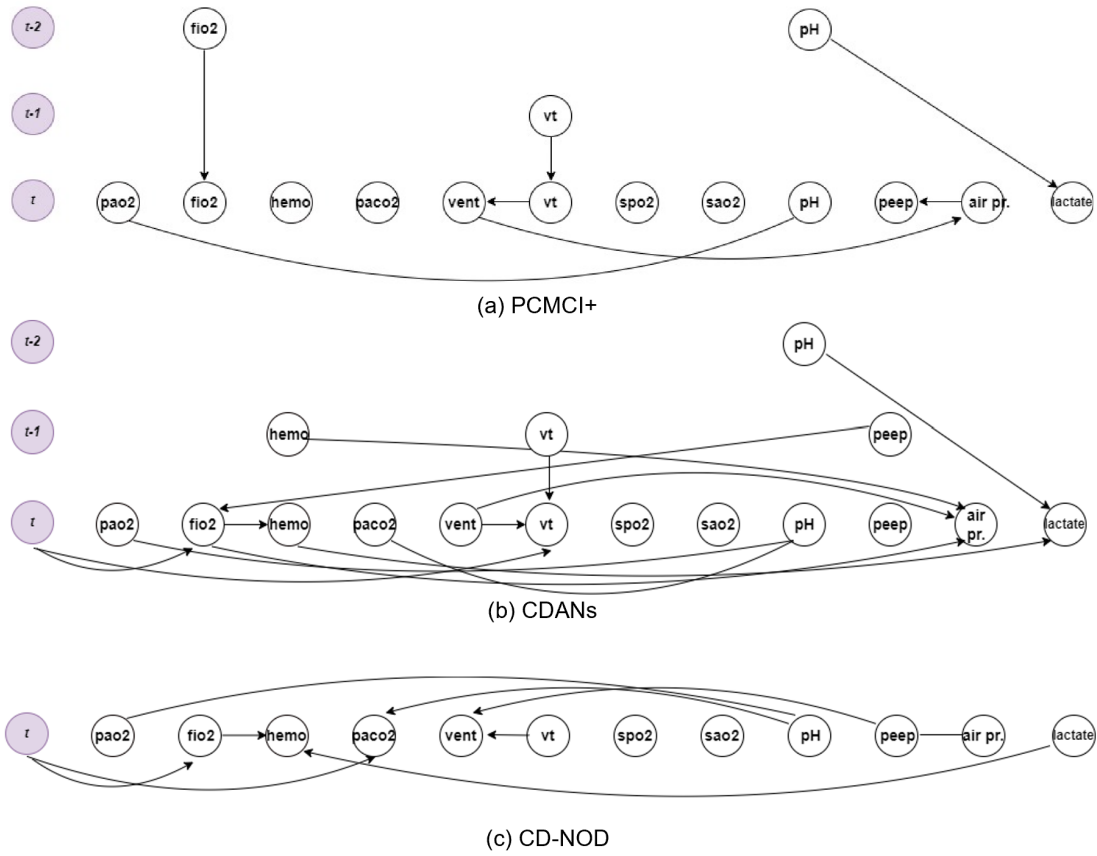


Figure 7: Recovered causal graphs of oxygenation therapy data over 12 time series variables in Intermittent Mandatory Ventilation (IMV) patients. The dataset, collected every 4 hours for 2 weeks, comprises 88 time points for patients receiving IMV. The comparison of algorithms is not possible due to the absence of a true causal graph.

