ESCOLA
SUPERIOR
DE SAÚDE
POLITÉCNICO
DO PORTO

P.PORTO

M

MESTRADO
EUROPEAN MASTER OF MEDICAL TECHNOLOGY AND HEALTHCARE BUSINESS

Causal Discovery from Time Series Data

Fernanda Ribeiro de Almeida

09/**2022**

P.PORTO

ESCOLA SUPERIOR DE SAÚDE

Causal Discovery from Time Series Data

Author

Fernanda Ribeiro de Almeida

Supervisor(s)

PhD. Luís Coelho ISEP

MSc. Vânia Guimarães Fraunhofer AICOS

PhD. Vitor Rolla Fraunhofer AICOS

Dissertation presented to fulfill the requirements to obtain of Master in **Medical Technology and Healthcare Business** from the Escola Superior de Saúde do Instituto Politécnico do Porto.

Acknowledgements

This thesis is the conclusion of a two-year study within the European Master's in Medical Technology and Healthcare Business. I would like to thank the Polytechnic Institute of Porto, the University of Applied Sciences of Hamburg and the Lille Institute of Health Engineering for the immersive growth experience not only academically but also culturally. My special gratitude to my advisor Luís Coelho for his essential support and cheerful spirit.

With their integral role in this work, I am sincerely grateful to my supervisors Vânia Guimães and Vitor Rolla for introducing me to the fascinating science of causality research and for the outstanding guidance full of valuable lessons throughout this course. I am also deeply grateful to Fraunhofer AICOS for providing a unique scientific research environment and leading people into excellence.

Finally, my acknowledgments go to everyone who has been fundamental in making the master's program an interesting, enriching and delightful experience.

Abstract

The drive to understand the laws that govern the universe and ourselves in order to expand our view of reality is deeply rooted in humanity. In science, this urge is a robust process filled with challenges and opportunities given the rapidly growing technology-driven volume of time series data. Causal discovery supports science in an innovative and fast-growing manner with the essential goal of uncovering mathematical orders directly from observational data translated into *causal* association networks. This scientific tool pledges to accelerate growth in various fields, including life sciences.

This work approaches the topic of causal discovery on two levels. First, we address the theory of *constraint-based methods* on detecting and quantifying causal relations, covering how the methods work, the challenges they face, and the opportunities they present. Second, we explore the PCMCI method with an implementation on both synthetic and real-world data.

The results of this work found in applying causal discovery in real physiological signals data may provide insights into the prospects and difficulties of causal structure search in healthcare Big Data and, moreover, the advantages of using causal models in prediction.

Keywords: Causal discovery, causal relations, time-series data, physiological signals, MIMIC III.

Resumo

A vontade de compreender as leis que governam o universo e a nós próprios, a fim de expandir a nossa visão da realidade está profundamente enraizada na humanidade. Na ciência, esta vontade traduz-se em um processo robusto cheio de desafios e oportunidades, dado o rápido crescimento do volume de dados de séries temporais impulsionado pela tecnologia. A Descoberta Causal apoia a ciência de uma forma inovadora e em rápido crescimento com o objectivo essencial de descobrir ordens matemáticas directamente a partir de dados observacionais que se traduzem em redes de associação *causal*. Esta ferramenta científica promete acelerar o crescimento em vários campos, incluindo as ciências biológicas.

Este trabalho aborda o tema da descoberta causal a dois níveis. Em primeiro lugar, abordamos a teoria dos *constraint-based methods* sobre a detecção e quantificação das relações causais, abrangendo a forma como os métodos funcionam, os desafios que enfrentam, e as oportunidades que apresentam. Além disso, exploramos o método PCMCI com uma implementação em dados tanto sintéticos e em dados do mundo real.

Os resultados deste trabalho encontrados na aplicação da descoberta causal em dados de sinais fisiológicos reais podem fornecer uma visão das perspectivas e dificuldades da pesquisa da estrutura causal em grandes volumes de dados de saúde e, além disso, as vantagens da utilização de modelos causais em previsões.

Palavras-Chave: Descoberta causal, relações causais, séries temporais, sinais fisiológicos, MIMIC III.

Contents

List of Figures

List of Tables

List of Acronyms

CCU	*Coronary Care Unit*
CMIknn	Conditional Mutual Information based on KNN estimator conditional independence test
CSD	*Causal Structure Discovery*
CSRU	*Cardiac Surgery Care Unit*
DAG	*Directed Acyclic Graph*
DBP	*Diastolic Blood Pressure*
GPDC	Gaussian Processes and Distance Correlation conditional independence test
HR	*Heart Rate*
ICU	*Intensive Care Units*
KNN	*K-Nearest Neighbor*
MAP	*Mean Blood Pressure*
MCI	*Momentary Conditional Independence*
MICU	*Medical Intensive Care Unit*
MIMIC	*Medical Information Mart for Intensive Care*
NICU	*Neonatal Intensive Care Unit*
ParCorr	Partial Correlation conditional independence test
RMSE	*Root Mean Squared Error*
RR	*Respiration Rate*
SBP	*Systolic Blood Pressure*
SICU	*Surgical Intensive Care Unit*
SpO2	*Peripheral Oxygen Saturation*
TSICU	*Trauma Surgical Intensive Care Unit*

1. Introduction

1.1 Context and Motivation

The persistent quest to understand the causes behind the phenomena we observe is constantly driving human knowledge and making it possible to answer the questions of "Why". Yet, it is hard to define the terms and notions instinctively embedded into human understanding and language, such as cause and effect. What is the exact implication of stating that X causes Y? Or that smoking causes cancer? It is not that smoking always results in cancer, nor that smoking is always necessary for developing cancer. The implication is that if someone starts to smoke, they will become more likely to develop cancer, and if a smoker quits smoking, they will become less likely to develop cancer. It implies some underlying mechanism connecting the cause (smoking) to the effect (cancer). Regardless of the exact connection, we get a changed effect outcome if we manipulate or intervene in the cause. Instead, if we manipulate or intervene on the affected variable, there is no changed outcome on the causing variable [1].

Science is often confronted with such critical cause-and-effect questions. In order to answer those questions, scientists are challenged with identifying causal relations, the laws or regularities that govern the behavior of observational data [2]. This reasonable effort to explain the causes of things and events is a major goal in science, for possessing this essential knowledge allows humans to intervene in order to cause a desired effect.

Machine learning, deep learning, and all sorts of data analysis methods have decidedly made plenty of our understandings and advances possible. By reason of their exceptional capacity to unveil patterns and correlation and excellent predictive accuracy, there is a significant focus on the scientific community to explore its capacity and make use of it. Unquestionably valuable yet, this strategy provides few insights into the causal mechanisms that govern the dynamics of a system and, therefore, is limited in regards to the ability to predict the outcome of new, previously unseen manipulations or interventions [3].

Possibly, in this limitation lies the challenges of developing what is called Strong AI, which means to develop artificial intelligence with intellectual capability functionally equivalent to a human's. Judea Pearl, one of the fathers of causality, is convinced that a mature causality theory will give rise to a more robust generation of AI capable of counterfactual thinking and that can learn to communicate in the language of cause and effect, as he confidently states in the Book of Why:

> This new generation should explain to us why things happened, why they responded the way they did, and why nature operates one way and not another. More ambitiously, they should also teach us about ourselves: why our mind

> clicks the way it does and what it means to think rationally about cause and effect, credit and regret, intent and responsibility [4].

In various disciplines of science, including healthcare, merely achieving high prediction accuracy in and by itself is commonly not the elemental research question and final goal, it is to discover the causes or mechanism of a system. Moreover, with massive data availability and increased complexity, causal discoveries are usually far from being intuitive and straightforward. In many cases, it is difficult, highly expensive, ethically unacceptable, or perhaps impossible to conduct properly designed interventional experiments from which it would be possible to make causal relations inferences. Clinical research, as an example, is predominantly focused on causal relationships. Hypothesis-driven clinical research frequently assumes a set of causal relationships within a causal structure and estimates the effect magnitude of those relationships (i.e., causal inference). In such a case, drawing a causal conclusion is valid due to prior knowledge reliability that the relationships between variables are indeed causal [2].

However, when there is no knowledge of causality, the causal structure itself has to be discovered. Hence, developing methods to identify causal relationships from purely observational data is desirable; this process is known as causal discovery or causal structure search. Causal discovery is a relatively new science that can infer and quantify potential causal relationships from raw data, without the necessity to intervene in systems, by analyzing characteristics and mathematical properties of purely observational data [5]. Ultimately, finding systems underlying causal relations means the possibility of using them. After identifying the causal structure model of a system, it is possible to estimate the effects of an intervention and causal relations and, thus, condition a system to deliver desired outcomes [3].

It is essential to conceive the difference between inferring causality and inferring correlation. While both explore interdependent relationships of variables in a system, correlation does not imply causation. There are two distinct cases of interdependence. In the case that the system variables evolve in synchrony, with indirect but symmetric connections, it is inferred correlation. Now, if a variable drives another one, they are connected with a causal relationship. In other words, correlation means there is a statistical association between variables. Causation conveys that a change in one variable causes a change in another variable. And thus, correlations, if not coincidental, are most often merely manifestations of the more fundamental causal processes and relations within the system [6, 7, 4].

Methods for causal discovery are an innovative and exciting recent field of science with both practical and theoretical challenges, discussed with details in Section 2.5, nevertheless, causal discovery has a great potential to revolutionize research. Causal discovery has

been advancing rapidly for two main reasons: the great algorithmic advances in mathematics and computer science over the last several years and the tremendous availability of big data produced with technology and science. In the last decade there has been substantial progress in mathematically representing causal networks, with key algorithms discovery's for finding causal networks with efficient processing of massive amounts of data. And more than ever, the immeasurable growth of data that contains greater variety, arrives in increasing volumes and with great velocity, specially in domains such as bio-informatics, medical, neuroscience and financial applications. These developments have led to rising interest in the task of discovering causal networks directly from observational time series and may boost scientific advances [8, 9].

1.2 Objectives

This thesis aims to investigate the advance of causal discovery constraint-based methods that learn causal association networks directly from observational data. With this perspective, it is aimed first to point out the progress and limitations of the science of causation. This reflection is followed by a discussion of applications where causal discovery methods have already led to important insights into life sciences and biomedical themes to understand how the causal revolution will impact the future of science. Next, we elucidate the principles of causal discovery and how the main constraint-based methods work.

Following, around a promising constraint-based causal discovery method named PCMCI, we perform theoretical assumptions simulations in a synthetic multivariate time-series to verify the implications of performing a causal search on non-compliant data. Finally, a case study for a real-world application was selected to verify the potential and challenges of practical causal investigations and to analyze the prospects of utilizing a causal structure in predictive models.

From those developments, this work intends to answer the following objectives:

1. Elucidate the goals and advances of causality research.
2. Decode the principle on which the constraint-based causal discovery methods derive causal structures, how causality is represented, and what challenges the science of causality faces.
3. Expose applications and potentials of causal discovery.
4. Perform causal structure search on synthetic data, manipulating different data conditions.
5. Perform causal structure search on real-world data.

6. Utilize causal structure in predictive models.

1.3 Dissertation structure

This work is composed of five chapters, structured as follows: The first Chapter introduces the topic of this project and defines the pursued main goals. The second Chapter discusses fundamental theoretical concepts and covers practical applications and limitations. The third Chapter shows the implication of different data features on causal discovery through simulations. In the fourth Chapter, we present a real-world application of causality in the context of physiological parameters of patients admitted to critical care units. Finally, in the fifth Chapter, we provide an overview of the main conclusions of this research, including achievements, limitations, and future work.

2. Background and Related Work

This chapter reviews the main theoretical concepts related to time-series causal discovery within causality research and is organized as follows. Section 2.1 introduces the distinction between correlation and causality. Section 2.2 divides the two main tasks of causality research. Section 2.3 focuses on causality discovery in time series data and defines the representation of causal structures. Section 2.4 reviews the properties of constraint-based causal structure discovery (CSD) methods, followed by section 2.5, where practical issues are presented. Section 2.6 displays actual application examples. Moreover, in concluding the chapter, key evaluation metrics are presented in Section 2.7.

2.1 Correlation vs Causation

As mentioned in the introduction, correlation does not implicate causation. Rushing into causation conclusions from correlated data can be very misleading and problematic. It is important to identify those differences for drawing sound scientific conclusions from research.

2.1.1 Correlation

Variables are correlated when there is a mutual statistical relation, whether causal or not, to which the variables progression are coordinated with one another. The correlation coefficient expresses the strength or degree of their connection without making a statement of cause and effect. Two variables can be highly correlated without a conclusive causal link between them for mainly three reasons: the directionality problem, coincidental associations, and the third variable problem [6].

The directionality problem occurs when two variables may actually be causally related. However, it is seemingly impossible to determine which variable causes changes in the other, which is the cause, and which is the effect. For example, vitamin D levels are correlated with depression, but it is not clear whether low vitamin D levels cause depression or whether depression causes lower vitamin D intake [6].

Coincidental associations result from the correlation of variables that happen only by chance. To illustrate, there is real data evidence that the number of Nicholas Cage films is highly associated with the number of deaths by dawning [10]. Yet, there is no causality here, Nicholas Cage appearance in films did not cause the deaths, and neither did the deaths caused the films. If there were a misinterpretation of causality, an intervention in one variable would be expected to have an effect on the other.

The third variable problem occurs when a confounding variable affects two or more variables, making them seem causally linked even though they are not. For instance, ice cream sales and crime rates are closely correlated but not causally related. Instead, a third variable, hot temperatures, drive both variables separately, creating a confounding relationship between ice cream sales and violent crime rates [6].

The third variable phenomenon is a well-known statistical puzzle in science, and it is often demonstrated with the Simpson's Paradox, also called Yule-Simpson effect. A phenomenon where it would be possible to draw two opposite inferences from the same data, depending on how one grouped the data prior to the statistical analysis [7].

A medical trial comparing the success rate of two treatments for kidney stones while searching for the best treatment option faced Simpson's Paradox [11]. Table 2.1 portrays a similar situation and presents the number of prescribed treatments and their success rate in two scenarios for treating small and large kidney stones. The compared treatments were treatment A, open surgical procedure, and treatment B, percutaneous nephrolithotomy closed surgical procedure. The numbers in parentheses denote the success cases over the total size of the group.

Table 2.1: Kidney stone treatments

Stone size	**Treatment A**	**Treatment B**
Small	93% (81/87)	87% (234/270)
Large	73% (192/263)	69% (55/80)
Total	78% (273/350)	83% (289/350)

The data shows a better success rate in both small and large kidney stones when treatment A is applied. Yet, when considering both subgroups simultaneously, the total data points to a better success rate when treatment B is applied, leading to an apparent paradoxical result. The issue is that a confounding variable that affects the success rate of the treatments was not accounted for, in this case, the stone size. Small kidney stones have a better success rate than large stones, and therefore small stones are more likely to be assigned to treatment B, the less invasive option. So, the less effective treatment B appeared to be more effective than A because it was more often applied to the small stones. Hence, the inadequate comparison could lead to confusion or false conclusions [11].

So the Simpson's Paradox may arise if there is (at least) one confounding variable that has not been accounted for in the variables correlational analysis, resulting in a misleading conclusion. The paradox can be resolved when confounding variables and causal relations are appropriately addressed in the statistical modeling. So, to distinguish the causal effect

of treatment from the causal effect of the stone size on the success rate, the confounding variable needs to be addressed. The causal effect of the treatments in the outcome is therefore what remains after adjusting for every confounding variable [7, 11].

To identify the exact causes that leads to changed outcome and quantify those links is the shared purpose of several medical trials. And that is the precise ambition of inferring causality, not correlation. Section 2.4 details causality in time series data.

2.1.2 Causation

The key step that made possible the rise of causality research was to express causal relations mathematically or derive causal information from statistical data and therefore make it possible to test the essence of apparently linked variables. Having defined that correlation does not imply causation, the contra-position is that causation means correlation, in a way that statistical independence between variables indeed implies an absence of direct causation.

Two random variables X and Y are statistically independent of one another if for each value of (X, Y) denoted by (x, y), $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x).\mathbb{P}(Y = y)$ And it is denoted by $X \perp\!\!\!\perp Y$ or $(X \perp\!\!\!\perp Y)D$; otherwise, they are statistically dependent. If X and Y are statistically dependent only by a confounding variable Z, then for a given Z, the two random variables X and Y are conditionally independent. Therefore, for any value of (X, Y, Z) denoted by (x, y, z), the $\mathbb{P}(X = x, Y = y|Z = z) = \mathbb{P}(X = x|Z = z).\mathbb{P}(Y = y, Z|z)$, and is expressed by $X \perp\!\!\!\perp Y|Z$ or $(X \perp\!\!\!\perp Y|Z)D$, otherwise they are conditionally dependent [12].

Therefore, if X and Y are statistically correlated, either X causes Y, or Y causes X, or X and Y have a common cause, confounder. If X and Y have a common cause Z (only), then conditioning on Z would make X independent of Y.

The first major effort to mathematically express the theoretical definition came with the proposal of Granger Causality, which later triggered a whole literature on causal discovery test methods. It was introduced by Granger in 1969, with a statistical hypothesis test for determining whether one time series is useful to forecast another. It is fair to think of granger causality as a predictive causality. When testing if variable X causes Y, the granger causality test evaluates if the prediction of future values of Y is improved by modeling the prediction on the past values of both variables, rather than only Y own past values. If there is an improvement, then variable X evolves over time, causing the evolving variable Y. It is not a test to define if X causes Y, but if X forecasts Y [13]. Figure 2.1 illustrates this concept.

