International Journal of Current Science Research and Review

ISSN: 2581-8341

Volume 06 Issue 11 November 2023

DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789

IJCSRR @ 2023

A Review of Causal Identifiability Techniques across Different Observational Datasets

Gabriel Terna Ayem[1*], Salu George Thandekkattu[2], Augustine Shey Nsang[3]

[1, 2, 3]Computer Science Department, School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria.

Orcid: [0000-0001-9889-0184][1], [0000-0003-3957-3554][2]

ABSTRACT: We present an aggregation of the causal identifiability solutions techniques and their assumptions as advanced in extant literatures with datasets of odd origins, which do not necessarily conform to the independent and identically distributed (i.i.d) dataset, multinomial datasets and the Gaussian datasets settings; alongside their concomitant assumptions. The transformation process in data generation can sometimes be a desideratum of datasets of the following forms: linear and non-Gaussian, nonlinear & non-Gaussian, datasets with missing values, datasets tainted with selection biases, datasets with whose variables forms cycles, datasets with heterogeneous/nonstationary variables, datasets with confounding or latent variables, time-series datasets, deterministic datasets, etc. The study begins proper in section 2 after the introduction with the basic background into the concept of causality with observational data. The concept of graph as an embodiment of the background knowledge with structural causal model (SCM) is explicated in section 3; followed by the basic assumptions employed especially with common observational data settings in section 4. An exposition into the categorization of the algorithms used in causality is presented in section 4. Section 5 aggregates and expounds the causal identifiability techniques and their associated assumptions athwart varying datasets; which is the crux of the study and a recapitulation of same is presented in table 1. This study's main contribution is to present an aggregate review of the causal techniques and their assumptions across different data settings especially in data settings of odd origins, as reviews such as this are grossly lacking in extant literatures.

KEYWORDS: Causality, causal identifiability techniques, observational datasets

INTRODUCTION

From time immemorial till date, human actions, processes and indeed scientific explorations have been predicated on the premise of cause and effect. In the primordial era, the savaged and primitive man sought ways to articulate and uncover this phenomenon of cause and effects; and not having equipment, enough facts or the sine-quo-non to ascertain this phenomenon of knowing what actions (causes) that produces the effects especially in incidences that were agonizing to him such as certain ebullitions of some sicknesses concomitant with mysterious deaths. Thus, the ability to know the right action to influence his environment or predict his future made man an idiosyncratic specie from the rest of the animals. Thus, driving the savaged man from his initial state of higgledy-piggledy to embrace the practice of magic, astrology and certain fetish ways to achieve the causation phenomenon in order to overcome his bewildered state. Gradually, as societies evolved and advanced, and mankind himself advanced from his primitive and savaged state to his current state of scientific and technological advancement. Thus, establishing his hegemony on earth over and above every other specie; the same motives of trying to influence his environment and predict his future still stands. Nonetheless the methods of achieving it have evolved; as magic arts wanes to scientific logic, and astrology metamorphosed to astronomy and other technological innovations such as computer predictions, simulations etc., became the modern genies that are aberrations from the fetish ways of predicting the future. Albeit, in this current era, the science of trying to ascertain causality or causation in human processes and actions is still a daunting and a nontrivial task; as the traditional scientific way of ascertaining this act is resident with the randomized controlled experiment or randomized controlled trial (RCT) method. This RCT method and idea is credited to Fisher [1]. Thus, this standard framework for causal discovery known as RCT always involves setting some (usually half) of the sampled population of study and given them a treatment (an intervention) under the same conditions, while the second half of the study population is left untreated (not intervened on) or controlled under the same or similar conditions, in order to slay any possible

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 06 Issue 11 November 2023
DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789
IJCSRR @ 2023

confounding or lurking variable, which is often the factor that jeopardizes a proper juxtaposition of these two sampled population in the RCT experiments. As fascinating as this method of RCT is, there are events and circumstances that makes this kind of experiments too expensive, infeasible or even unethical to perform. A good instance is to perform a RCT on a hypothesize query that seeks to uncover the health benefits, or otherwise of smoking on a certain population. This is an unethical experiment to conduct under RCT, because it would involve setting half of the population under review to smoke (treated) and the other not to smoke (controlled). Hence, with this obstacles posed by RCT, many researchers have resorted to the discovery and inferring of causal structures from purely observational dataset, or from a combination of both data and RCT [2, 3].