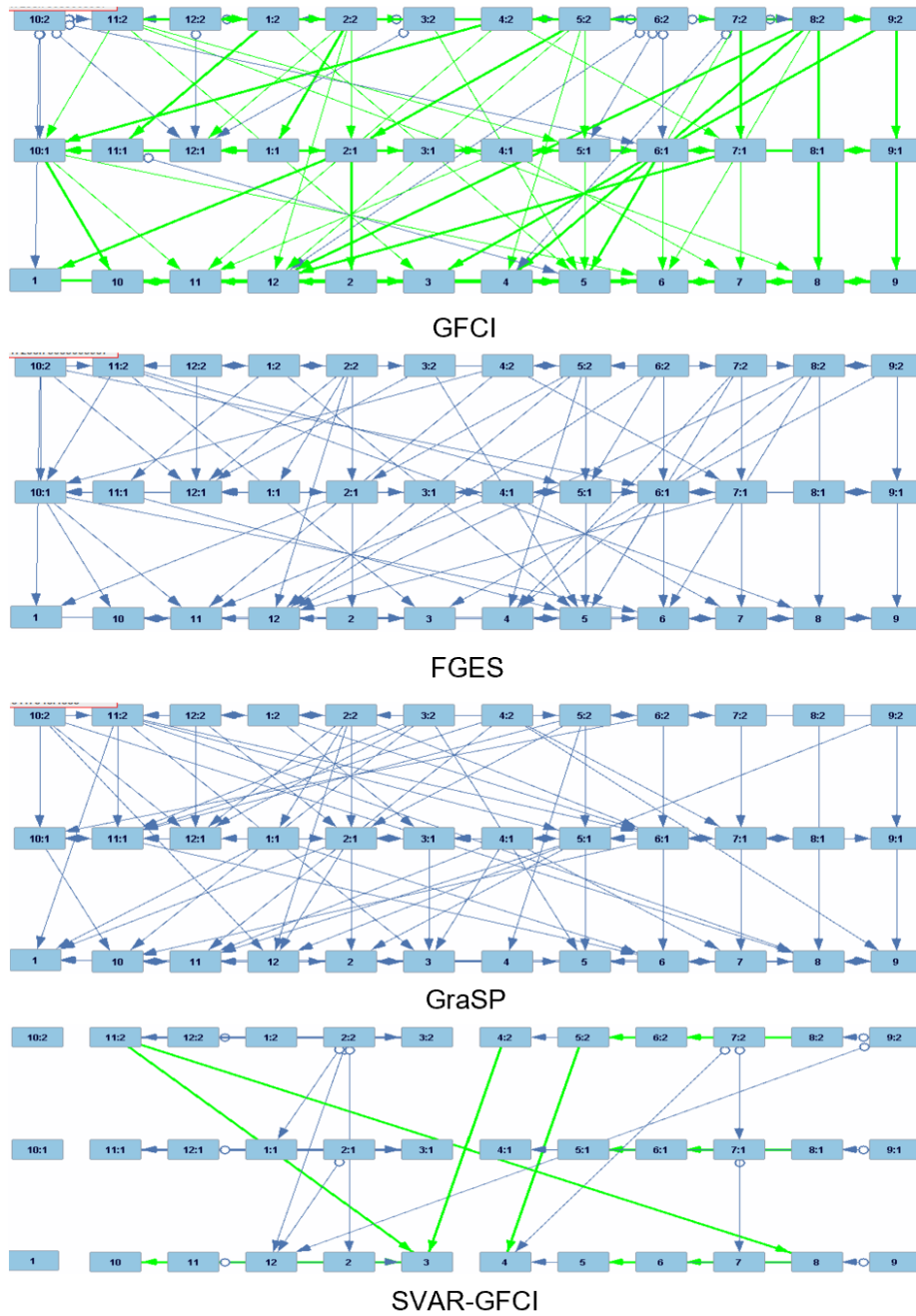


Figure 8: Recovered causal graphs of oxygenation therapy dataset of GFCI, FGES, GRaSP, and SVAR-GFCI algorithms. All graphs are generated using the tetrad package. Here, numeric values are imposed instead of the variable names. Numeric values are given following the variable names used earlier. Also, the lagged period is represented by the number followed by the colon sign. For example 4:1 indicates $PaCO_2$ with lag 1.