The principles from granger causality relationships are:

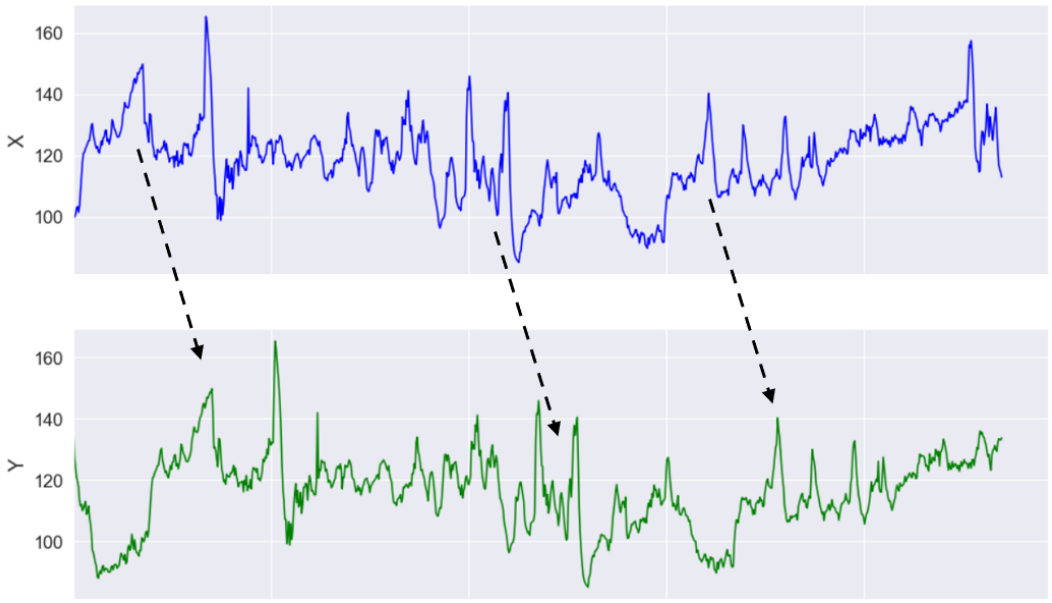

Figure 2.1: Illustration of time series X that Granger-causes time series Y with a time lag

1. The cause happens prior to its effect.

2. The cause has unique information about the future values of its effect.

Considering $\mathbb{P}$ to denote the probability, A any non-empty set, $\mathcal{I}(t)$ the information available at time t t in the whole universe, and $\mathcal{I}_{-X}(t)$ the information available in the modified universe where X is excluded. The two two principles are tested by the following mathematical hypothesis (Equation 2.1) in the identification of a causal effect of X on Y [13]:

$$\mathbb{P}[Y(t+1) \in A \mid \mathcal{I}(t)] \neq \mathbb{P}[Y(t+1) \in A \mid \mathcal{I}_{-X}(t)] \tag{2.1}$$

From a pure mathematical test and expression of causality came the Markov idea of representing causal structures by employing Bayesian Networks to represent probabilistic causality models graphically, called directed acyclic graphs (DAGs). The graph is therefore structured from the likelihood that any of several possible known causes was the contributing factor of an event that occurred [14]. This DAG representation of causality would later shape causal discovery.

The Markov property defines that the probability distribution of future states of a process depends only on the current state and not on past information. Therefore, a Markov process does not remember the past if the current state is given. Hence, it is denoted as a process with memoryless property [14].

2.2 Causality research

Causality research is often essential to derive scientific conclusions from research. Knowing causal relationships allows us to understand and describe mechanisms, predict the results of interventions in a system, and possibly control events and outcomes, justifying the growing interest of the scientific community that has recently started to emerge in the life sciences. This growth is confirmed by the increasing number of published scientific papers on causality research on PubMed [15], a search engine that accesses mainly the MEDLINE database (Figure 2.2).

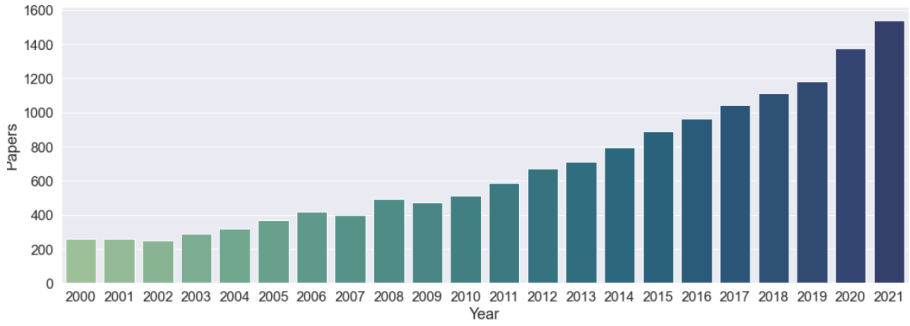

Figure 2.2: Number of published papers related to causality research on life sciences and biomedical topics over the years (Keywords of query: "causal discovery" or "causality research" or "causation" or "causal inference")

Here we divide causal research into two main tasks, causal discovery and causal inference. Briefly, causal discovery aims to identify or infer causal relationships between variables in the data, and causal inference seeks to estimate the effect of an intervention or treatment on a set of parameters. The central theme of this work is causal discovery.

Of course, an appropriate research design can distinguish between correlations and causal relationships. Causal links between variables can be indeed demonstrated with randomized controlled experiments that can establish directionality from variables changes and eliminate the influence of third variables [6].

Since controlled experiments are often impossible, unethical, or highly expensive, causality research brings causal discovery algorithms to infer and quantify potential causal dependencies from time series data without intervening [16].

Causal research is growing to become a mature scientific approach, and it is vital to establish an accurate terminology that can clearly express the definitions and meanings of causality research. Table 2.2 presents key nomenclatures.

Table 2.2: Causality nomenclature [17]

Terminology	Synonyms	Definition
Causality	Causal relations Causation	Causal relation between variables
Causal Effect		Strength of a causal relation
Causal Inference	Learning Causal Effects Forward causal inference Forward causal reasoning Causal treatment effect estimation	Identification and estimation of causal effect
Causal Discovery	Causal Structure Search Causal Learning Causal Search Causal Network Reconstruction	Inferring causal structures from data
Causal Graph	Causal Diagram	A graph with variables as nodes and causality as edges
Confounder variable	Confounding	Indirect causing
Parent variable		Directly causing
Children variable		Directly caused by
Descendant variable		Directly or indirectly causing
Ancestor variable		Directly or indirectly caused by
Spurious links	Spurious associations	Illegitimate causal link, false detection

2.2.1 Causal inference

Given a causal relationship between variables, for example, X being a driver of Y, a natural goal is to infer the exact effect of this causal relationship. Causal inference or causal treatment effect estimation will estimate the effect on Y given any changes in the cause X. With that, it is possible to answer crucial decision-making questions in various fields by estimating from data the effect of interventions, treatments, or policies [18].

For better comprehension, this science domain can help to resolve questions like: Which medicine is more effective in treating a specific disease? Was an enforced policy effective? Alternatively, find the one that would be the most effective at reducing emissions from many different climate change policies. Instead, suppose there is a rise in mental health disorders; one possible substantial cause is social media. In that case, causal inference tries to evaluate and quantify the direct contribution of social media to the problem [18].

Furthermore, some studies on causal inference have already answered similar questions. For example, Abadie et al. investigated the effectiveness of a large-scale tobacco control program passed in 1988, Proposition 99, that increased California's cigarette excise tax by 25 cents per pack to reduce cigarette consumption. The study demonstrated the applicability of the synthetic control in the causal effect estimation, and the results suggested

that the policy was more effective than prior estimates had reported [19].

In another investigation, Bica et al. tackled the extensive challenge in medicine for estimating treatment effects in the presence of hidden confounders. The study proposed a framework to deconfounder the analyzed data before the causal inference estimation. Hidden confounding was present in the data set as patient comorbidities, and several lab tests were not included. The results showed that the framework could remove the bias from hidden confounders when estimating treatment responses for antibiotics, vasopressors, and mechanical ventilators on the white blood cell count, blood pressure, and oxygen saturation conditional on the patient's history [20].

2.2.2 Causal Discovery

In section 2.2.1, causal inference was briefly discussed with the prospects it brings. The great challenge of causal inference in real-world applications is that it requires prior knowledge of the causal structure, which is often not accessible. This is where causal discovery comes in handy. Causal discovery aims to infer the causal structure from raw data. In other words, given a data set, derive a causal model that describes it [21].

Informally, causality is defined as a relationship between two variables, X and Y, such that changes in X lead to changes in Y. The main difference between association and causality is the possibility of confounding. Suppose there is no direct causal relationship between X and Y. Rather, a third variable, Z, causes both X and Y. In such a case, a change in X does not cause a change in Y, even though X and Y are strongly related. More precisely, a causal link is a direct effect between X and Y that remains after adjusting for confounding. Confounding can be observed or unobserved (latent) [2].

Taking Figure 2.3 as an illustration, from given data, either observational or simulated data, linear or nonlinear, and structured in many dimensions, causal discovery assumes that there are some uncovered dynamics governing variable A in respect to time so that A can be written as a function of a subset of the other variables contemporaneously and some noise modeled as describing unresolved processes. This subset of variables that are the causal drivers of variable A is denoted as parental variables. And the same function reasoning can be done for the other variables [22].

There is a wide variety of algorithms exploring different causal discovery approaches, all able to reveal the underlying causal model that governs the data structure, with different mathematical and computational approaches justified by specific statistical data characteristics. The methods can vary tremendously. Still, it is possible to categorize the algorithms in their general common ground. Usually, the causal discovery approaches for

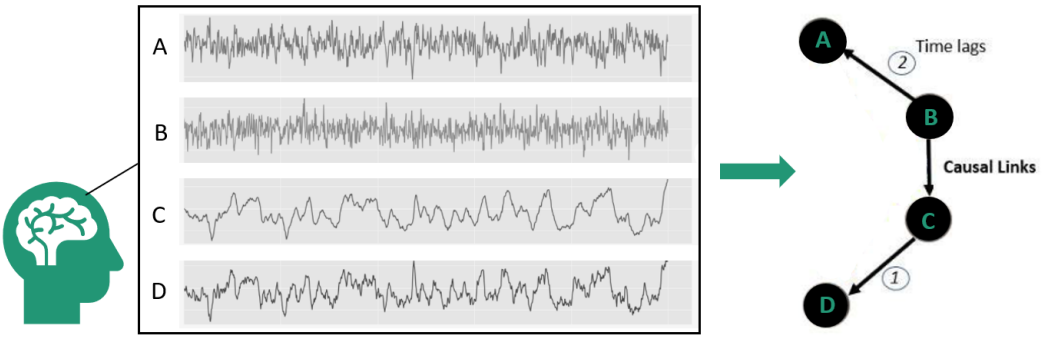

Figure 2.3: Illustration of the task of Causal Discovery for multivariate time series: From a time series data set from a given analysed system, such as neuronal signals, Causal Discovery aims to find the representative causal network, accounting for cause-and-effect time lag, that influence the behavior of the observed data.

time series data are divided into three main categories: constraint-based, score-based, and hybrid-based [3].

Generally, the constraint-based algorithms construct the causal structure based on conditional independence constraints. In contrast, the score-based algorithms generate several candidate causal graphs, assign each a score, and select a final graph based on the scores [2]. Hybrid-based algorithms explore a combination of score and constraint-based algorithm properties [2, 23]. Regardless of the approach, the algorithms need to narrow down the possible underlying causal structures of a given data set. Thus, some assumptions are necessary regarding the raw data itself, and the most common assumptions are described in section 2.4.1. That being done, we are left with a finite number of possible causal structures to be handled with [12].

Table 2.3 provides different proposed causal discovery methods within the different categories for more details. This paper covers the principles and progress of constraint-based computational methods for causal discovery and discusses applications.

Table 2.3: A overview of causal discovery methods

Methods	**Algorithm**
Constraint-Based	PC [24] PCMCI [25] Fast Causal Inference (FCI) [26] Inductive Causation (IC) [27]
Score-Based	Greedy Equivalence Search (GES) [28] Greedy Interventional Equivalence Search (GIES) [29]
Hybrid-Based	Structural Agnostic Modelling (SAM) [30] Causal Additive Model (CAM) [31] Causal Generative Neural Network (CGNN) [32]

2.3 Causal Discovery for Time Series

The type of data of interest for causal discovery in this work is time series. Causal discovery for time series data refers to understanding and identifying inter-dependencies amongst individual data components collected in adjacent periods that can be contemporaneous or between distinct observations. This task allows various applications in multiple fields, such as economy, medical, and earth system science, which are largely interested in understanding the progression of dynamical systems. For instance, causal discovery for time series can be used to identify the performance indicators of stock analysis, uncover the mechanisms of pathogenicity of a recently mutated virus, or discover the causal relations between the external drivers of climate change and climate variables [12].

Differently than purely cross-sectional data, which presents the observation of subjects at one point or period of time, or for which the analysis has no regard for differences in time among the observations, time series data are categorized by observations about a single subject at multiple points or periods, indexed in time order. This distinction leads to significant prospects differences in causal discovery models. For cross-sectional data, causal discoveries do not exploit causal precedence (X causes Y if X happens before Y) since the data represents a single point in time [33].

Causal structure search of time series data can overcome the problems found in cross-sectional data by benefiting from this causal precedence in the time component. Time plays the most important and explicit role here. When faced with an undirected causal link between two variables, it is safe to assume that the relationship direction flows from past to future, as events in the present cannot cause events in the past [33].

By discovering those causal relations, a causal graphic model shines a light on the underlying causal association's network that rules within the data. A causal graphic model visually represents the results of a causal structure search for better and easier comprehension. The causal graph uses nodes to represent the variables of interest in the time series, possibly in different points of time—time lags—and arrows to represent when one function of one variable appears to be a causal relation in the argument of another variable [3].

2.3.1 Time Series DAGs

Direct acyclic graphs (DAGs) are used to represent a causal model and provide visual qualitative information about all conditional (in)dependencies of a set of variables. Causal DAGs are mathematically grounded yet simple and easy to understand while maintaining consistency. Each variable, in this case, a time series, is represented by a vertex (also called a node), and the causal connections between them are represented by the edges (also called

a link). Each vertex necessarily has a specific direction and thus represents a single directed flow from one to another. The representation is acyclic, meaning there is no path in the graph leading back to the original vertex because there are no cyclic causal relationships. The structure is repetitive in time due to the assumption of causal stationarity [25, 34].

An edge $X^i_{t-\tau} \to X^j_t$ with $\tau > 0$ is referred to as lagged and the integer τ is its lag. Edges $X^i_t \to X^j_t$ are called contemporaneous. The order of a process, denoted by p_{ts}, is the causal-effect maximum lag.

For illustration, consider the following linear structural causal process shown in Figure 2.4 composed of four time series, detailed in Equations 2.2, 2.3, 2.4, 2.5 with maximum causal lag of two, where η refers to an independent random variance that adds bias to the components.

$$A_t = 0.5A_{t-1} - 0.6B_{t-1} + \eta^0_t \tag{2.2}$$

$$B_t = 0.3B_{t-1} + \eta^1_t \tag{2.3}$$

$$C_t = 0.8C_{t-1} + 0.4B_{t-1} + 0.6D_{t-2} + \eta^2_t \tag{2.4}$$

$$D_t = 0.9D_{t-1} - 0.8A_{t-1} + \eta^3_t \tag{2.5}$$

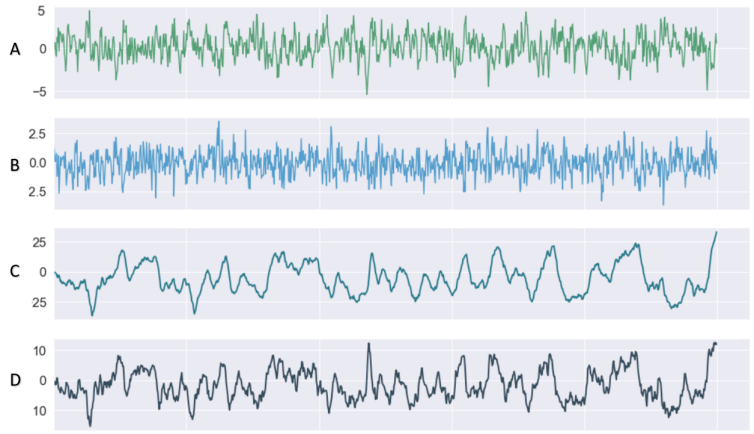

Figure 2.4: Synthetic multivariate time series plot where y-axis represents variables amplitude and x-axis represents instances in time

The associated categorical DAG and time varying DAG of this process is shown in Figure 2.5.

Through the categorical DAG in Figure 2.5 I, we can visualize the three main different types of paths of links in a DAG: Chains, forks, and inverted forks. It is defined as a chain the path of $B_{t-1} \to A_{t-1} \to D_{t-2} \to C_t$, as a fork $D_{t-2} \to C_t \leftarrow B_{t-2}$, and as an inverted fork $A_t \leftarrow B_{t-1} \to C_t$. In the last path, B_{t-1} is denoted as a common driver of A_t and C_t. The

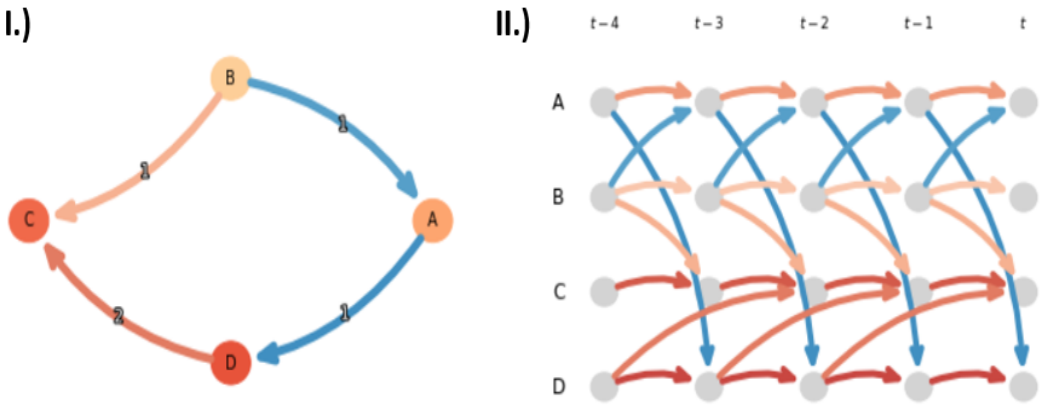

Figure 2.5: I) Categorical DAG II) Time varying DAG of the synthetic system [2.5]

number on every path informs the lag between the variable's relationship. Different edge representations between the nodes characterize the different types of causal relationships. In the given example, only oriented causal parental relationships were present. All possible structure representations are explained in Table 2.4

The time-varying DAG in Figure 2.5 II, in principle, extends to the infinite past and future. However, due to the repetitive structure as imposed by causal stationarity, it is sufficient to restrict to a time window at least as large as $[t - p_{\text{ts}}, t]$ (for the above example, this means it would have been sufficient to show two-time step less) [25].