However, in spite of the successes recorded by causal identifiability with observational data especially datasets that are independent and identically distributed (i.i.d), a lot of other datasets setting exists that are not generated and contrived from the i.i.d perspective. Data settings such as: time-series, deterministic, feedbacks, heterogeneous/nonstationary, missing value, measurement error, selection bias etc., that are still obfuscated and painted in shades of grays when it comes to causality in observational data. Hence, extensive research works in these areas is a desideratum for a panoramic view of the causal phenomenon in observational datasets.

1.1 Study Contributions

The major contribution of this work is to collate and succinctly present from extant literatures the few various solutions advanced by researchers in datasets which do not necessarily conform to the independent and identically distributed (i.i.d) dataset, multinomial and the Gaussian settings; alongside their concomitant assumptions, i.e., datasets of linear and non-Gaussian, nonlinear & non-Gaussian, datasets with missing values, datasets tainted with selection biases, datasets with whose variables forms cycles, datasets with heterogeneous/nonstationary variables, datasets with confounding and latent variables, time-series datasets, deterministic datasets, etc. the study of these odd datasets settings when it comes to causal identification with observational dataset will help researchers both new and old in this field to have a working understanding of the methods and techniques involved in determining causality in these divert dataset settings; as comprehensive reviews on the aggregation of different solutions types for different dataset settings as presented in this work is grossly inadequate as far as we can search.

1.2 Structure and Overview

In order to paint a clear picture of the study, we try to give a concise but elaborate concept of causal modeling in observational studies, by defining the two forms of causality, which are causal discovery and causal inference; and this is presented in section 2 under the bold heading captured as “Basic concept of causal models. Section 2 also elucidates the two frameworks for executing causality in dataset, i.e., the structural causal model (SCM) framework and the potential outcome or Rubin causal model (RCM) framework. A comparison of both frameworks is advanced. Section 2 ends with how interventions are done with SCM in observational datasets. Since our emphasis is on SCM, a major component of this model is the graph (specially the directed acyclic graph – DAG). Thus, section 3 is dedicated to the concept of graph and its concomitant features that makes causal identifiability in observational settings possible and plausible. A composition of the causal graph is explicated in this section, followed by the cardinal connections that exists in them, which are the collider, the chain/mediator and the fork. A succinct background into the popular backdoor adjustment criteria and how causal connections can be ascertained in graph is also presented. The section ends with the Bayesian network factorization (BNF), the mathematical parts that connects the probabilities of variables in the graph with their parent probabilities and how the causal interventions are done with BNF is also explicated. Section 4 presents the major assumptions that drives these DAGs and their accompanying datasets. The Markov condition that connects variables probabilities with their parents (conditional) probability is demystified. The causal sufficiency assumption that ensues there are no latent or confounding variables is presented; followed by the acyclicity assumption that precludes variables in a graphs from forming cycles is also elucidated. Finally, the causal faithfulness assumption that ensues the variables relations in the graph is symmetric with the distribution concludes the section. Section 5 seeks a categorization of the causal algorithms into constraint-based, score-based and functional causal model (FCM), in order to identify where each algorithm alongside its dataset setting belongs. Section 6 which is the crux of the matter seeks to identify each data setting and the type of algorithm and assumptions advanced to solve causality in them. Ten data settings type are identified with each algorithm solution(s) and assumptions accompanying it. Section 7 concludes the study with a recapitulation.

International Journal of Current Science Research and Review

ISSN: 2581-8341

Volume 06 Issue 11 November 2023

DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789

IJCSRR @ 2023

2 BASIC CONCEPT OF CAUSAL MODELS

In this section, the various forms of causality are defined, followed by the two major framework used for causality, which are the structural causal model (SCM) framework and the potential outcome or Rubin causal model (RCM) framework; with a juxtaposition of both frameworks. The section concludes with how causal interventions are executed in dataset with the SCM framework.