The color tones in the links illustrate the effect sizes (red links have a positive effect, blue links have a negative effect, and grey links are spurious). The node colors denote the autocorrelation strength. Both direct acyclic graphs represent the same causal links, but the time-varying DAG conveys the information of lagged relationships in a more straightforward way [25].

Table 2.4: Interpretation of Causal Structure illustrations [2]

	Present Relationships	**Absent Relationships**
$A \to B$	A is causal parent of B	B is not an ancestor of A
$A - B$	A is causal parent of B or B is causal parent of A	
$A \leftrightarrow B$	There is a unmeasured confounder of A and B	A is not causal parent of B and B is not causal parent of A
$A \circ\!\!\to B$	Either A is a causal parent of B or there is an unmeasured confounder of A and B	B is not a causal ancestor of A
$A \circ\!\!-\!\!\circ B$	One of the following holds: 1. A is causal parent of B 2. B is causal parent of A. 3. There is an unmeasured confounder of A and B. 4. Both 1 and 3. 5. Both 2 and 3.	

2.4 Constraint-Based Method

Often called Conditional Independence Testing methods as well, Constraint-Based algorithms use the idea that two statistically independent variables are not causally linked. By modeling conditional (in)dependence and joint probability distribution of the variables through analysis of the probabilistic relations entailed by the Markov property of the Bayesian network, a DAG represents the results. The Bayesian networks assess an event that occurred and predict the likelihood that any one of several possible known causes was the contributing factor to the event. Given an observational data set defined on a variable set, this methodology consists of two key tasks: first, searching for (in)dependencies for the skeleton phase and then orienting dependencies [34].

2.4.1 Assumptions and Theorem

The statistical modeling is possible when a number of assumptions are taken regarding the data.

- Causal Sufficiency: Assumes that the analyzed set of variables is causally sufficient for a process. That is, for every pair of variables that have their observed values in a given data set, all their common causes also have observations in the data set. Thus, there are no unobserved variables influencing directly or indirectly any of the variables in the set, nor is the data sampled at too coarse time intervals relative to the causal links [9].

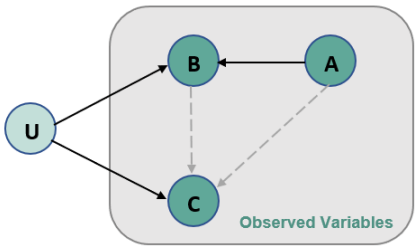

Figure 2.6: Illustration of the consequences of violated Causal Sufficiency assumption due to unobserved variable. U is an unobserved variable thus the direct influence of U in B and C are also unobserved, leading to spurious links (dashed edges) between the observed variables [9]

- Causal Markov condition: The causal Markov condition, also called Markov memory, states that previous states of a variable or process are irrelevant to predicting subsequent states as long as the current state is known. Thus, the joint distribution of a time series process X containing a set of variables with graph G satisfies the causal

Markov condition only if each variable is independent of its non-successors and dependent on its parents [35, 18]. More formally:

$$\mathbb{P}(X_{n+1} = x|X_0, X_1, X_2, \ldots, X_n) = \mathbb{P}(X_{n+1} = x|X_n),$$
$$\mathbb{P}(X) = \prod \mathbb{P}(X|\hat{\mathcal{P}}(X))$$

- Faithfulness: All dependencies in a DAG must be defined under the criteria of conditional independence, implying that the independence in the data is not due to chance but rather to a causal structure. In other words, two variables that are independent on another subset of variables are not connected by a causal link in the graph. The joint distribution of a time series process X and its graph G are faithful to each other only if all independent relations satisfied by the Markov condition are represented in the graph [9].

- Causal Stationarity: The assumption of causal stationarity requires that if the causal mechanisms of a process vary over time due to its dynamics, only one regime belonging to the dynamic process, e.g., a particular season in the climate system, is considered for the conditional independence to be stationary. A common violation of causal stationarity occurs when the dynamical process is based on a trend caused by a latent perturbation [22, 36].

- No Instantaneous effects: Based on temporal causality, this assumption assumes that there are no zero time-lag causal effects, such as $A_t \rightarrow B_t$, because if the process is sampled with sufficient resolution, only delayed causal effects need to be considered.

  Temporal priority makes the causal process asymmetric in time and helps orient a causal relationship when two variables are known to be causally related. However, the temporal difference between two events associated with two-time series cannot be observed if the sampling frequencies of the time series are small. Therefore, two events that occurred at different times may be seen instantaneously in the observed time series. Instantaneous causal relationships, sometimes referred to as contemporaneous causal relationships, correspond to causal relationships between causes and effects that occur at different times but appear to be immediate [34].

- Parametric Assumptions of Conditional Independence Tests: The constraint-based algorithms are virtually based on the (in)dependencies found in the data set, which result from the defined conditional independence test inserted into the algorithms. The options are numerous and are based on the properties of the analyzed dataset. The tests' assumptions that can be chosen in the causal discovery of the method explored in this work, PCMCI, further detailed in Section 2.4.3, are given in Table 2.5.

Table 2.5: Conditional independence test assumptions extracted from [37]

Tests	Assumptions
ParCorr	Univariate variables with continuous, linear Gaussian dependencies
GPDC	Univariate variables with continuous, additive dependencies
CMIknn	Multivariate variables with continuous, general dependencies

2.4.2 PC algorithm

One of the most popular and pioneer advanced constraint-based methods for causal discovery is the PC algorithm, proposed in 1991, named after its authors Peter Spirtes and Clark Glymour. The PC algorithm provides a search architecture based on statistical procedures for deciding conditional independence. The statistical decision procedure might be a hypothesis test for conditional independence or a method based on the difference in fitting scores, such as the Bayesian Information Criterion (BIC) between models. The algorithm relies upon three assumptions, the Markov condition, the Faithfulness assumption, and the Sufficiency assumption [3].

Let D be a dataset comprised of variables A, B, and C as variables. Suppose the PC statistical decision procedure finds A and B statistically interdependent. However, independent when conditioned on a third variable, C, $A \perp\!\!\!\perp B|C$. Then, under the assumptions that there is no influence of the variables past joint probability—Markov condition—, and that independence arises from the causal structure—Faithfulness—, and that there are no unobserved confounding variables: A and B are not, and cannot be, directly causally related. Two variables are only considered directly causally related (with an edge in between) if and only if there is not any subset of the remaining system variables conditioning on which they are independent [3].

Taken the true causal structure shown in Figure 2.7 (I) as an illustration, the PC algorithm steps to find this essential graph goes as following [3]:

1. Creation of what is called a DAG skeleton, a complete undirected graph of which there is an edge connecting every pair of variables, as shown in Figure 2.7 (II);

2. Elimination of edges:

   (a) Test pairwise independencies and for the unconditionally independent variables as $A \perp\!\!\!\perp B$ eliminates the edge between the variables, resulting in the graph in Figure 2.7 (III);

(b) Test independencies of pairs of variables conditioned on their adjacent variables and remove the edges between the conditionally independent pairs, as $A \perp\!\!\!\perp D|C$ and $B \perp\!\!\!\perp D|C$, resulting in Figure 2.7 (IV)

(c) Continue checking independencies conditional on subsets of variables of increasing size until there are no more adjacent variables. In the considered example, C and D are not independent conditional on A or on B or on both A and B, therefore $C \perp\!\!\!\perp D|A, B$ is false and there are no further statistical decisions to make. Similarly for A and C, and for B and C.

3. Orientation of the remained edges:

(a) For each triple of variables as (A, B, C) such that A and C are adjacent, B and C are adjacent, and A and B are not adjacent, orient the edges A – B – C as A $\rightarrow$ C $\leftarrow$ B, the so called fork structures, resulting in the graph in Figure 2.7 (V).

(b) For each triple of variables such that B $\rightarrow$ C – D, and B and D are not adjacent, orient the edge C – D as C $\rightarrow$ D, defining the so called v-structures. In Figure 2.7 (VI), Y $\rightarrow$ Z – W is oriented as Y $\rightarrow$ Z $\rightarrow$ W finishing the orientation of the causal propagation.

It is possible that in some processes, none of the orientation rules apply to a particular undirected edge; therefore, that edge remains undirected in the output. This implies that although the two variables are known to be adjacent, it cannot be determined in which direction the edge points, i.e., it is not known which is the cause [3].

In the given example, if the chosen conditional independence tests reveal the correct relations, then the PC algorithm, from the defined general sets of rules, derives the true causal DAG of the system components since there is no other DAG that would represent the (in)dependencies. However, in many cases, there is more than one DAG that would encode the same statistical interpretations [3]. Those are said to belong to the same true Markov Equivalence Class. In this scenario, the PC Algorithm will return all Markov equivalent graphs instead of one final DAG, as illustrated with the original causal structure (I) of Figure 2.8.

Since the algorithm relies fundamentally upon finding conditional independence, statistical power, known as sensitivity, is its biggest challenge [25]. If two variables are indeed independent, the power for detecting this independence depends on:

1. The sample size of the dataset

2. The number of variables and, therefore, the conditioning dimension of the set

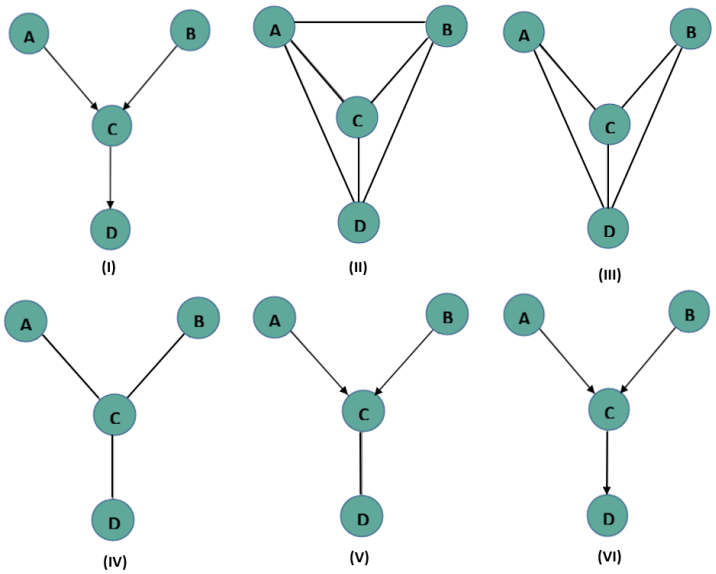

Figure 2.7: Illustration of the PC algorithm steps. (I) Original true causal graph. (II) Complete undirected graph (DAG Skeleton). (III) The $A-B$ edge is removed because $A \perp\!\!\!\perp B$. (IV) The $A-D$ and $B-D$ edges are removed because $A \perp\!\!\!\perp D|C$ and $B \perp\!\!\!\perp D|C$. (V) After finding v-structures. (VI) After orientation propagation. [3]

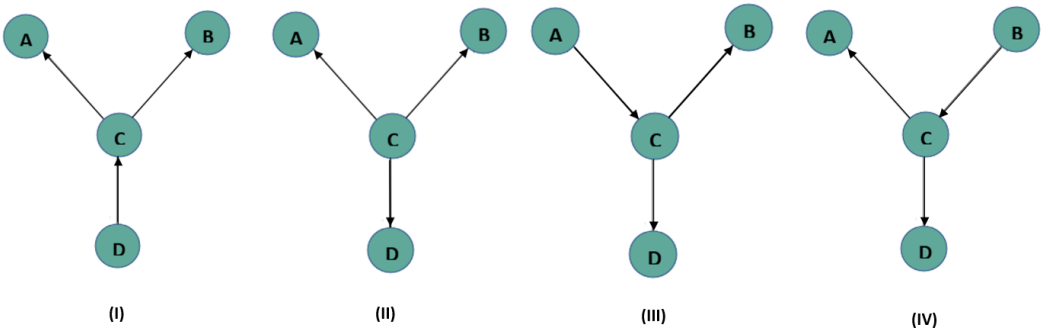

Figure 2.8: Illustration of Markov Equivalence Class. (I) Original true causal graph. (II), (III) and (IV) are Markov equivalent of (I). I, II, III and IV belong to the same Markov Equivalence Class because the conditional independencies of $A \perp\!\!\!\perp B|C$, $A \perp\!\!\!\perp D|C$ and $B \perp\!\!\!\perp D|C$ hold true for all.

3. The significance level of the independence tests
4. The effect size, i.e., magnitude, of the test statistic in the population (The strength of the causal relations)

A fundamental shortcoming of the original PC algorithm is that it is order dependent and thus not stable. Moreover, since it relies on the faithfulness assumption, the algorithm does not account for possible unobserved/latent confounders. Moreover, the fact that the number of independence tests is exponential to the number of considered variables can be time-consuming with high computational power and reduced scalability [3]. The algorithm described next is based on PC and provides advances to tackle those issues.

2.4.3 PCMCI

The PCMCI algorithm is a recent development proposed by Runge et al. to address some of the weaknesses entailed by the PC algorithm. It is adapted to operate with highly interdependent time series systems. The algorithm is constructed under the following assumptions: Causal Markov Condition, Faithfulness, and Causal sufficiency. Additionally, the time lagged causal discovery assumes Causal Stationarity, no instantaneous causal effects (i.e., no instantaneous causal links), and no measurement error [25].

The method consists of two main stages. In the first stage, a Markov set discovery algorithm based on a version of the PC algorithm for time series is used to select the parent sets for all time series variables as detailed in Section 2.4.2.

Then, in the second stage, it carries out the momentary conditional independence (MCI) test to search for time-shifted parents for each pair of variables by testing the independence of all variable pairs with time delays conditioned on the previously selected parent sets. Let X^N be a time series set of N variables, the test of all pairs $(X^i_{t-\tau}, X^j_t)$ with $i, j \in (1, ..., N)$ and time delays $\tau \in (1, ..., \tau_{max})$ establishes a lagged causal link $(X^i_{t-\tau} \to X^j_t)$ in the causal DAG if and only if: $(X^i_{t-\tau} \not\perp\!\!\!\perp X^j_t)|\mathcal{P}(X^j_t)$ where $\mathcal{P}$ denotes the causal Parents set [25].

The maximum time delay is chosen according to the maximum physical time delay expected in the complex system. If a relevant time lag that could explain a dependence between two other variables is not included, the assumption of causal sufficiency would be violated. This additional condition for the lagged parents accounts for autocorrelation and leads to correctly controlled false positive rates at the expected level, even in the presence of confounding factors, and a large extent to an approximately correct result even in the presence of highly interdependent time series [25].

The MCI test controls false positive detection in highly interdependent time series and reduces the number of independence tests it needs to perform. The critical advantage of PCMCI over PC is that the MCI test accounts for autocorrelation, keeping false-positive rates at the expected level [25].

From Causality to Time Series Prediction

More than uncovering the underlying mechanisms that govern data and systems, which evoke answering many of the questions of scientific investigations, the knowledge of the causal parents of a process can also lead to an optimal scheme for prediction.

Predictive modeling is widely used to predict future behavior from data mining of a given past and present data input. Those models can be set to assess risk and detect significant anomalies or acute events in advance, possibly allowing for anticipated interventions to reduce the out-turn of those events. In practice, a prediction using the entire set of a variable undergoes extreme dimensionality and overfitting, meaning that in many cases, variables data aggregation does not effectively convey helpful information by fitting noise in the time series. Thus, it causes an overly complex predictive model that becomes highly adapted to the data on which it was trained but performs poorly with unknown data for the lack of generalizing capability; thus, it is unfit to be applied in real contexts [25]. Besides, those models are often inadequate to anticipate how the relation between past and future values will change if new interventions or changes are undertaken.

To mitigate and overcome this implication, Runge et al. propose using the variables parents set from the found causal structure, once determined in the Markov condition, as the optimal predictors to forecast a process. This optimal prediction scheme consists of three steps performed separately for each size of steps ahead h into the future [25]:

1. Estimate the parents predictors $\mathcal{P}_{t+h}$ from the observed time series with the PC algorithm. Figure 2.9 illustrates the optimal predictors of C_{t+1} and C_{t+2} of the causal structure represented.

2. Rank predictors using forward selection of their multivariate mutual information. This process is mathematically detailed in [25].