2.1 Causal: Discovery & Inference definitions: Causality can be defined as the process by which one or more independent variables (an event, process, object or state) can produce or influence the outcome of one or more variables (usually called a dependent variable); that is to say the c*ause*' wholly or partly is responsible for the effect and the '*effect*' is partly or wholly dependent on the cause [4-6]. It is imperative to note that the concept of causality is indeed broad and diverse, and encompasses different fields and disciplines, such as statistics, machine learning, data mining, epidemiology, economics etc. [5, 7]. Causality is divided into two main branches. Viz. (i) causal discovery and (ii) causal inference. *Causal discovery* is defined as the process of extracting or inferring causal knowledge, relationship or causal structure from datasets, by analyzing observational data, based on some graphs (Direct Acyclic Graph [DAG]) as in the case of SCM and by the use of some statistical tools like the probability distribution and the use of structural equations [5, 8]. While *causal inference* is defined as the process of inferring causality based on some assumptions that a particular treatment or intervention was actually the cause of the observed outcome [4, 9] from the causal discovery process in a dataset. For example ascertaining from the observation data that the intake of aspirin (intervention or treatment) was the reason for stopping the headache (the outcome) on the test subjects from which data was collected [10]. In general, causal models in observational studies are designed to mimic the RCT experiment which is the standard framework for carryout causality [11]. In spite of the distinction made between causal discovery and causal inference, as explicated and delineated above; in this study, words or phrases such as: causality, causation, causal identifiability, casual discovery, shall be interchangeably used to mean both.

2.2 Causal Model: It is an abstraction of mathematics that describes quantitatively the relations of causality that exist among variables in an observable dataset [7]. These mathematical models are derived from the domain and background knowledge embodied in the DAG, and they evince the causal relations within the observable dataset [11-13].

2.3 Types of causal models: Two types of casual models exists for causality, which are (i) Structural causal model (SCM) proposed by Pearl [12] and (ii) Potential outcome framework also called Rubin causal model (RCM) [14, 15].

2.3.1 An SCM: The framework for causality based on SCM gives a holistic understanding of the theory of cause and effect. It is composed of two parts: the causal diagram (or graph) that encodes background domain knowledge and assumptions of the distribution (the dataset), and the Bayesian network factorization (BNF) or structural equations part, which models or algorithmised (mathematically) the relations among the study variables based on the causal assumptions from the graph [7, 16-18]. This works focuses more on the SCM with a more detail explication of the connections of the graphs and the dataset in subsequent sections.

2.3.2 The potential outcome framework: Causality employ by this framework composed of a pair of variable set (t, y), where t stands for the treatment and y stands for the potential outcome. Formally, the potential outcome is defined as: given a pair of variables (t, y), the potential outcome $t_i(y)$, shows what the outcome would be, if treatment t were to be applied in that individual variable i instance [7, 19]. Hence, this scenario differentiates potential outcome from observed outcome, because not all potential outcomes can be observed, but rather all potential outcome have the potentials to be observed [19]. The observed outcome is dependent on the value assigned by the treatment. The RCM framework is employed to help articulate and solve the fundamental problem of causality, which is missing data [20]. Thus with this framework, only the potential outcome of one individual instance can be observed at a time. Hence, we can define the Individual treatment effects (ITE) which translates to the causal impact on the individual mathematically as: $\tau_i = y_i(1) - y_i(0)$; which means the potential outcome of an individual instance τ_i under two varying conditions, i.e., treatment ($y_i(1)$) and control ($y_i(0)$). This can also be extrapolated to Average Treatment Effect (ATE) on a certain arbitrary population as the expectation of the ITE of the entire population $(i = 1, 2 \ldots, n)$ as:

$$\tau = E_i[y_i(1) - y_i(0)] = \frac{1}{n}\sum_{i=1}^{n}(y_i(1) - y_i(0)) \quad (1)$$