3. Forecast the unobserved future value of the target variable in $t + h$ using any prediction method.

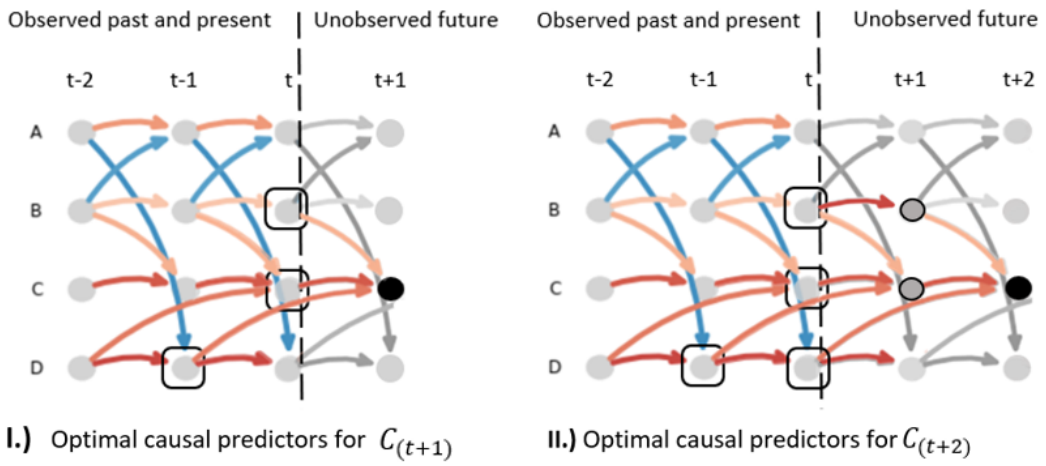

Figure 2.9: Estimation of the parents predictors $\mathcal{P}(C_{t+h})$. For the one time step ahead prediction of C_{t+1} in (I) the optimal predictors are simply the parents of C. In (II) for predicting C_{t+2}, while D_t is still observed, the processes B_{t+1} and C_{t+1} already lie in the future and are not available, however, part of the information in B_{t+1} and C_{t+1} can still be recovered by measuring their parents.

2.5 Practical Issues in Causal Discovery

2.5.1 Statistical power and computational complexity

From a mathematical and computational perspective, algorithmic scalability poses a significant challenge to causal constraint-based methods. The complexity of analyzing systems increases significantly for high-dimensional data—data with too many variables compared to the observation size—and for highly interdependent variables and larger sample sizes. The complexity curve for the time series PC algorithm is polynomial. When analyzing a system with N variables under the defined τ_{max}, the complexity can amount to:

$$N \sum_{\tau_{max}-1}^{N} N\tau_{max} = N^3\tau_{max}^2 \text{ [25]}$$

In the PCMCI algorithm, the complexity is the aggregation of the PC condition selection stage with the MCI step, which involves $N^2\tau_{max}$ tests, resulting in $N^3\tau_{max}^2 + N^2\tau_{max}$ tests. The more variables are considered, the more reliable the decoding of potentially spurious associations is; therefore, the more credible becomes the found causal structure. However, high dimensionality compromises statistical power to establish a true causal relation. Similarly, although a larger sample size is beneficial to making more reliable causal discoveries, the computational time in searching for causal structures increases as the sample size increases. In addition, with low statistical power, weak causal relationships are not well detected, and control of false positives at the desired significance level is compromised in MCI tests [3, 16, 22].

2.5.2 Data challenges

1. Data collection and integration
   Complex systems continue to struggle with data acquisition and integration. For example, the quality of Big Data in healthcare is affected by inconsistencies and instabilities in data collection, as measurements often have missing values for various reasons, from equipment malfunction to operator error. When data entries are not missing at complete random, the conditional independence relationships in the observed data may differ by the underlying causal process [3]. In addition, data come from various sources, including administrative records, clinical registries, electronic health records, biometric data, patient-reported data, medical imaging, biomarker data, clinical trials, and others [38]. With variables often collected in different periods, integrating these data sources results in high dimensionality, incongruence, incompleteness, and complexity. In addition, ethical and legal issues may also pose challenges [22].

2. Unobserved variables
   Arguably, it is difficult to determine the relevant driving forces of a complex system, and even if all driving forces are correctly defined, they often cannot be measured, so the possibility of unobserved variables must be considered in the causal interpretation of the estimated graph since they may render nonstationarity and spurious links [25].

3. Data sub-sampling and time aggregation
   If the data collection sampling is less frequent than the underlying time-dependent causal effects, causal links appear to be contemporaneous and even cyclic due to insufficient time resolution [22].

4. Measurement errors
   Uncertainty in the observed values of the variables, given the ubiquity of measurement error caused by instruments or proxies utilized in the measuring process, can significantly alter the output of various causal discovery methods. Biosignals data collection, for example, is especially challenged by all kinds of measurement errors, such as observational noise, systematic biases, and missing values. The issue of measurement errors is addressed in depth in the study of Zhang et al., [39], where sets of identifiability conditions are proposed to recover optimal causal models from measurement-error-contaminated data.

5. Selection bias
   Selection bias is a crucial problem in statistical analysis that occurs when the probability of including a data point in the sample depends on some property of the point, as mentioned in the Simpsons paradox. If not corrected, selection bias, often distorts the results of statistical analysis and causal discovery and inference [3, 39].

6. Non-stationary processes
   In real-world applications, data is often composed of non-stationary or heterogeneous systems, in which, over time, the data has a distributional shift of the underlying generation process, i.e., causal structure [3].

2.6 Application of Causal Discovery

Learning causal relationships is a major task in the biological and natural sciences. However, compared to other fields where causal detection plays a more prominent role in scientific work, such as economics, geosciences, and social sciences, the potential of causal detection in healthcare is still nascent, especially for time series data. This section presents examples of real-world applications in this area where causal detection has been essential for obtaining meaningful insights.

2.6.1 Application to Alzheimer's pathophysiology

Shan and Ma systematically examined whether causal discovery algorithms can discover known causal relationships directly from observational data for Alzheimer's disease (AD) [2]. Here, the evaluation of the algorithmic discovered structures was possible given the well-established evidence that provides a "gold-standard" causal graph of AD progression for comparison.

The methods investigated were Fast Causal Inference (FCI) and Fast Greedy Equivalence Search (FGES). Causal structure searches were performed in three different scenarios [2]:

1. Without including background knowledge in the algorithms
2. With the addition of trivial background knowledge
3. Using longitudinal data samples

Their work presents promising results: the CSD methods succeeded in discovering graphs that almost matched the gold standard causal graph highlighting the biological mechanisms behind the AD biomarker cascade. However, some critical drawbacks were also revealed. In the case where the input was time series data, the computational costs were relatively high, leading to the use of a much smaller sample. A smaller sample size lowers the statistical power of causal discoveries. When the sample size was reduced by 50 percent, edges that were consistently detected in the total sample were also consistently detected in the reduced sample, with similar precision and recall. The total number of edges detected decreased when the sample was reduced by 75 percent. While the most frequently detected edges continued to be detected, the number of "noise edges," i.e., edges detected in only a few bootstrap iterations, increased [2].

In summary, the results suggest that dedicated causal discovery algorithms outperformed structural equation modelling (SEM)[1] in discovering causal structures. In real-world data analysis, data quality impacted the correctness of the discovered structure. Moreover, dedicated CSD methods managed to discover graphs that nearly coincided with the gold standard. Incorporating prior knowledge and using longitudinal data improved the discovered result by preventing the algorithms from making some potential errors. The work suggests that causal discovery algorithms should be used for best results with longitudinal data providing as much prior knowledge as possible [2].

2.6.2 Genes regulators of flowering time in Arabidopsis Thaliana plants

The transition to flowering, controlled by a complex and intricate gene regulatory network, plays a vital role in the reproduction of plant offspring. In [40], authors aimed to discover the critical causal genes that control and affect the flowering transition of the Arabidopsis Thaliana by dynamic network analysis of time-course gene expression data.

First, 47 expression profiles of the species, each with 21,326 genes and corresponding flowering time data, were collected. From the raw data, a causal network was designed with causal discovery, presenting the genes candidates that caused the most substantial changes in flowering time. According to the causal network, 25 out of the 21,326 genes were detected to be causally responsible for the flowering time. 5 out of the 25 were known regulators of flowering [40].

To validate if the remaining 13 were indeed flowering regulators, they genetically manipulated those genes in the seeds to insert a mutation. From the modified seed types, four grow to have a statistically significant shorter mean flowering time than the control group.

[1] SEM is a statistical technique used to analyze effect-size of variables structural relationships [2].

The study managed to shrink the candidate genes from 21,326 to 25, allowing the manipulation of the remaining, which would be very time-consuming and perhaps impossible to do for all the genes. Overall, 31 percent of the gene-altered seeds produced the desired outcome, considered a very high-efficiency performance [40].

2.6.3 Constructing Brain Connectivity Model Using Causal Network Reconstruction Approach

The study of brain activity and connectivity is an ongoing paradigm that can provide deep knowledge of neurophysiology and ultimately forward insights into how the brain works and the mechanisms of cognitive functions.

The study of Saetia et al. pioneers the investigation of brain connectivity through causal discovery. Here, the interaction between brain regions and information flow was done by applying the Human Connectome Project motor task-fMRI Dataset to the causal discovery algorithm PCMCI to analyze the causal search approaches compared to prior knowledge [36].

Using a motor task-fMRI dataset was justified due to solid scientific knowledge of the mechanisms of the brain's motor functions. The study selected a set of regions of interest to find its causal structure model. Overall, the implementation allowed for a detailed brain connectivity analysis. It attenuated the third variable problem since the absence of direct links can be interpreted as the absence of direct causal associations, and a signal of possible unobserved confounding variable [36]. Nevertheless, there were some challenges and limitations.

The model did not detect the frontal lobe mediation pathway caused by motor movement planning activity, as expected due to prior knowledge. This absent detection may have been caused by the dataset's sampling frequency, which was acquired with an fMRI with a repetition time of 2.8s. If the causal effects are more frequent, the causal associations are detected as instantaneous, which are not regarded as causal links [9]. Moreover, the intrinsic mathematical analysis is sensitive to signal noise and variance. Thus the variation of brain activity due to brain plasticity augments the complexity of causal discovery.

This project also points out that the approach is applicable on an individual level. Nevertheless, the method does not allow for the construction of group connectivity that would enable an interpretation of the brain mechanisms for the general population since modeling connectivity from data across individuals is disputed [36, 41]

The study concludes that despite the faced challenges, causal discovery benefits the interoperability of complex dynamical systems such as neurophysiology and that the implementation of time-lag into the analysis of brain activity allows the distinction between

direct and indirect pathways, which may be a limitation of classical approaches [36].

2.6.4 Causal Factors of Anxiety and Depression in College Students

Huckins et al. described the causal and concurrent networks between stress's critical psychological health factors and self-esteem in anxiety and depression in college students. Although the concurrent relationships between these variables are well defined, the causal relationship between the factors is not well understood. Acknowledging how causal factors affect the development of mental health conditions over time may provide critical information for targeted treatment or, perhaps more efficiently, preventive interventions for individuals at risk for depression and anxiety [42].

The method involved surveying two cohorts of students over 40 weeks with ecological snapshots (EMAs), and responses were applied as longitudinal data to the PCMCI algorithm. Results provided insights into the temporal dynamics between mental health factors. Stress was found to be a causal predictor of anxiety. In contrast, low self-esteem was a causal predictor of depression and, to a lesser extent, anxiety, indicating that reducing high-stress levels may reduce subsequent increases in self-reported anxiety. The study concludes that continued testing and expansion of models for the interactions between these and other psychological state measures can improve the identification of critical causal factors. Moreover, further provide potential treatments or strategies to mitigate them, especially among groups at higher risk than the general population [42].

2.7 Evaluation Metrics

2.7.1 Pattern Evaluation Metrics

Having found a causal model to represent a system, a most necessary task is to evaluate the result and determine how reliable the model represents reality. Several metrics can be used to assess the learned causal models regarding the ground truth model by comparing their network patterns if the ground truth casual model is available. Some pattern metrics are presented in Table 2.6.

2.7.2 Evaluation Metrics for Regression

However, if a ground-truth model is unavailable, causal structure models can be evaluated regarding their performance in classification or regression tasks [33]. In these cases, the traditional statistical performance metrics are adopted. Several metrics can be chosen to evaluate the performance of the regression. Often used ones, and later performed in the

Table 2.6: Causal pattern evaluation metrics (Extracted from [33])

Metric	**Description**
Missing edges	Number of edges that are present in the original model but not in the generated
Extra edges	Number of edges that are present in the generated model but not in the original
Incorrect adjacencies (undirected edges)	Number of undirected edges that are present in the generated model but not in the original one
Correct directed edges	Number directed edges present in the generated model that were correctly directed
Incorrect directed edges	Number directed edges present in the generated model that were incorrectly directed
Structural hamming distance	Sum of missing edges, extra edges, and incorrectly directed edges
Adjacency precision	Adj Precision = Correctly predicted adjacencies a / Predicted adjacencies b
Adjacency recall	Adj Recall = Correctly predicted adjacencies / True adjacencies c
Arrowhead precision	Arrhd Precision = Correctly predicted arrow heads d / Predicted arrow heads e
Arrowhead recall	Arrhd Recall = Correctly predicted arrow heads / True arrow heads f

case study, are defined in Table 2.7 considering n as the number of data points, y_i the observed values, $\overline{y}_i$ the mean of y_i, and $\hat{y}_i$ the estimated values.

Table 2.7: Causal regression evaluation metrics [43]

Metric	Description	Interpretation
R-Squared	$R^2 = 1 - \frac{\sum_{i=1}^{n} \hat{y}_i - \overline{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2}$	Strength of relationship between the variance of predicted and actual values, the closer to one, the more accurate the prediction
Mean Absolute Error (MAE)	$MAE = \frac{\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert}{n}$	Mean of absolute distance between predicted and actual values
Maximum residual error	$r = max(y_i - \hat{y}_i)$	Maximum distance between predicted and actual values
Mean Squared Error (MSE)	$MSE = (\frac{1}{n}) \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$	Mean distance between predicted and actual values. The closer to zero, the more accurate the prediction
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{(\frac{1}{n}) \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$	Root of MSE, the closer to zero, the more accurate the prediction
Normalized Root Mean Square Error (NRMSE)	$NRMSE = (\frac{RMSE}{\sigma_y})$	RMSE normalized through the standard deviation of actual values

3. Simulations

In this chapter, we develop pertinent data simulations under different sets of conditions and perform the constraint-based causal discovery method PCMCI causal search to identify the impact of specific conditions of the data on the causal outcome.

Here, we aim to observe the PCMCI causal detection power when the input data complies with the algorithm assumptions and the implications of violating those assumptions. As detailed in the previous chapter, besides the Markov condition and Faithfulness assumption in which the causal DAG of a given process is constructed, the PCMCI method requires further assumptions, such as causal sufficiency, causal stationarity, and the absence of instantaneous causal effect. Using a synthetic process makes it possible to extract comparative insights on causal outputs since we have access to the ground truth causal structure.

3.1 Materials and Methods

All synthetic multivariate time series in this work were generated on the basis of random seeds with a vector auto-regressive (VAR) process to structure the evolving relationship between the multiple variables in time in a way that each variable is a linear function of past lags of itself and past lags of the other variables, in addition to an independent random variance that accounts for error [44].

We used the synthetic causally linked multivariate time-series already presented in Equations 2.2, 2.3, 2.4, and 2.5, which are reintroduced below:

$$A_t = 0.5A_{t-1} - 0.6B_{t-1} + \eta_t^0 \tag{3.1}$$

$$B_t = 0.3B_{t-1} + \eta_t^1 \tag{3.2}$$

$$C_t = 0.8C_{t-1} + 0.4B_{t-1} + 0.6D_{t-2} + \eta_t^2 \tag{3.3}$$

$$D_t = 0.9D_{t-1} - 0.8A_{t-1} + \eta_t^3 \tag{3.4}$$

We performed a set of interventions on the data to violate the stated assumptions. After discussing the causal search performance in the compliant synthetic data, we examine the reverberations of causal structure search in non-compliant data.

3.1.1 Causal Sufficiency Transgression

The causal sufficiency assumption requires that the analyzed set of variables is causally sufficient for a process. Thus, we implemented two experiments, one to hide input variables

and the other to sub-sample the input data, before searching for their causal structure. By doing so, we were violating the causal sufficiency assumption of the given process. Each variable was hidden in succeeding runs. Afterward, all the variables were again inserted in the causal search but with a sampling reduced by half and by one-third.

3.1.2 Causal Stationarity Transgression

In order to test the implications of the causal stationarity assumption, we intervene by appending a different causal mechanism regime to the original causal process. The new causal mechanism is composed by the four-time series defined in Figure 3.1. As shown in Figure 3.1, at a certain time the resulting data points are governed by a different causal structure, violating the causal stationarity assumption.

$$A_t = 0.5A_{t-1} - 0.6B_{t-1} + \eta_t^0$$

$$B_t = 0.3B_{t-1} + \eta_t^1$$

$$C_t = 0.8C_{t-1} + 0.4B_{t-2} + 0.6D_{t-2} + \eta_t^2$$

$$D_t = 0.9D_{t-1} - 0.8A_{t-1} + \eta_t^3$$

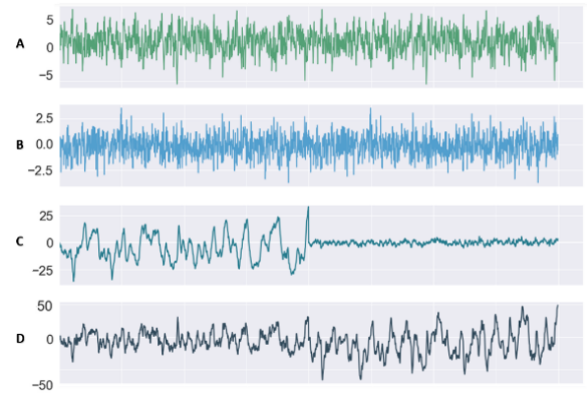

Figure 3.1: Synthetic multivariate time series that violates the Causal Stationarity assumption with y-axis representing variables amplitude and the x-axis representing instances in time

3.1.3 Independent sub causal processes

In this experiment, we investigate how the causal discovery method reacts when there is a combination of two independent causal mechanisms. Two variables, E and F, that are not causally linked to the original process were added to the causal search input data, resulting in the multivariate time-series presented in Figure 3.2.