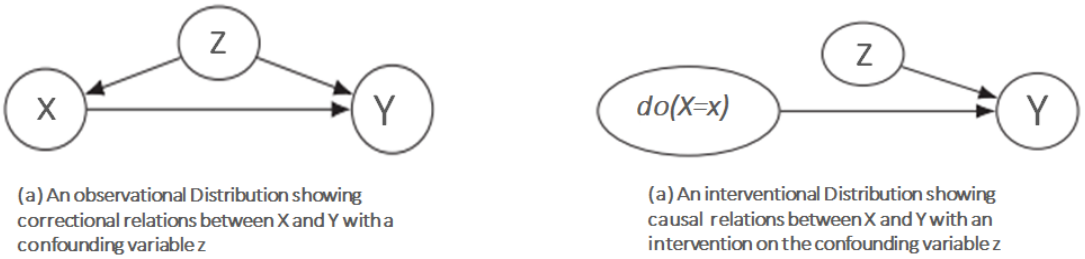
Also, ATE can be taken for a subpopulation of the data that is conditioned on, like in the case of an intervention and it is known as the *conditional average treatment effects* (CATE) [19].The RCM does not necessarily require a graph or DAG. The RCM framework is cardinal focus of the work but the rather the SCM with its avalanche exposition of its DAG.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 06 Issue 11 November 2023
DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789
IJCSRR @ 2023

In comparison, the SCM and the RCM frameworks are both alike when it comes to the logic and the assumptions that drives them [4]. They however differ slightly, as the potential outcome or RCM framework does not explicitly identify or define the casual effects of some instrumental variables in the distribution, aside the special variable of treatment; and the knowledge about the complete graph structure is not always a desideratum; which is its downside. This downside of the RCM framework is also its advantage, as researcher can elect to develop estimators of certain variables and model them in the distribution without necessarily considering variables of non-interest. While in the SCM framework, all variables (both observed and unobserved) can be identified, studied and modelled by the use of the complete graph (DAG) or others as its relates their connections in the distribution. The pro of the SCM is in its detailed approach to causal modeling. This can also be a setback when it comes to the quick and timely identification of causality in variables of interests in the distribution, as it would be preferable to use the RCM framework in such an instance, instead of a model like the SCM that would necessitate a complete graph structure of the entire variables in the distribution.

2.4 Causal Relations with SCM: Determining the causal relations that exists among variables in an observational study in a purely probabilistic distribution is an ambiguous and daunting task. If a conditional probability distribution such as $P(Y|X)$ for instance, represent the conditional probability distribution of obesity (Y) given a particular level of sugar intake (X). This distribution relation is ambiguous in terms of an experimental setting (RCT) where sugar intake was ascertained by randomization or by merely through an observational process. In his book on causality, Pearl [12] in order to differentiate the mere conditional observational probability distribution (I,e., statistical association/correlation) and interventional conditional probability distribution (which is a causal association), introduced the do-operator of the do-calculus to differentiate interventional distribution from observational'. Hence, the expression $P(Y|X)$ can now be regarded as mere conditional observational association which depict how the probability of Y (obesity) will change, if someone were to observe the sugar intake (X). While $P(Y|do(X = x))$ is regarded as the interventional conditional probability distribution (which is a causal association), depicting the probability of obesity (Y) given that a measure unit of sugar (x) were taken (purposefully and not observed). Hence, making the observation and intervention distinct: $P(Y|X = x) \neq P\big(Y\big|do(X = x)\big)$. The practical difference between the two may be the existence of a variable(s) Z (individual gene tar for instance) that may be confounding the relations, which exists in some back-door path: See figure 1 DAG for confounding relations. In the intervention distribution, the causal effects is determined given difference values of the treatment/control X (i.e., when sugar is taken and when sugar is not taken) and this can be measured and compared in the interventional distribution, written as: $P\big(Y\big|do(x = 1)\big)$, and $P\big(Y\big|do(x = 0)\big)$ where 1 and 0 stands for treatment and no treatment (control) respectively for an individual instance, which is called the *individual treatment effect* (ITE). Thus, when this process involves all sampled or all instances of the population, the causal intervention is defined in terms of the average treatment effects (ATE) for the instances of the population. Written in terms of the expectation as: $\tau(1,0) = E[Y|do(x = 1)] - E[Y|do(x = 0)]$. Also, conditional average treatment effects (CATE) can be taken for subpopulation group in a similar manner as well. Thus, it can be seen that this kinds of intervention model's the RCT experiment that determines causality in observational dataset [21, 22]. In spite of the clear distinction describing and differentiating these two processes by Pearl [12], not every dataset can be neatly categorized into this distinction of observational and interventional dataset, as some experiments may not clearly or wholly show the value of the variable that is intervened on in the dataset. Thus, due to this two distinctions, which are obfuscated in the distributions, it has become imperative to represent causal models explicitly in terms of directed acyclic graph (DAG) or simply causal graph as proposed by Pearl [21]. Causal graph in SCM are very essential component which make it easier to identify the causality from dataset; hence, we discuss them in the next section.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 06 Issue 11 November 2023
DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789
IJCSRR @ 2023