3.2 Results

The Causal Structure found without intervening in the original data is shown in Figure 3.3. The original synthetic data did not pass through any data distortion and therefore complied with every assumption of the PCMCI method.

$$A_t = 0.5A_{t-1} - 0.6B_{t-1} + \eta_t^0$$

$$B_t = 0.3B_{t-1} + \eta_t^1$$

$$C_t = 0.8C_{t-1} + 0.4B_{t-2} + 0.6D_{t-2} + \eta_t^2$$

$$D_t = 0.9D_{t-1} - 0.8A_{t-1} + \eta_t^3$$

$$E_t = 0.7E_{t-1} + 0.8F_{t-2} + \eta_t^3$$

$$F_t = 0.5F_{t-1}$$

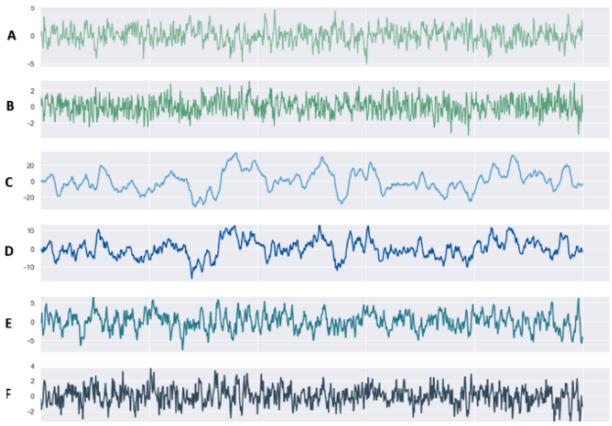

Figure 3.2: The two independent causal process, first one composed by A, B, C and D, and the other one by E and F, with y-axis representing variables amplitude and the x-axis representing instances in time

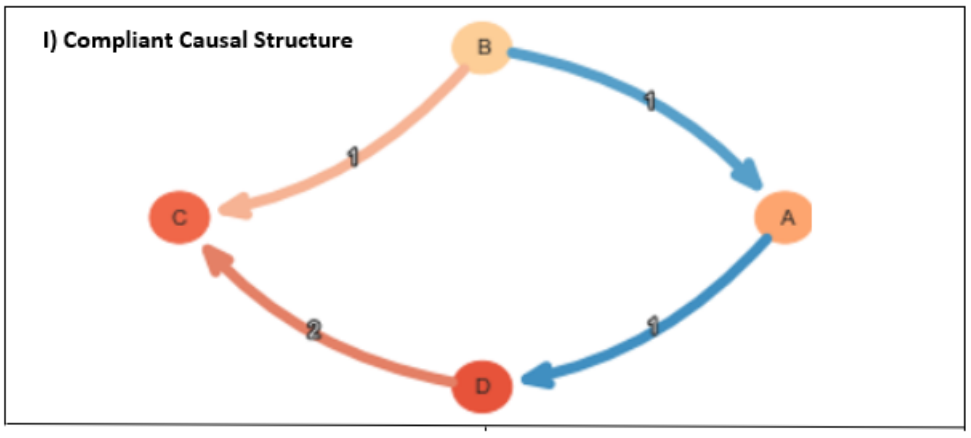

Figure 3.3: PCMCI causal structure found in the original generated data

3.2.1 Causal Sufficiency Transgression

Figure 3.4 contains the causal search outputs under the different conditions of causal sufficiency violation.

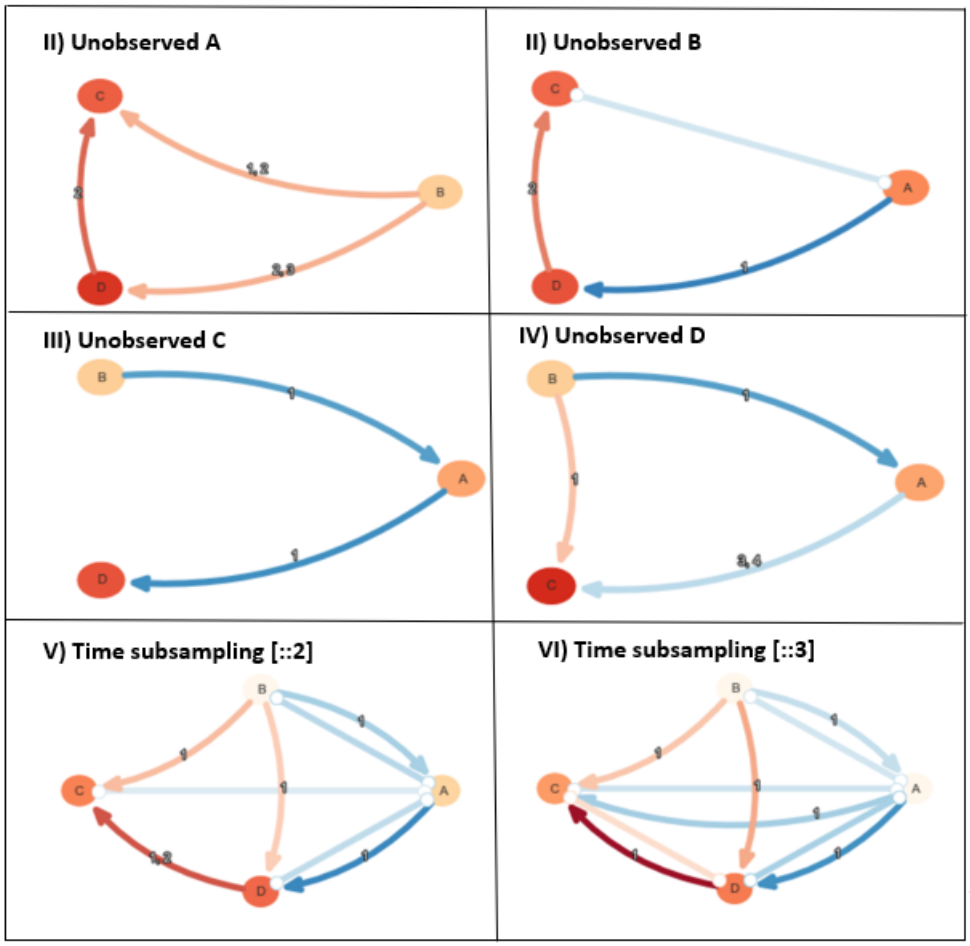

Figure 3.4: I) Causal Structure found by utilizing data compliant to all assumptions. Causal Structure found when violating the Causal Sufficiency assumption of I) by unobserving the variables II) A, III) B, IV) C, V) D, and by time sub-sampling of factor V) 2 and VI) 3.

3.2.2 Causal Stationarity Transgression

Figure 3.5 presents the resulting DAG of the causal search for the process shown in Figure 3.1 that violates the causal stationarity assumption.

3.2.3 Independent sub causal processes

The result of applying the PCMCI causal structure search in a system containing two independent sub causal processes is shown in Figure 3.6.

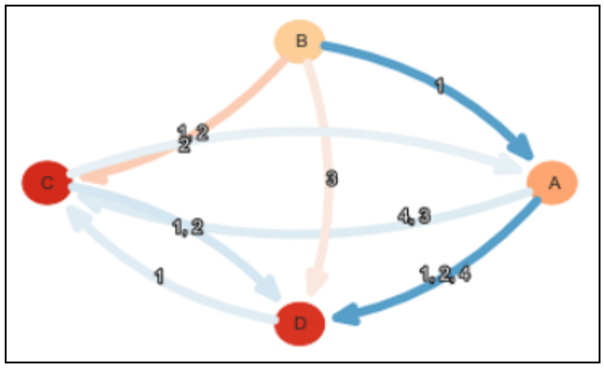

Figure 3.5: Resulting DAG of the non stationary process

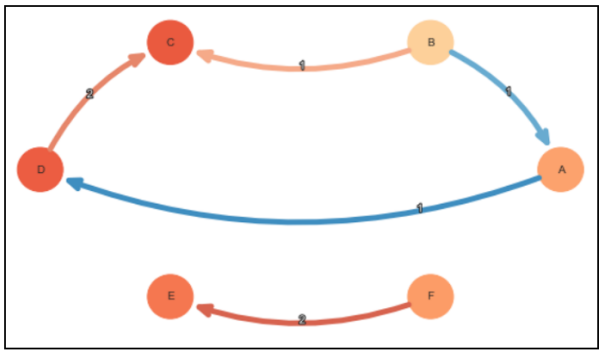

Figure 3.6: Causal structure found by PCMCI on the intervened data containing two independent causal processes

3.3 Discussion

When the input data was completely compliant with the method assumptions (Figure 3.3), the found causal structure successfully detected all causal links and their respective effect time lag entailed in the process that generated the synthetic time series. This result confirms the method's ability to detect causal relationships directly from time-series data when performed in compliance with the algorithm assumptions. However, the causal structure results under data that violates the PCMCI assumptions are uncertain.

The violation of the causal sufficiency assumption led to numerous false detection. In this specific causal structure, hiding variables data from the causal search implicated in indirect causal pathways to be detected as direct links (Figure 3.4). Moreover, When B—the common driver of A and C—was hidden (Figure 3.4 I), the relation between A and C was shown as undirected, which signals a possible latent driver in the process. The implications of sub-sampling the data were more dramatic with the appearance of several spurious links.

Moreover, the causal search for the system containing causal stationarity transgression, meaning a system where distinct causal regimes govern the data, has an interesting result (Figure 3.5). The algorithm correctly finds the present causal relations but interprets them to belong to the same causal structure instead of outputting the presence of two causal regimes. This finding indicates the necessity of dividing non-stationary data into their regimes prior to the PCMCI causal structure search.

Lastly, when processing independent sub-causal processes, the PCMCI method successfully distinguished the independence of the processes as shown in Figure 3.6. Furthermore, the found causal structures are loyal to the undertaken data governance structure. The result suggests that the method can simultaneously process different systems underlying mechanisms that connect variables.

Those implications are essential, signaling the necessity of using high-quality data while using data-driven causal discovery methods in real-world applications. However, the required data assumptions are often not verifiable and met in reality. In those cases, data causal search still enables to infer causality from DAGs, but it requires a careful interpretation of CSD results relative to their data assumption foundations.

4. Case study

Having covered the theory and examined the power of the PCMCI algorithm for causal discovery on synthetic data, we move to apply the method to a real database in the field of healthcare. The dataset was selected to verify the potentials and challenges of causal discovery analysis in a real-world scenario where the method may help identify physical mechanisms from observed time series and improve the interpretability of predictive models.

4.1 Materials and Methods

In addition to the Dataset described in section 4.1.1, the materials used throughout the development of this project are described in Table 4.1.

Table 4.1: Python packages used in throughout the development of this project.

Library	Version	Description
Tigramite [37]	5.1	Causal time series analysis
Seaborn [45]	0.11.2	High-level tools for Data visualization
NumPy [46]	1.21.5	Arrays and matrix computing mathematical tools
Pandas [47]	1.4.2	Data analysis and manipulation tools
Scikit-Learn [48]	1.0.2	Machine learning tools
SciPy [49]	1.7.3	Mathematical methods and scientific computing tools
TSFEL [50]	0.1.5	Automated pre-processing and feature extraction tools for time-series

4.1.1 Dataset

Due to intensive care units enhanced capacity to gather data through monitoring of continuous or repeated measurements of patient's health on numerous parameters, we choose to work with a data subset of the MIMIC III database [51].

MIMIC III is a complete database of critical care information constituted of deidentified, multivariate high-resolution of more than 40,000 patients admitted to the ICU of the Beth Israel Deaconess Medical Center in Boston, Massachusetts. Figure 4.1 represents an overview of the subdivisions of the database. [51]. The dataset consists of an extracted subset of information from the Medical Information Mart for Intensive Care (MIMIC III) Database used in [51].

First, we completed compliance with MIMIC III regulations. To gain access to the database, researchers completed the Data or Specimens Only Research Training course from the Collaborative Institutional Training Initiative (CITI program) [53]. MIMIC III [54, 51] documented

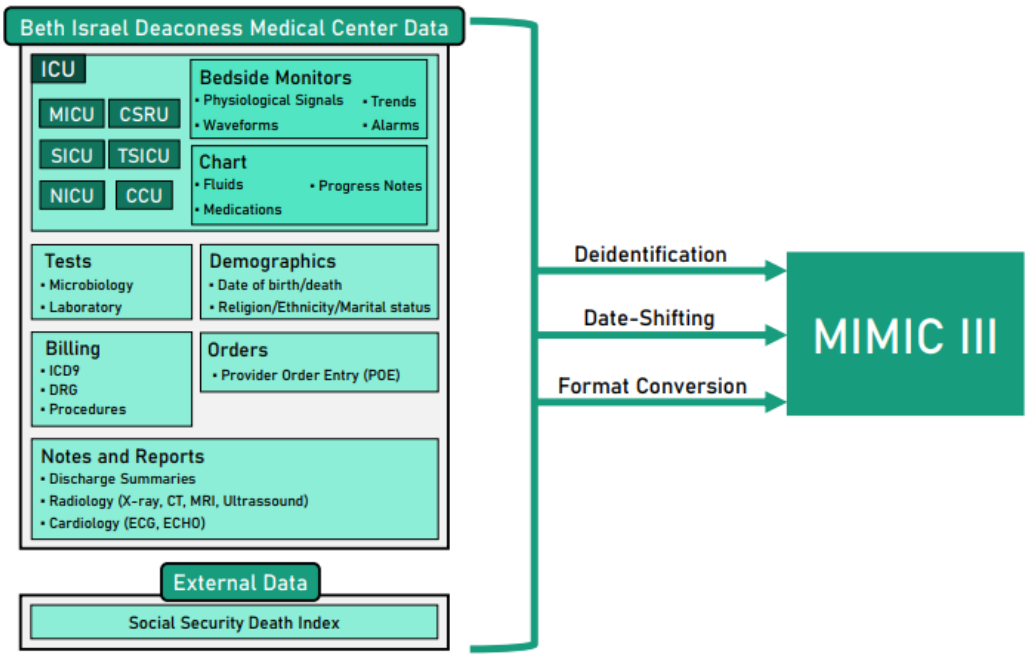

Figure 4.1: Constitution of the MIMIC III Database (Extracted from [52]).

different categories of the patient's information, physiological signals measurements, used medications, demographic data, imaging reports, and microbiology records, among others. The data were retrieved from over 40.000 patients who stayed in critical care units from six different intensive care sub-units of the Beth Israel Deaconess Medical Center: the Medical Intensive Care Unit (MICU), the Cardiac Surgery Care Unit (CSRU), the Surgical Intensive Care Unit (SICU), the Trauma Surgical Intensive Care Unit (TSICU), the Coronary Care Unit(CCU), and the Neonatal Intensive Care Unit (NICU).

The used Database subset contains records on six physiological parameter signals, respiration rate (RR), peripheral oxygen saturation (SpO2), mean blood pressure (MAP), systolic blood pressure (SBP), diastolic blood pressure (DBP), and heart rate (HR) with simultaneous information on all six of these biosignals, added to the information of administrated drugs. The detailed data extraction and fitting criteria steps are described in [52]. This selection results in a set of 597 records from the CCU, 647 from the CSRU, 832 from the MICU, 61 from the SICU, and 68 from the TSICU, totaling 2,205 ICU records acquired tough the CareVue Information System, corresponding to adult patients.

4.1.2 Dataset preprocessing

The data preprocessing described next was necessary to confidently shape the data assumptions the PCMCI algorithm was designed upon. Since we are interested in detecting causal relations here, three parameters were selected for analysis, HR, SBP, and DBP,

for their well-understood interconnected mechanisms, although the cardiovascular system is highly complex and sensitive. The resting heart rate inversely influences the cardiac stroke volume, which, in turn, drives diastolic blood pressure (Starling's law) when all other factors remain constant. Additionally, it is known that the mechanism by which diastolic blood pressure affects systolic blood pressure includes the stroke volume, the corresponding pulse pressure, and the overall peripheral resistance [55, 56].

The raw signal records of the data subset contain several biological artifacts and outliers generated in their collection. Therefore, each record had to go trough an extensive signal processing stage involving multiple data transformations, to produce clean data samples from which meaningful information may be extracted. To remove the outliers of our multivariate time series, we turn to considered acceptable thresholds for each physiological signal defined in literature. The defined values were the following: For the heart rate time series, points below 10 bpm and higher than 200 bpm were defined as outliers. For the systolic and diastolic ambulatory blood pressure, points below 10 mmHg and 200 mmHg were also defined as outliers [57].

All outliers were dealt with as follows. Instances with no more than two minutes of consecutive outliers were replaced by means of interpolation, and detected outliers instances for more than two consecutive minutes were removed. The records were then split and separated into sub-records whenever there was an identified data time gap, guaranteeing an adequate continuity of the samples. This processing step resulted in 6664 samples from the unit CCU, 10739 from the CSRU, 11295 from the MICU, 1128 from the SICU, and 669 from the TSICU. Figure 4.2 presents an example of a cleaned sample.

Next, all the samples that received any medication were eliminated, to ensure that any confounding variable did not influence the causal relationship of our variables of interest. This step resulted in 2717 samples from the unit CCU, 2428 from the CSRU, 5206 from the MICU, 454 from the SICU, and 281 from the TSICU, totaling 11086 samples not under influence of medicine.

As suggested in [25], we selected only the ones of size larger than 1000 instances or samples collected for more than 16 hours and 36 minutes. This preprocessing step resulted in the final 10 samples from the CCU, 1 sample from the CSRU, 33 samples from the MICU, and 1 sample from the SICU. No sample of the TSICU was eligible after this condition.

Since the cardiovascular system of the patients could have suffered alterations after undergoing surgical procedures, the samples of the surgical units were dismissed, and from the remaining samples, only one sample from each patient was selected. This process resulted in 4 samples of CCU and 19 of MICU for the causal analysis. Finally, all final selected samples were split into train and test subsets.

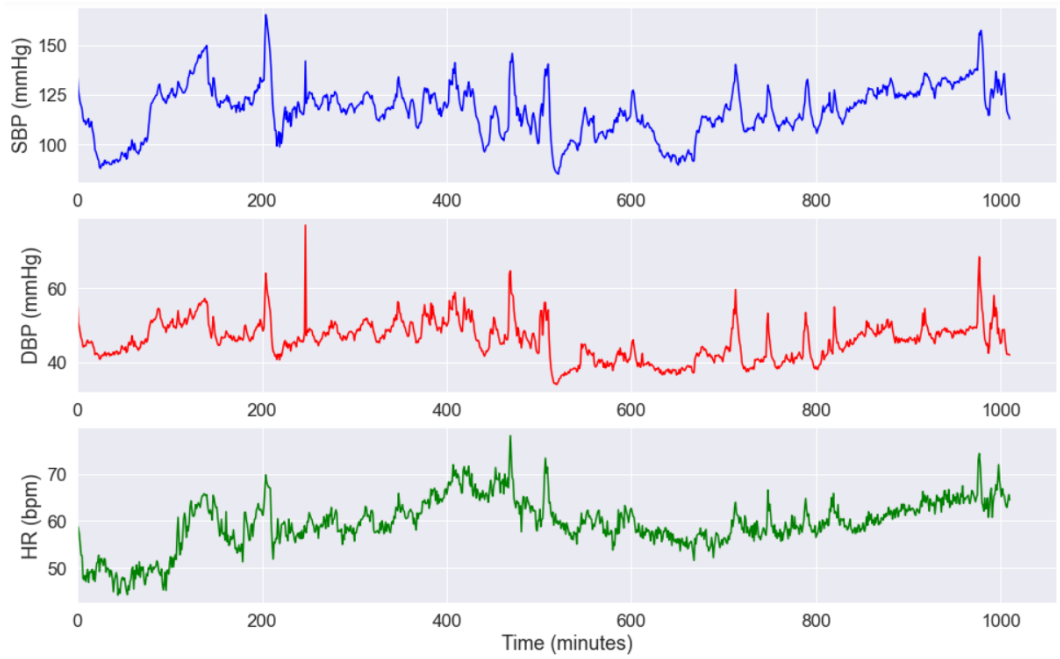

Figure 4.2: Cleaned sample after preprocessing. SBP in mmHg represented in blue, DBP in mmHg represented by red, in HR in bpm represented in green. The x-axis represent time, in minutes

Splitting the sample datasets into train and test subsets is fundamental to optimize a predictive model on a primary data source and evaluating the model performance on unknown data simulating practical applications. The defined split was as follows: the first 60% of the sample's time compose the training sets, and the final 40% compose the testing sets.

4.1.3 Causal Discovery

Given the nature of the database that contains data of patients in critical states, we postulated a great variability in their causal structure. Moreover, we understand that the sampling frequency of the data collection is too coarse for an enriched causal discovery. Nevertheless, we here investigate if a constructed causal model of every sample enhances the forecasting of their physiological parameters. With that perspective, only the sample's training subset was used in the causal discovery stage in order to derive a causal model exclusively from the data used to train our predictive model.

In the analysis, we consider a $\tau_{max} = 2$ given that the data time sampling is every minute and that the correlation of the cardiovascular parameters perished for tests longer than two minutes. The chosen significance level at which the statistical tests are restrained by the threshold to derive decisions and construct a causal graph was $\alpha = 0.05$. Since we cannot assume linear inter-dependencies in the data, we define the CMIknn as the most

suitable and primary conditional independence test used to analyze variables in(dependencies) in the data.

The computational complexity and processing time of CMKnn is relatively large when compared to the other conditional independence tests. Commonly, running k-nearest neighbor searches using k-dimensional search trees on GPU may result in poor performance due to branching and memory access inapt for GPU hardware. Consequently, the run-time of the individual CI-test and hence the run-time of constraint-based causal discovery algorithms increase with higher computational complexity [58]. The time consumption and complexity of searching the optimized causal ancestors for the predictive model using the CMKnn test were considered impractical. Consequently, we also performed the causal search with a simpler computational test, ParCorr, to later derive the predictive model from.

4.1.4 Prediction

The prediction of changes in time series data obtained from observations is extremely relevant in medical systems, especially in physiological parameters where the accurate determination of changes is particularly critical and important in supporting clinical decisions regarding the delivery of therapeutic interventions [59].

Monitoring and predicting the heart rate of ICU patients is critical for detecting and managing the heart's function irregularities at a very early stage [60]. For this reason, we defined heart rate as our variable of interest for prediction. From the constructed causal model of each sample, we extracted the optimized causal predictors, as explained in 2.4.3, then used them to develop a predictive model for our target variable. Two different approaches to model each sample training subset were taken: linear regression and k-nearest neighbors regression. The same modeling approaches were repeated without utilizing the optimized predictors, but the entire available information.

Linear regression modeling is a widely used linear approach to minimize the residual sum of squares between the observed targets in the dataset and the targets predicted by a linear approximation [61]. KNN relies on a distance-based approach. Given a sample characterized by a set of features/variables, the proximity between their instances may be defined by how similar those instances are to the values of the related features. A KNN regression uses this definition of proximity and produces a prediction of a new sample accordingly to the k observations in the same neighborhood whose characteristics are the most approximated to those of the unknown sample. [62].

Therefore, for each different step ahead of prediction, we forecast from each sample the heart rate variable four times as outlined next:

1. Linear regression utilizing uniquely the causal predictors within the analysed window size of the sample.
2. Linear regression utilizing the whole set of information within the analysed window size of the sample.
3. K-Nearest Neighbor regression utilizing uniquely the causal predictors within the analysed window size of the sample.
4. K-Nearest Neighbor regression utilizing the whole set of information within the analysed window size of the sample.

Steps ahead

Considering the average effect onset and duration of commonly administrated cardiac medications in ICU patients in general and as well as in the dataset contained in Table 4.2, we decide to analyze the performance of the forecasting in a range from 3 to 60 minutes ahead.

Table 4.2: Commonly administrated cardiac drugs in ICU patients effect time in minutes [63] [64], [65], [66], [67]

Drug	Onset	Duration
Dopamine	5	10
Epinepine	5-10	240
Norepinepine	1-2	1-2
Phenylepine	10-15	15
Vasopressin	30-60	10-20
Labetolol	5-15	120-240

Training Process

The training process refers to the identification of tendencies and patterns in the training set of the samples, allowing for the establishment of a mapping between the inputs and outputs that gets more accurate with experience.

Testing and Performance Evaluation

After the training process, when the model was fitted for the samples train sets, the predictive model is then applied to new data, the samples test sets to test their generalization

results. The predicted time series results were evaluated in comparison to their original samples with the metrics R^2, MAE, $NMRSE$, $MaxError$, $RMSE$ and MSE, previously defined in section 2.7.2. Finally, we estimated the average and standard deviation of the whole set of predictions grouped by the size of the steps ahead.

4.2 Results

4.2.1 Causal Discovery

The result of every sample PCMCI causal structure search with the CMKnn conditional independence test under the mentioned conditions in section 4.1.3 is displayed in figure 4.3. Figure 4.4 exhibits the found causal model for every sample under the same described conditions but by utilizing the ParCorr conditional independence test. In all panels, the node colors render the auto-dependency strength, and the edge colors render the cross-link strength at the lag with maximum absolute value [25]. The causal structures in Figure 4.4 were utilized to derive a prediction model specifically for the samples that generated them.

4.2.2 Prediction

The full metrics performance of forecasting the entire test set in the different sizes of steps ahead by linear regression and k-nearest neighbors are exhibited in Table 4.3 and Table 4.4, respectively. Figure 4.7 illustrates different steps ahead results of the causal forecasting in a given sample for the training and testing data set. And Figures 4.5 and 4.6 show the progression in performance of root mean square error (RMSE) in the different steps ahead.

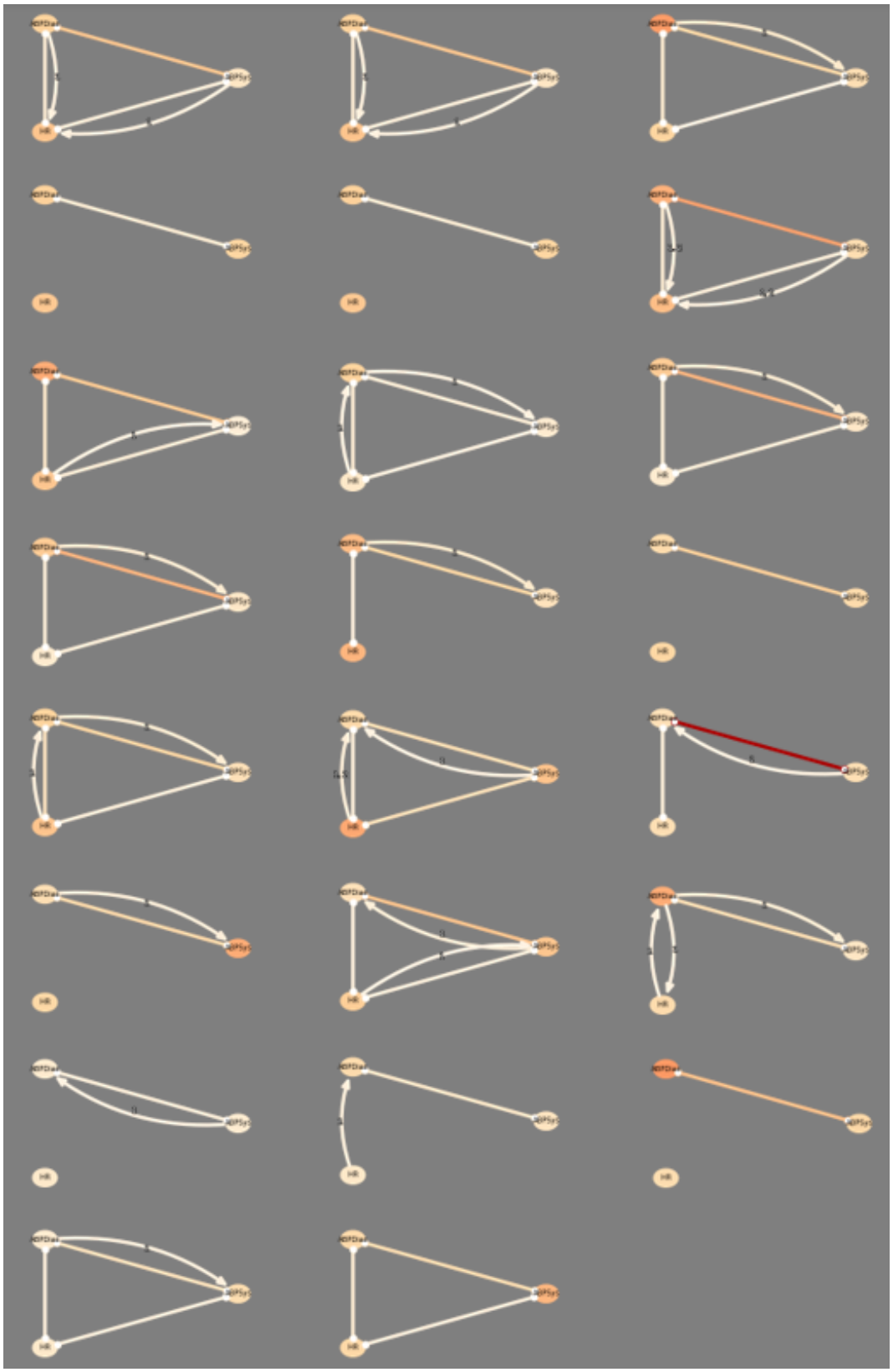

Figure 4.3: Causal structure of all analysed sample found by the PCMCI method when utilizing the CMKnn independence test.

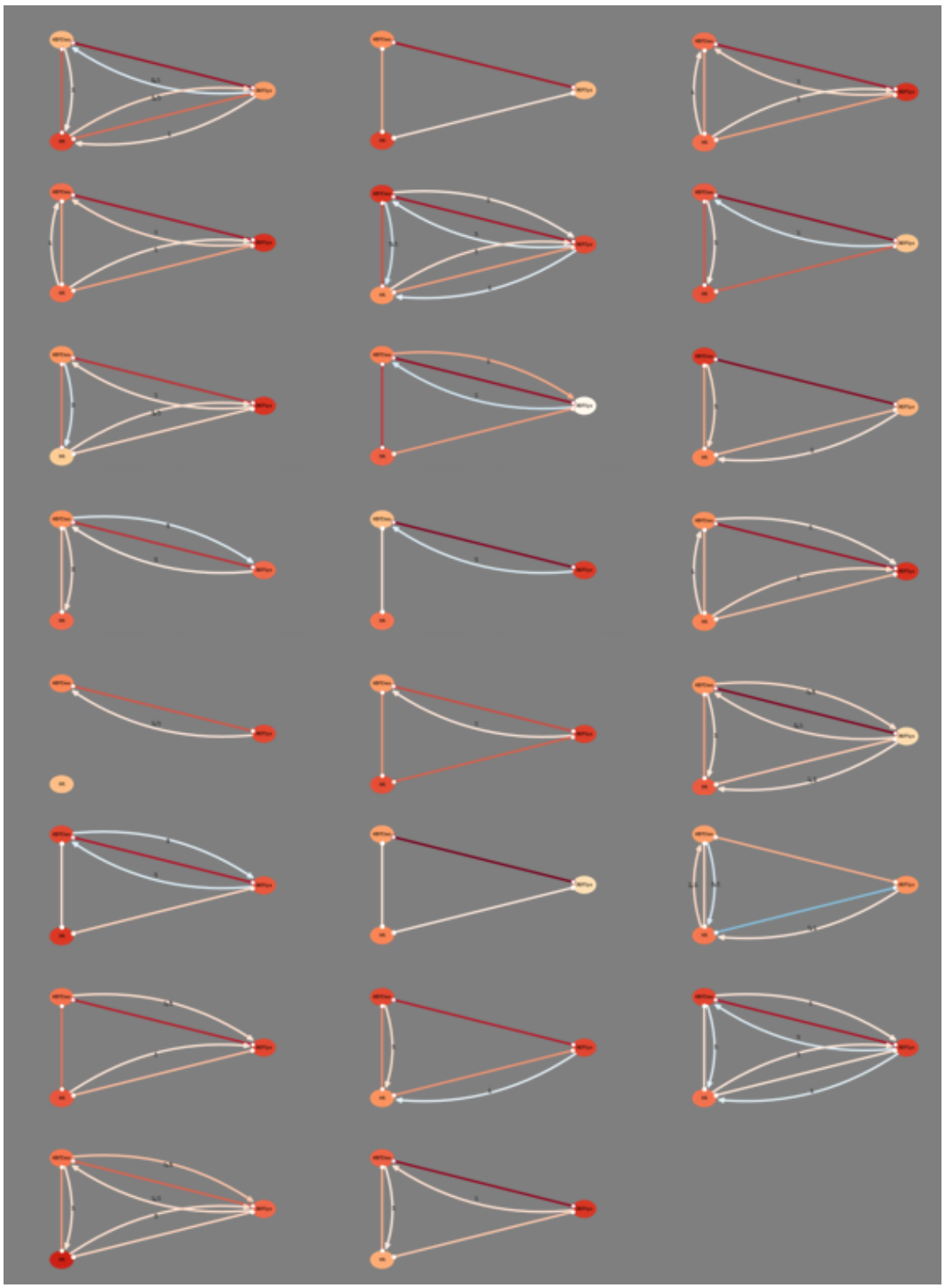

Figure 4.4: Causal structure of all analysed sample found by the PCMCI method when utilizing the CMKnn independence test.