(a) An observational Distribution showing correctional relations between X and Y with a confounding variable z

(a) An interventional Distribution showing causal relations between X and Y with an intervention on the confounding variable z

Figure 1: Depicts the observational statistical correlational relations distribution (a) and the causal interventional relations (b)

3. CAUSAL GRAPH

This section presents causal graphs as is applicable in SCM. Fundamental concepts in graph such as the popular backdoor adjustment criteria and the Bayesian network factorization (BNF) are elicited and explicated.

3.1 Causal graph Composition: A causal graph (denoted as $G = (V, E)$, consists of two or more nodes (also called vertices) representing a random variable sets $(V), where\ V = X_1, X_2, X_3, \ldots X_n$ and a number of connecting lines among the nodes called edges (E). These random variables may include the observed and unobserved (if the exists) variable alongside the treatment and outcome variables. In figure 2: 1A is an undirected graph due to the lack of directional arrows on them. While 1B the graph is directed because of the arrow direction. And 1C shows a directed graph with a cycle [19] and finally 1D shows and intervention graph on variable C. A directed edge from A to B (written as: A $\rightarrow B$) is interpreted as, B is caused by A or (A is the potential cause of B) [7]. Hence, with a causal graph an hypothesized causal query can clearly be modelled through the causal pathways in the graph, and all dependent/independent relations as it relates all variables associated with the query are known. And this graph model can be factorized using the Bayesian network factorization or the structural equations; based on some assumptions to obtain a causal estimand of the conditional probability distribution from which it can be used with the observed dataset to ascertain the causal estimate of the hypothesized query [21, 23].

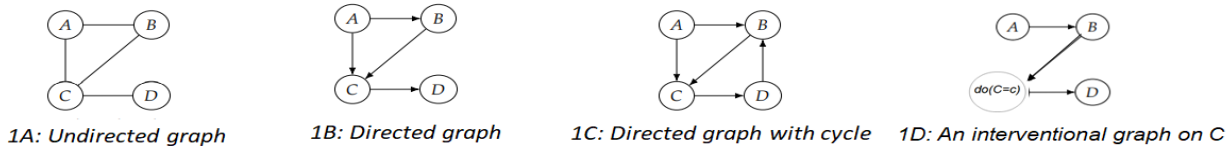


Figure 2: Shows an undirected, directed, directed with cycle, and intervention graph

A *path* in the graph is an oriented order of adjacent edges irrespective of the direction of the adjoining nodes. For instance, $A - C - B$ is considered as a path in figure 2 1A and $A \longrightarrow C \longleftarrow B$ is also a path in figure 2 1B. A directed path is that in which all edges are directed or pointing in the same direction. E.g., the path, $A \longrightarrow C \rightarrow B$ in figure 2 1B is regarded as directed. Most causal algorithms work best with the directed acyclic graphs (DAGs) condition as shown in figure 2 1B and a few causal algorithms work with the cyclic graph condition as shown in figure 2 1C [7, 16-18].

3.2 Three Cardinal Relations in Graphs: A descendant of a node A is a node $C \in V$, such that there is direct edge from A to C (written as: $A \longrightarrow C$) in the DAG G. This corresponds to A being an ancestor (parent of) C. The progenies (A and B) of a node C, are the nodes in V with a directed edges connecting C, (designated as: $A \longrightarrow C \longleftarrow B$). This child and two parents relationship designated as $A \longrightarrow C \longleftarrow B$, is also called a *collider* [5, 9] or *immorality* [13, 19] is the first basic relation that can exists among variables represented in DAG. A second relation exists called a *mediator or chain*, where a parent node A (usually exogenous) that produces a child node C, where C in turn produces another child B (which is a grand descendant of A) [13, 21, 23]. finally, a third relationship exists where a node C, which is a single parent having two descendants A and B (written as: $A \longleftarrow C \longrightarrow B$) is called a *fork* or common cause confounder. Thus, these three relations (collider, chain/mediator and fork) are the three common relations that exist in an observational dataset and can be mirrored or expressed in a DAG, forming the building block or structure in causal graph for determining relationship (causal or associational) in an observational settings [8, 13, 16, 21, 23, 24].