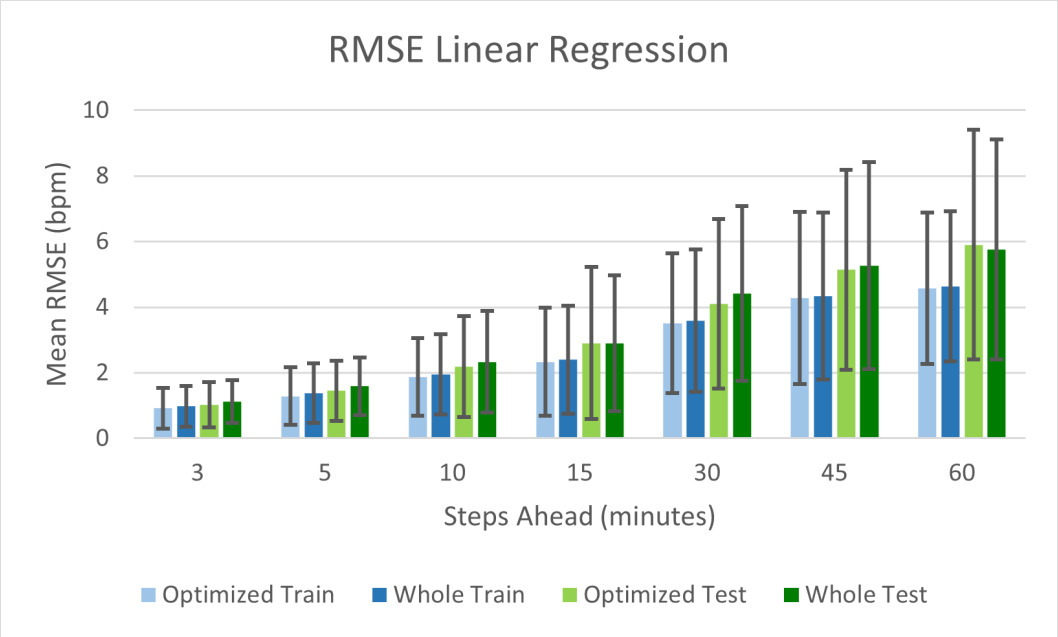

Figure 4.5: Linear Regression RMSE Progression for increased steps ahead prediction

Figure 4.6: KNN Regression RMSE Progression for increased steps ahead prediction

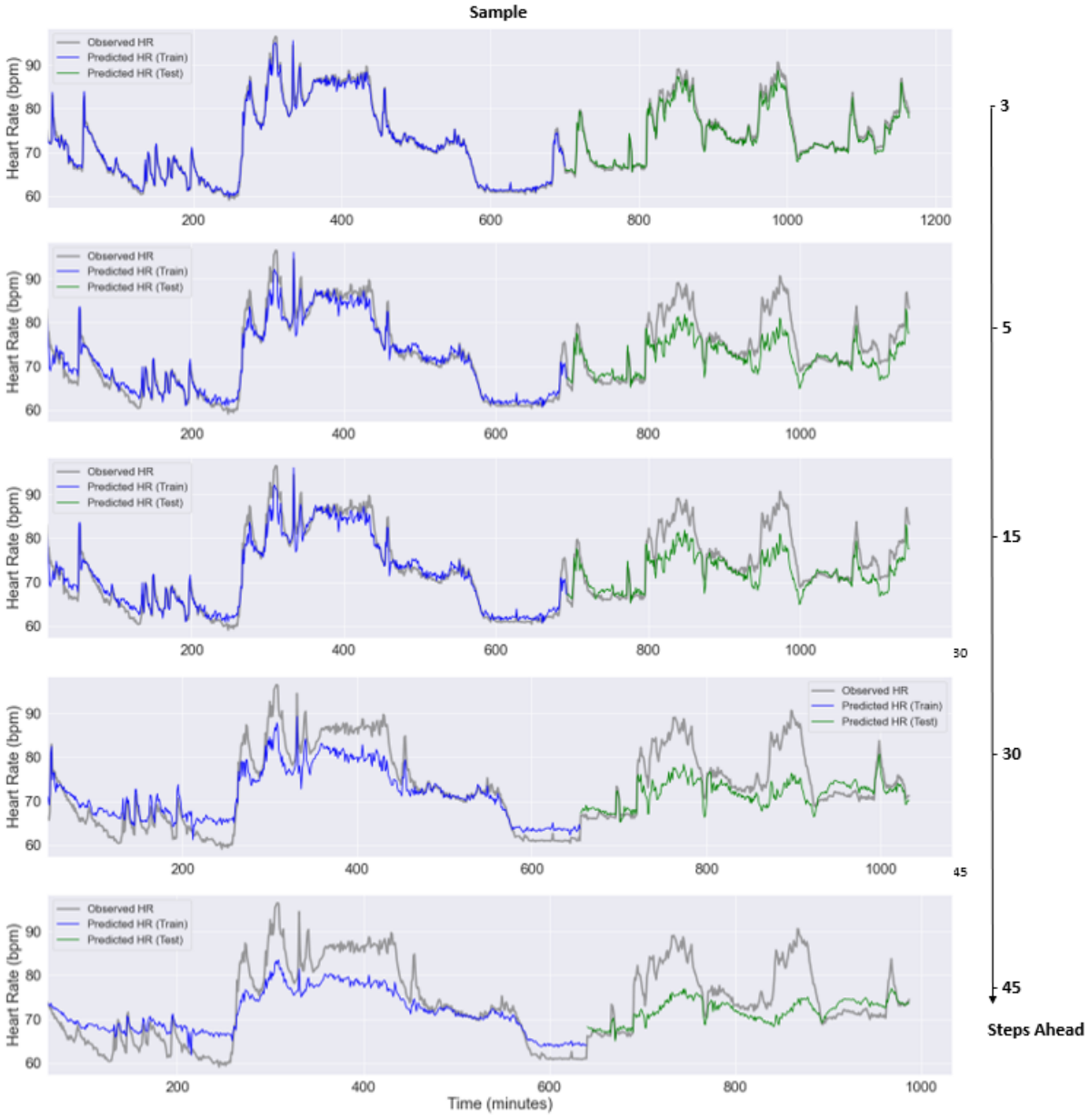

Figure 4.7: Sample Causal forecasting progression for increasing steps ahead. Original time series is represented in grey, predicted in training data represented in blue, and predicted in testing data represented in green.

Table 4.3: Forecasting results of mean and standard deviation for different steps ahead into the future by **linear regression** with the optimized causal predictors and whole predictors

			Steps Ahead						
Metric	Set	Predictors	3	5	10	15	30	45	60
R2	Train	Optimized	0.97 ± 0.04	0.94 ± 0.07	0.89 ± 0.10	0.83 ± 0.16	0.62 ± 0.28	0.46 ± 0.34	0.33 ± 0.37
		Whole	0.97 ± 0.04	0.93 ± 0.07	0.88 ± 0.10	0.82 ± 0.16	0.61 ± 0.29	0.44 ± 0.34	0.32 ± 0.33
	Test	Optimized	0.93 ± 0.09	0.87 ± 0.16	0.66 ± 0.56	0.29 ± 1.34	-0.81 ± 4.54	-2.27 ± 9.21	-3.79 ± 13.99
		Whole	0.92 ± 0.10	0.85 ± 0.17	0.63 ± 0.61	0.33 ± 1.28	-1.02 ± 4.78	-2.34 ± 9.11	-4.03 ± 15.51
MAE	Train	Optimized	0.67 ± 0.50	0.94 ± 0.70	1.38 ± 0.90	1.72 ± 1.17	2.61 ± 1.51	3.19 ± 1.66	3.39 ± 1.45
		Whole	0.71 ± 0.50	1.00 ± 0.72	1.42 ± 0.91	1.77 ± 1.19	2.69 ± 1.53	3.24 ± 1.67	3.46 ± 1.47
	Test	Optimized	0.79 ± 0.54	1.09 ± 0.70	1.69 ± 1.20	2.32 ± 1.90	3.36 ± 2.33	4.18 ± 2.67	4.86 ± 3.26
		Whole	0.85 ± 0.52	1.19 ± 0.70	1.76 ± 1.22	2.24 ± 1.67	3.56 ± 2.33	4.28 ± 2.81	4.73 ± 3.14
NMRSE	Train	Optimized	0.11 ± 0.07	0.16 ± 0.09	0.23 ± 0.11	0.28 ± 0.13	0.43 ± 0.18	0.54 ± 0.18	0.60 ± 0.17
		Whole	0.12 ± 0.07	0.17 ± 0.10	0.23 ± 0.11	0.29 ± 0.14	0.45 ± 0.17	0.54 ± 0.18	0.61 ± 0.16
	Test	Optimized	0.18 ± 0.13	0.24 ± 0.15	0.38 ± 0.30	0.54 ± 0.51	0.84 ± 0.90	1.10 ± 1.26	1.32 ± 1.58
		Whole	0.20 ± 0.13	0.27 ± 0.16	0.40 ± 0.30	0.52 ± 0.46	0.90 ± 0.89	1.12 ± 1.25	1.31 ± 1.64
MaxError	Train	Optimized	4.94 ± 3.17	6.24 ± 3.72	8.73 ± 5.21	9.77 ± 7.19	13.40 ± 7.86	16.09 ± 11.63	16.75 ± 10.72
		Whole	5.29 ± 3.13	7.10 ± 3.80	9.39 ± 5.50	10.44 ± 7.15	13.94 ± 8.41	15.83 ± 9.64	16.51 ± 10.35
	Test	Optimized	4.74 ± 4.25	6.83 ± 6.71	8.92 ± 7.68	10.26 ± 8.46	13.11 ± 7.68	16.04 ± 14.30	17.79 ± 14.73
		Whole	5.71 ± 4.50	7.88 ± 6.57	10.09 ± 8.12	11.19 ± 8.17	13.63 ± 7.87	16.03 ± 12.04	18.12 ± 14.03
RMSE	Train	Optimized	0.92 ± 0.62	1.28 ± 0.88	1.87 ± 1.19	2.33 ± 1.65	3.50 ± 2.13	4.28 ± 2.62	4.57 ± 2.30
		Whole	0.97 ± 0.62	1.38 ± 0.90	1.94 ± 1.22	2.40 ± 1.65	3.59 ± 2.17	4.34 ± 2.54	4.63 ± 2.30
	Test	Optimized	1.02 ± 0.70	1.45 ± 0.91	2.18 ± 1.53	2.90 ± 2.31	4.11 ± 2.58	5.14 ± 3.05	5.90 ± 3.50
		Whole	1.12 ± 0.65	1.58 ± 0.88	2.33 ± 1.55	2.90 ± 2.08	4.41 ± 2.66	5.26 ± 3.15	5.75 ± 3.35
MSE	Train	Optimized	1.21 ± 1.84	2.39 ± 3.90	4.86 ± 6.40	8.04 ± 12.16	16.64 ± 19.19	24.91 ± 32.31	26.00 ± 26.39
		Whole	1.32 ± 1.87	2.67 ± 4.06	5.21 ± 6.82	8.36 ± 12.15	17.40 ± 20.02	24.99 ± 29.15	26.50 ± 26.27
	Test	Optimized	1.51 ± 1.97	2.89 ± 3.15	7.02 ± 8.95	13.55 ± 20.06	23.26 ± 24.67	35.31 ± 34.53	46.56 ± 48.93
		Whole	1.65 ± 1.93	3.25 ± 3.18	7.71 ± 9.84	12.52 ± 17.78	26.21 ± 26.88	37.19 ± 38.10	43.83 ± 47.81

Table 4.4: Forecasting results of mean and standard deviation for different steps ahead into the future by **K-nearest neighbor** regression with the optimized causal predictors and whole predictors

			Steps Ahead						
Metric	Set	Predictors	3	5	10	15	30	45	60
R2	Train	Optimized	0.87 ± 0.10	0.84 ± 0.12	0.79 ± 0.13	0.76 ± 0.17	0.76 ± 0.17	0.74 ± 0.17	0.72 ± 0.19
		Whole	0.87 ± 0.10	0.86 ± 0.11	0.82 ± 0.13	0.80 ± 0.15	0.84 ± 0.13	0.83 ± 0.12	0.86 ± 0.11
	Test	Optimized	-0.61 ± 4.55	-1.11 ± 5.10	-1.73 ± 5.56	-1.92 ± 5.17	-1.97 ± 4.20	-3.10 ± 7.25	-4.48 ± 11.54
		Whole	-1.78 ± 6.06	-2.49 ± 7.34	-2.43 ± 6.19	-2.51 ± 5.18	-2.85 ± 4.72	-2.98 ± 4.07	-5.32 ± 13.26
MAE	Train	Optimized	1.36 ± 0.72	1.50 ± 0.83	1.79 ± 0.97	1.85 ± 0.99	1.87 ± 1.03	1.90 ± 1.02	1.94 ± 1.14
		Whole	1.31 ± 0.71	1.38 ± 0.76	1.53 ± 0.86	1.65 ± 0.93	1.46 ± 0.83	1.48 ± 0.84	1.30 ± 0.79
	Test	Optimized	2.88 ± 1.97	3.46 ± 2.11	4.27 ± 2.36	4.77 ± 2.55	5.22 ± 2.59	5.75 ± 2.86	6.42 ± 3.81
		Whole	3.69 ± 2.20	4.29 ± 2.49	4.71 ± 2.32	5.13 ± 2.54	5.69 ± 2.46	6.08 ± 2.91	6.73 ± 3.47
NMRSE	Train	Optimized	0.23 ± 0.10	0.25 ± 0.12	0.29 ± 0.12	0.31 ± 0.14	0.32 ± 0.14	0.32 ± 0.14	0.34 ± 0.16
		Whole	0.22 ± 0.10	0.23 ± 0.11	0.26 ± 0.12	0.27 ± 0.12	0.25 ± 0.12	0.25 ± 0.11	0.23 ± 0.11
	Test	Optimized	0.73 ± 0.84	0.88 ± 0.89	1.08 ± 0.97	1.19 ± 0.94	1.27 ± 0.81	1.44 ± 1.12	1.61 ± 1.41
		Whole	0.99 ± 1.05	1.14 ± 1.18	1.22 ± 1.06	1.30 ± 0.98	1.43 ± 0.92	1.49 ± 0.84	1.73 ± 1.49
MaxError	Train	Optimized	13.83 ± 10.50	14.03 ± 10.37	14.55 ± 10.14	14.38 ± 9.93	14.48 ± 9.12	15.48 ± 10.84	14.02 ± 10.85
		Whole	13.50 ± 9.54	12.72 ± 7.13	12.71 ± 6.33	13.78 ± 10.12	12.30 ± 8.85	13.42 ± 10.29	11.76 ± 10.60
	Test	Optimized	17.43 ± 15.79	18.77 ± 15.79	20.68 ± 15.04	20.75 ± 14.75	22.21 ± 14.18	22.78 ± 14.58	24.32 ± 15.23
		Whole	18.85 ± 15.31	20.20 ± 15.07	20.90 ± 14.94	20.75 ± 13.93	23.31 ± 13.99	23.81 ± 14.40	25.40 ± 14.88
RMSE	Train	Optimized	2.13 ± 1.15	2.33 ± 1.34	2.76 ± 1.59	2.79 ± 1.41	2.77 ± 1.45	2.91 ± 1.55	2.90 ± 1.73
		Whole	2.08 ± 1.10	2.14 ± 1.09	2.37 ± 1.20	2.57 ± 1.43	2.24 ± 1.17	2.33 ± 1.28	2.09 ± 1.53
	Test	Optimized	3.88 ± 2.31	4.54 ± 2.51	5.44 ± 2.58	5.91 ± 2.75	6.44 ± 2.82	7.03 ± 3.15	7.77 ± 4.07
		Whole	4.80 ± 2.62	5.43 ± 2.80	5.93 ± 2.58	6.34 ± 2.72	7.01 ± 2.72	7.49 ± 3.23	8.18 ± 3.76
MSE	Train	Optimized	5.80 ± 6.22	7.17 ± 8.58	10.01 ± 12.41	9.69 ± 9.55	9.68 ± 10.05	10.76 ± 10.57	11.24 ± 13.11
		Whole	5.49 ± 5.78	5.71 ± 5.83	6.99 ± 7.20	8.57 ± 9.89	6.34 ± 7.25	7.01 ± 8.03	6.62 ± 11.74
	Test	Optimized	20.13 ± 24.48	26.65 ± 30.67	35.97 ± 34.93	42.19 ± 41.70	49.10 ± 46.23	58.90 ± 51.09	76.16 ± 74.19
		Whole	29.66 ± 32.32	37.04 ± 36.87	41.56 ± 35.08	47.30 ± 42.74	56.23 ± 44.31	66.00 ± 58.36	80.49 ± 67.29

4.3 Discussion

4.3.1 Causal Discovery

The results in Figure 4.4 and 4.3 confirm a great variability in the causal structure of patients in critical states.

The causal search of the PCMCI coupled with the CMKnn independence in many samples identifies an absence of causal links between the physiological parameters, which implies statistical independence between the variables by the faithfulness assumption. Instead, the causal structures found by coupling the ParCorr independence test detect more dependence between the parameters but generally fail to determine the causal relations' direction.Similar to the results obtained simulating causal discovery in a non-sufficient dataset (3.1.1), the lack of directionality detection may indicate the presence of an unobserved confounding variable.

Moreover, the causal structures found with the CMKnn independence test (Figure 4.3) exhibit weaker causal links between the variables than those found in the ParCorr search results (Figure 4.4). This result may also signal the presence of latent confounding affecting the observed variables.

The sampling rate of the dataset highly influences causal detection. Here, with the knowledge of the ground truth rhythmic causal structure of the cardiovascular system, we believe that the analyzed samples violate the causal sufficiency assumption by the reduced sampling frequency. A data set sampled in heartbeats would surely deliver more information in the methods of mathematical analysis of dependencies.

In summary, those results provide insights into the fundamental challenge of the causal sufficiency assumption in applying exclusively observed time-series data to data-driven constraint-based causal discovery methods.

4.3.2 Prediction

As evidenced by Figures 4.5, 4.6, and 4.7 for all predictive models, the prediction performance deteriorates as the prediction horizon increases. We also observed decay of performance between the training set and testing set. Figure 4.7 of a sample prediction progression conveys this deterioration more intuitively and visibly. This trend is standard in predictive models for both simulated and real-world problems, and it is due to the fact that predictive modeling is based on the current values of the data. Thus, by increasing the multi-step ahead, the information gap increases [68]. The practical implication in a clinical setting is that while short horizon predictions are significantly trustworthy, farthest predictions are less reliable.

Comparing Figures 4.5 and 4.6, the predictions horizon progression indicates a better overall performance of the linear regression over KNN. This result may be due to a limitation of our KNN training step since the hyperparameter number of neighbors was not optimized. Moreover, the prediction performance between the training set and testing set has a stronger deterioration on KNN regression than on linear regression, suggesting that the KNN training model expressed larger overfitting. In addition, within the KNN modeling, the predictions yielding the whole predictors suffered from a larger training overfitting than the causal predictions due to low useful information data aggregation, as theorized by Runge et al. in [69] and detailed in 2.4.3.