International Journal of Current Science Research and Review

ISSN: 2581-8341

Volume 06 Issue 11 November 2023

DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789

IJCSRR @ 2023

3.3 Causal Connection & the Backdoor Adjustment Criteria in a Graph*: D-separation* and *d-connection* are the processes that define a sets of variable V's connectivity in a causal graph G [24]. The D in the d-separation and d-connection stands for *dependency* and it is a process of establishing independency or dependency from two or more variables that are independent or otherwise on a third variable C in in a DAG which is a reflection in the dataset. For instance, in the case of a fork ($A \longleftarrow C \longrightarrow B$), or a chain/mediator ($A \longrightarrow C \rightarrow B$), the variable C is a link between both A and B. Hence, once you condition on the linking variable C, you will block or close the dependency relationship that exist between paths A and B. That is to say, paths A and B will become independent conditioned on C, written as: $A \perp\!\!\!\perp B|C$. Albeit the reverse is the case, when it comes to the collider or immorality structure ($A \longrightarrow C \longleftarrow B$), as the paths A and B are already independent or blocked in their current state (i.e., $A \perp\!\!\!\perp B \nmid C$: A is independent of B not conditioned on C), without the need for conditioning on any variable including C. Hence, once you condition on C, a relationship between A and B is induced (i.e., A and B becomes dependent conditioned on C. written as: $A \perp\!\!\!\not\perp B|C$). This process of blocking the flow of unwanted association on non-causal pathways in order to determine causality only through a causal pathway is called the *backdoor adjustment criteria* [25, 26]. Pearl [24], defined the process of d-separation and d-connection for backdoor adjustment criteria in a DAG G formally as follows: A path connecting two variables A and B is said to be d-separated or blocked if and only if: (i) the path contains a fork such as : ($A \longleftarrow C \longrightarrow B$) or chain/mediator such as: ($A \longrightarrow C \rightarrow B$) that has been conditioned on C. Written as: (A$\perp\!\!\!\perp_G$ B|C), and (ii) the path between A and B contain a collider on C, such as ($A \longrightarrow C \longleftarrow B$) that has not been conditioned on, alongside any descendant of collider C, that is not conditioned on as well. Written as: ($A \perp\!\!\!\perp_G B \nmid C$) or just $\perp\!\!\!\perp_G B$. This same process of d-separation and the backdoor adjustment criteria from the graph G can be utilized to determine dependencies/independencies of variables in the distribution (or dataset), which is a factorization of the d-separation in the graph using the Bayesian Network Factorization (BNF). The d-separation in the distribution is written as: A$\perp\!\!\!\perp_p$ B|C, or A$\perp\!\!\!\not\perp_p$ B|C for independency and dependency conditions respectively, similar to the d-separation in the graph with the subscript $_P$ to distinguish it from the graph's d-separation criteria, which is represented by the subscript $_G$. This can further be used to determine causal relations in the distribution as whole.

On the other hand, a path from A and B through C, is said to be ***d-connected***, unblocked or open when it is not d-separated [23, 24].

3.4 The Bayesian Network Factorization (BNF) in Graphs*:* The DAGs are interpreted in two part. i.e., the probabilistic and the causal interpretations. The probabilistic inference sees the directional arrows on the DAG G as showing a probabilistic dependences or associations among the variables of study, while the lack of arrows corresponds to the conditional independence asserted by the study variables [27]. Based on some assumptions, the simplest being the *Markovian condition*, which states that each study variable is considered independent of all its non-descendants in the graph with the exception of its direct parent. Usually written as $A \perp\!\!\!\perp B|C$. Hence, based on the assumption, the joint probability distribution function $P(v) = P(v_i, \dots, v_n)$ factorizes based on the BNF as:

$P(v) = \prod_i^n P(v_i|pa_i)$ (1b)

Where $v_i = 1, \dots, n, and\ pa_i$ denotes the parent of the variable $v_i$in the graph [7, 24, 27].