Clinicians must be confident that the predictive models are trustworthy to sustain decisions on the best clinical pathway for each patient. Health care is a domain with unique challenges that require reliable information to base decisions upon, for the implications may impact the health state of patients [70]. Usual consternation of predictive models are:

1. The lack of model transfer ability to new and distinct data.

2. The potential algorithmic bias or lack of reliability in the models.

The results of this work reveal the potential of causal discovery in tackling the mentioned issues. Overall, using exclusively optimal causal predictors resulted in higher predictive performance levels than all available information on forecasting the variable in all measured metrics. The KNN predictive modeling using the whole predictors faced higher over-fitting, as shown in Figure 4.6, with a higher decline in performance from training data to testing data, meaning that the predictions with exclusive causal predictors enjoyed better generalizability to handle new previously unseen data, surpassing issue 1. In association with better forecasting results, the predictions based on causal mechanisms lean toward insights into how the algorithmic modeling produced those predictions. Those advantages create a more trustworthy methodology, especially essential in the health care domain, excelling issue 2.

5. Conclusion

In the following, we provide an overview of key findings and achievements to conclude this dissertation.

In numerous investigations, searching for causal relationships between variables is critical to understand the connected mechanisms within a system. This knowledge improves the design of interventions able to generate appropriate outcomes. Thus, a tool to recover causal networks directly from observational data is precious. CSD methods suited to detect lagged causality overcome cross-sectional data shortcomings by leveraging temporal causal prioritization and there is an increasing practice of time series data collection. Moreover, the illustration of causal links with direct acyclic graphs, obtained from statistical (in)dependencies in multivariate time series data, improves the interpretability of causal structures [26].

A careful interpretation is nevertheless required since the method relies on several strong and often unverifiable assumptions, commonly not met in real-world data. Life sciences data is usually composed of artifacts and contain missing values that may violate the method requirements. When relying exclusively on observational data, it is challenging to presume that all relevant variables have been collected and observed in the appropriate time scale. However, in contexts where the problem of confounding is mitigated—such as randomized controlled trials—the combination of observational and experimental data can optimize the discovery of causal relationships. Moreover, the causal analysis of high-dimensional data can be a crucial advantage in drawing causal inferences without the need of interventions [16, 71]. Furthermore, recent methods already support causal discovery that does not rely on the assumption of causal sufficiency, and it shows that the young science of causality research is progressing rapidly.

This work can serve as an overview of constraint-based method functionalities containing the various essential theoretical aspects of causality research, a significantly new domain of science. Giving the described goals in 1.2 this work achieved: In Chapter 2 to highlight the progress of constraint-based causal detection methods from PC to PCMCI algorithm, to explain the causal search procedure, to show how causal networks are represented, and their practical challenges. In addition, we have examined practical applications where causal detection have led to interesting findings.

Moreover, in Chapter 3, by searching for the causal structure of simulated data with different characteristics, we provided evidence of the implications of violating causal discovery methods assumptions. Finally, in our case study (Chapter 4), we confirmed the challenges of uncovering causal structure from real-world observational data. Challenges such as to presume that no unobserved variables, directly or indirectly, influenced any of the

analyzed dataset's variables, and that the sampling frequency was sufficient to evidence their causal effect, in addition to computational complexity and time-consuming aspect of causal search, mainly when using the CMKnn test.

These practical problems in causal discovery convey two essential aspects for optimizing causal research. First is the necessity of the researcher's expertise to identify all significant variables to be included in the causality discovery analysis minimizing selection bias. Second, high-quality data collection is necessary to minimize measurement error and achieve adequate sampling frequency.

Our work on the prospects of using a causal structure in prediction models for actual healthcare data is exceptional in the literature and yields remarkable findings. Using causal ancestors to predict the heart rate variable resulted in better predictive performance than employing all variable associations. Moreover, models based on causality increase the explanatory and generalizability power of predictive algorithms. Those findings indicate that a mature causality theory can overcome some limitations of classical machine learning methods, as envisioned by Judea Pearl (1.1).

Finally, the combined achievements from the different studies of this work show that the benefits of causal discovery frameworks, along with high-quality Big Data, improve the interpretability of complex systems. CSD methods can be particularly beneficial in an exploratory domain, such as the search for causal relationships for further investigation and validation. Furthermore, moving from an associative and purely predictive machine learning perspective to a causal-informative model improve algorithm's robustness and generalizability. Although one must be cautious about inferring causality from causal structural modeling, CSD methods raise efforts to boost scientific progress.

5.1 Limitations and Future work

Given the knowledge and limitations gained in this work, several developments may be taken towards progress.

The theoretical review of this work centered exclusively on constraint-based causal detection methods. A more comprehensive study of the different CSD methods, comparing their performance in different scenarios, would allow researchers to select the most appropriate method for each data feature based on their advantages and limitations.

In our case study implementation, optimizing the hyperparameters of the predictive model could improve heart rate prediction results and provide a fairer comparison between linear and KNN regression models. In addition, other classical machine learning algorithms for prediction could further exploit the use of causal antecedent predictors.

The case study focused primarily on causal discoveries on an individual level, and the prospects of utilizing causal structure knowledge in prediction. Implementing causal inference in the same data context would provide insights into cardiovascular treatments' causal effects on the physiological parameters of ICU patient's.

In addition, the causal result for the variables selected in our case study strongly suggests the possible presence of confounding variables affecting the analyzed system. The exploitation of additional variables within this perspective combined with a study on the strength of the variable's causal links could give insights into the dynamic relationship differences of the cardiovascular system in distinct health conditions. For example, the study by Reidl et al. indicates a significant increase in the influence of respiratory rate on cardiac rate in pregnant women with pre-eclampsia compared with healthy ones [72].

Finally, we hope for a future project to address the challenges of generating a group-level causal structure from cross-patient data [36, 41] in order to analyze the possibility of searching for a causal model that can interpret cardiovascular mechanisms for the general population.

5. References

[1] A. Hyttinen, "Discovering causal relations in the presence of latent confounders," 2013, Ph.D. dissertation, University of Helsinki. [Cited on page 1]

[2] X. Shen, "Challenges and opportunities with causal discovery algorithms: Application to alzheimer's pathophysiology," *Scientific reports*, vol. 10, 2020. [Cited on pages 1, 2, 11, 12, 15, 25, and 26]

[3] P. S. Clark Glymour, Kun Zhang, "Review of causal discovery methods based on graphical models," 2019. [Cited on pages 1, 2, 12, 13, 18, 19, 20, 21, 23, 24, and 25]

[4] D. M. Judea Pearl, *The book of why*. Penguin Books, 2019. [Cited on page 2]

[5] K. Singh, G. Gupta, V. Tewari, and G. Shroff, "Comparative Benchmarking of Causal Discovery Techniques," *arXiv:1708.06246 [cs, stat]*, Sept. 2017. arXiv: 1708.06246. [Cited on page 2]

[6] P. Bhandari, "Correlation vs. causation: Difference, designs and examples," 2021. [Cited on pages 2, 5, 6, and 9]

[7] A. Papana, "Connectivity analysis for multivariate time series: Correlation vs. causality," vol. 23, no. 12. [Cited on pages 2, 6, and 7]

[8] T. B. Murdoch and A. S. Detsky, "The Inevitable Application of Big Data to Health Care," *JAMA*, vol. 309, pp. 1351–1352, 04 2013. [Cited on page 3]

[9] J. Runge, "Causal network reconstruction from time series: From theoretical assumptions to practical estimation," vol. 28, no. 7, p. 075310. [Cited on pages 3, 16, 17, and 27]

[10] H. Chang, "Nicholas cage - spurious pool savior?," Jun 2021. [Cited on page 5]

[11] J. Sprenger and N. Weinberger, "Simpson's Paradox," in *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Metaphysics Research Lab, Stanford University, Summer 2021 ed., 2021. [Cited on pages 6 and 7]

[12] M. et al., "Causal inference for time series analysis: Problems, methods and evaluation," [Cited on pages 7, 12, and 13]

[13] S. L. Bressler and A. K. Seth, "Wiener–granger causality: A well established methodology," *NeuroImage*, vol. 58, no. 2, pp. 323–329, 2011. [Cited on pages 7 and 8]

[14] "Markov property - an overview | ScienceDirect topics." [Cited on page 8]

[15] "PubMed." [Cited on page 9]

[16] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Muñoz-Marí, E. H. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Schölkopf, P. Spirtes, G. Sugihara, J. Sun, K. Zhang, and J. Zscheischler, "Inferring causation from time series in earth system sciences," vol. 10, no. 1, p. 2553. [Cited on pages 9, 23, and 52]

[17] R. Guo, L. Cheng, J. Li, P. R. Hahn, and H. Liu, "A Survey of Learning Causality with Data: Problems and Methods," *ACM Computing Surveys*, vol. 53, pp. 1–37, July 2021. arXiv: 1809.09337. [Cited on page 10]

[18] P. Spirtes and K. Zhang, "Causal discovery and inference: concepts and recent methodological advances," vol. 3, no. 1, p. 3. [Cited on pages 10 and 17]

[19] A. Abadie, A. Diamond, and J. Hainmueller, "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 493–505, 2010. [Cited on page 11]

[20] I. Bica, A. M. Alaa, and M. van der Schaar, "Time series deconfounder: Estimating treatment effects over time in the presence of hidden confounders." [Cited on page 11]

[21] S. Talebi, "Causal discovery." [Cited on page 11]

[22] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Muñoz-Marí, E. H. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Schölkopf, P. Spirtes, G. Sugihara, J. Sun, K. Zhang, and J. Zscheischler, "Inferring causation from time series in earth system sciences," vol. 10, no. 1, p. 2553. [Cited on pages 11, 17, 23, and 24]

[23] J. M. Ogarrio, P. Spirtes, and J. Ramsey, "A hybrid causal search algorithm for latent variable models," vol. 52, pp. 368–379, 06–09 Sep 2016. [Cited on page 12]

[24] P. Spirtes and C. Glymour, "An algorithm for fast recovery of sparse causal graphs," vol. 9, no. 1, pp. 62–72. [Cited on page 12]

[25] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, "Detecting and quantifying causal associations in large nonlinear time series datasets," vol. 5, no. 11, p. eaau4996. [Cited on pages 12, 14, 15, 19, 21, 22, 23, 24, 39, and 43]

[26] P. Spirtes, C. Glymour, R. Scheines, S. Kauffman, V. Aimale, and F. Wimberly, "Constructing bayesian network models of gene expression networks from microarray data," [Cited on pages 12 and 52]

[27] P. Spirtes, C. Glymour, and R. Scheines, "Discovery algorithms without causal sufficiency," in *Causation, Prediction, and Search* (P. Spirtes, C. Glymour, and R. Scheines, eds.), Lecture Notes in Statistics, pp. 163–200, Springer. [Cited on page 12]

[28] J. Ramsey, M. Glymour, R. Sanchez-Romero, and C. 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," vol. 3, no. 2, pp. 121–129. [Cited on page 12]

[29] A. Hauser and P. Bühlmann, "Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs." [Cited on page 12]

[30] D. Kalainathan, O. Goudet, I. Guyon, D. Lopez-Paz, and M. Sebag, "Structural agnostic modeling: Adversarial learning of causal graphs." [Cited on page 12]

[31] P. Bühlmann, J. Peters, and J. Ernest, "CAM: Causal additive models, high-dimensional order search and penalized regression," vol. 42, no. 6, pp. 2526–2556. [Cited on page 12]

[32] O. Goudet, D. Kalainathan, P. Caillou, I. Guyon, D. Lopez-Paz, and M. Sebag, "Causal generative neural networks." [Cited on page 12]

[33] A. R. Nogueira, A. Pugnana, S. Ruggieri, D. Pedreschi, and J. Gama, "Methods and tools for causal discovery and causal inference," vol. 12, no. 2. [Cited on pages 13, 28, and 29]

[34] C. K. Assaad, E. Devijver, and E. Gaussier, "Survey and evaluation of causal discovery methods for time series," vol. 73, pp. 767–819. [Cited on pages 14, 16, and 17]

[35] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman, *Causation, Prediction, and Search*. MIT Press. [Cited on page 17]

[36] S. Saetia, N. Yoshimura, and Y. Koike, "Constructing brain connectivity model using causal network reconstruction approach," vol. 15, p. 619557. [Cited on pages 17, 27, 28, and 54]

[37] J. Runge, "Jakobrunge/tigramite: Tigramite is a python package for causal inference with a focus on time series data.." `https://github.com/jakobrunge/tigramite`. [Cited on pages 18 and 37]

[38] C. H. Lee and H.-J. Yoon, "Medical big data: promise and challenges," vol. 36, no. 1, pp. 3–11. [Cited on page 24]

[39] K. Zhang, M. Gong, J. Ramsey, K. Batmanghelich, P. Spirtes, and C. Glymour, "Causal discovery in the presence of measurement error: Identifiability conditions." [Cited on pages 24 and 25]

[40] D. J. Stekhoven, I. Moraes, G. Sveinbjörnsson, L. Hennig, M. H. Maathuis, and P. Bühlmann, "Causal stability ranking," *Bioinformatics*, vol. 28, pp. 2819–2823, 09 2012. [Cited on pages 26 and 27]

[41] A. Duggento, L. Passamonti, G. Valenza, R. Barbieri, M. Guerrisi, and N. Toschi, "Multivariate granger causality unveils directed parietal to prefrontal cortex connectivity during task-free MRI," vol. 8, no. 1, p. 5571. [Cited on pages 27 and 54]

[42] J. F. Huckins, A. W. DaSilva, E. L. Hedlund, E. I. Murphy, C. Rogers, W. Wang, M. Obuchi, P. E. Holtzheimer, D. D. Wagner, and A. T. Campbell, "Causal factors of anxiety and depression in college students: Longitudinal ecological momentary assessment and causal analysis using peter and clark momentary conditional independence," vol. 7, no. 6, p. e16684. [Cited on page 28]

[43] H. M and S. M.N, "A review on evaluation metrics for data classification evaluations," vol. 5, no. 2, pp. 01–11. [Cited on page 30]

[44] "11.2 vector autoregressive models var(p) models: Stat 510." [Cited on page 31]

[45] M. L. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021. [Cited on page 37]

[46] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, Sept. 2020. [Cited on page 37]

[47] W. McKinney *et al.*, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, vol. 445, pp. 51–56, Austin, TX, 2010. [Cited on page 37]

[48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau,

M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Cited on page 37]

[49] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. [Cited on page 37]

[50] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "Tsfel: Time series feature extraction library," *SoftwareX*, vol. 11, p. 100456, 2020. [Cited on page 37]

[51] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," vol. 3, no. 1, p. 160035. [Cited on page 37]

[52] B. Ribeiro, "Machine learning for the early detection of acute episodes in intensive care units:," Portugal, 2021. [Cited on page 38]

[53] "Research, ethics, and compliance training | CITI program." [Cited on page 37]

[54] Johnson, Alistair, Pollard, Tom, and Mark, Roger, "MIMIC-III clinical database." Type: dataset. [Cited on page 37]

[55] T. Seres, "Chapter 34 - heart failure," in *Anesthesia Secrets (Fourth Edition)* (J. Duke, ed.), pp. 236–243, Philadelphia: Mosby, fourth edition ed., 2011. [Cited on page 39]

[56] A. Vander, J. Sherman, and D. Luciano, "Human physiology: the mechanisms of body function. mcgraw-hill," 1998. [Cited on page 39]

[57] J. Lee and R. G. Mark, "An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care," *BioMedical Engineering OnLine*, vol. 9, no. 1, p. 62, 2010. [Cited on page 39]

[58] C. Hagedorn, C. Lange, J. Huegle, and R. Schlosser, "Gpu acceleration for information-theoretic constraint-based causal discovery," in *Proceedings of The KDD'22 Workshop on Causal Discovery* (T. D. Le, L. Liu, E. Kıcıman, S. Triantafyllou, and H. Liu, eds.), vol. 185 of *Proceedings of Machine Learning Research*, pp. 30–60, PMLR, 15 Aug 2022. [Cited on page 41]

[59] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," vol. 6, no. 1, p. 54. [Cited on page 41]

[60] K. T. Corley and C. M. Marr, "Cardiac monitoring in the icu patient," *Clinical Techniques in Equine Practice*, vol. 2, no. 2, pp. 145–155, 2003. Adult ICU. [Cited on page 41]

[61] "1.1. linear models." [Cited on page 41]

[62] A. Teixeira-Pinto, *2 K-nearest Neighbours Regression | Machine Learning for Biostatistics*. [Cited on page 41]

[63] "Intropin (dopamine) dosing, indications, interactions, adverse effects, and more." [Cited on page 42]

[64] "EpiPen, auvi-q (epinephrine) dosing, indications, interactions, adverse effects, and more." [Cited on page 42]

[65] "Levarterenol, levophed (norepinephrine) dosing, indications, interactions, adverse effects, and more." [Cited on page 42]

[66] "Biorphen, vazculep (phenylephrine IV) dosing, indications, interactions, adverse effects, and more." [Cited on page 42]

[67] "Vasostrict, ADH (vasopressin) dosing, indications, interactions, adverse effects, and more." [Cited on page 42]

[68] R. Chandra, S. Goyal, and R. Gupta, "Evaluation of deep learning models for multi-step ahead time series prediction," vol. 9, pp. 83105–83123. [Cited on page 50]

[69] J. Runge, R. V. Donner, and J. Kurths, "Optimal model-free prediction from multivariate time series," vol. 91, no. 5, p. 052909. [Cited on page 51]

[70] A. F. Markus, J. A. Kors, and P. R. Rijnbeek, "The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies," *Journal of Biomedical Informatics*, vol. 113, p. 103655, 2021. [Cited on page 51]

[71] M. J. Vowels, N. C. Camgoz, and R. Bowden, "D'ya like dags? a survey on structure learning and causal discovery," *ACM Comput. Surv.*, mar 2022. Just Accepted. [Cited on page 52]

[72] M. Riedl, A. Suhrbier, H. Stepan, J. Kurths, and N. Wessel, "Short-term couplings of the cardiovascular system in pregnant women suffering from pre-eclampsia," vol. 368, no. 1918, pp. 2237–2250. [Cited on page 54]