Thus, based on the BNF of equation (1), the graph in figure 2:1B for instance, the probability distribution of it (i.e.,1B) can be factorized and summarized, based on the Markov assumption as follows:

$P(A, B, C) = P(A)P(B|A)P(C|B, A)P(D|C)$ (2)

This contrasts the normal Bayesian probability distribution network which uses the chain rule without the graph and the Markov assumption, written as:

$P(A, B, C) = P(A)P(B|A)P(C|B, A)P(D|C, B, A)$ (3)

The difference in equation (2) and (3) is in the last product conditional probability of D, where equation (2) reduces the conditioning probability to only its immediate parent node C, based on the position of equation (1) and as captured in the graph of figure 2:1B. While equation (3) assumes no graph and factorizes the distribution using the chain rule. Hence, the probability of D, given (or conditioned on:) C, B and A are used as elicited in equation (3).

3.5 Causal Identifiability with BNF Intervention Graphs: The second interpretation of the graph is called a causal interpretation. In this scenario, the arrows direction in the DAG G represents the influence of causality among the variables. Here the BNF of equation (1) above is still essential, but the arrows are assumed to evince a separate process in the data generated. Hence, after eliciting causal path from the DAG G, the conditional probability of the distribution $P(v_i|pa_i)$ which is generated based on the

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 06 Issue 11 November 2023
DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789
IJCSRR @ 2023

graph G, and which is a statistical estimand, can be estimated from the data. The relations of conditional dependency expressed by the BNF formula of equation (1) does not necessarily leads to causal inference (due to the mixtures of confounding variables sometimes). However equation (1) can be extended to cater for interventions (which are causal in their implementation) as presented by Pearl in [12]. Using the do-operator of the do-calculus as an intervention on the desired variable (or node) the difference between mere conditional distribution (correction), written as: $P(Y|X = x)$, and the causal intervention of the conditional distribution, written as: $P(Y|do(X = x))$, in the graph and subsequently the data can be clearly distinguished. For instance, if the graph in figure 2, were derived from the query hypothesis of determining the effects of shoe size X on the reading ability Y of children. The age variable Z, confounds the relationship between reading ability Y and shoe size X, making them to have statistical correlation as shown in figure 1(a). But when you carry out an intervention on the shoe size X such as $P(Y|do(X = x))$, the age variable Z that confounds the relations is severed, and the conditional probability of the BNF produces an estimand which is given as $P\big(Y\big|do(X = x)\big) = P(Z)P(X|Z)P(Y|Z,X)$. Which is summarized by getting rid of the factor for probability of X in the BNF to get: $P\big(Y\big|do(X = x)\big) = \sum_z P(Y|Z,X)\,P(Z)$. With this causal intervention estimand, using the d-separation and the backdoor criteria, the shoe size X will be set to a treatment unit of 1 and no treatment (control) unit of 0, while conditioning on a certain age Z say 8years. Thus, the difference between the treatment and no treatment of shoe size $(X: 0,1)$ generated from conditioning on a certain age $(Z = 8)$ for the set of $\boldsymbol{Z}$ variable in the dataset can be calculated as the ATE, given mathematically in terms of their expectation as: $\tau(1,0) = E[Y|do(x = 1)] - E[Y|do(x = 0)]$, which translate to the causal estimate or causal inference estimation on the effect of shoe size X on reading ability Y in children. This estimate would likely be zero (no effect), thus killing the lurking variable (age) and exposing the spurious association (correlation) that exists between shoe size X and reading ability Y. Note however that if the confounding variable Z is unobserved or not part of the distribution (the dataset), the causal identification of X on Y cannot be feasible to obtain in the data, even though it is revealed in the graph. This do-operator which translate to intervention and causality in data differentiates mere association (correlation) that is used in machine learning algorithms.

With SCM, counterfactual hypothesized queries which are carried out on an individual level of the sampled dataset can also be estimated, using some techniques proposed by Pearl [10, 28] which transcend the do-operator of the do-calculus, which only work with i.i.d condition [29]. Although counterfactual causal effects would not be covered in this work.

4. MAJOR ASSUMPTIONS IN SCM

This section covers the four major assumptions often used for causality, especially with i.i.d datasets, thus driving the process of causality in observational data setting with the SCM framework. These assumptions are: (i) The Markov assumption, (ii) The Acyclicity assumption (iii) The Faithfulness assumption, (iv) The causal sufficiency assumption. These assumptions are summarized as follows:

4.1 The Markov assumption: This assumption states that, a parent node in a DAG G representing a variable is considered independent of all its non-descendant in the graph with the exception of its direct parent. This assumption ensures that causal estimand for the identification of the causal relations is generated from the graph to the data, using the BNF or the structural equation of functional causal model (FCM). This estimand which is modeled using the Markov condition when it is sufficient (i.e., all confounding variable identified), becomes the basis for which the probability distribution, which is a statistical estimand can be estimated from the dataset. Equation (1) is a representation of the Markov condition. The Markov assumption when combined with the causal edge assumption that states that: in a DAG G, all adjacent nodes are dependent; can generally be referred to as the *minimality assumption* [15, 19, 30].

4.2 The acyclicity assumption: It is the phenomenon that ensures that the set of adjoining variables nodes $\boldsymbol{V}$ in the causal graph does not form a cycle, a feedback loop or go back in time as shown in figure 2:1C, but are rather directed and acyclic as shown in figure 2:1B [31, 32].

4.3 The faithfulness assumption: This assumption is the opposite or converse of the Markov assumption. While the Markov assumption works from a given causal graph to the dataset by modelling the causal estimand from the graph and implementing same in the probability distribution $P(V)$; the faithfulness assumption utilizes a principle that seek to identify a causal graph and its

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 06 Issue 11 November 2023
DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789
IJCSRR @ 2023

associate casual estimands from a given dataset [19, 33]. That is, it seeks to move from condition probability distribution (in the dataset) such as: A$\coprod_P$ B|C to a DAG conditional probability structure such as: A$\coprod_G$B|C. The faithfulness assumption states that, if a variable A is independent of B, conditioned on C in a probability distribution in the dataset $P(A, B, C)$, written as A$\coprod_P$ B|C, then the variable A would be d-separated from B conditioned on C in the causal graph G as well, written as: A$\coprod_G$B|C. The assumption of faithfulness as regards the identification of causality in observational data is considered one of the simplest assumption and the violation of this assumption is pretty much a common phenomenon; as conditions and connections that doesn't show dependencies are enough to violate this assumption. For instance, as shown in figure 3, if the effects of two causals paths in a graph G cancels out themselves completely. I.e., $A \longrightarrow C \rightarrow D$ *and* $A \longrightarrow B \rightarrow D$; then the independent of the causal variables in the graph will remain stable. i.e., A and D becomes independent or d-separated from each other, written as: $A \coprod D$ [34].

4.4 the causal sufficient assumption: This condition states that in a given causal graph G, there are no variables confounding relationships that is unobserved among the study variables. That is to say, the causal sufficiency assumption ensures that all variables that may be confounding or having a hidden effect on the hypothesized query variable of treatment and outcome (t, y) are identified and explicitly shown on the graph, whether or not they are observed in the distribution of the dataset [35-37]. Hence, these four assumptions are the building blocks for causal discoveries in observational studies in constrain-based (CB) and the score-based causal identifiability techniques, with an i.i.d. distribution. However, other assumptions exist aside these ones, that are used in the FCM model settings and some datasets which are not i.i.d structured, and we will identify and explicate them in the relevant headings that they appear.

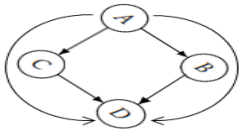


Figure 3: Showing an instance where the faithfulness assumption may be violated

5. CATEGORIES OF CAUSAL ALGORITHMS

In this section a clear exposition into the classes or categories of the standard algorithms developed for learning causal relations in dataset is presented and elucidated based on these three classes. Viz. (i) Constraint-based (CB), (ii) Score-based (SB) and (iii) Functional-based models.

5.1 Constraint-based (CB) Algorithm: It employs the statistical BNF approach, by using the Markov equivalent condition (MEC), the faithfulness assumption, the causal sufficiency assumption and the acyclicity (a DAG), which provides the causal graphs based on the conditional independencies of the set of variables found in the data distribution; by carrying out some hypothesis tests using statistical methods in order to identify the causality. Some techniques identified for the implementation of the conditional independent test are the G2 test [38, 39], and the Z-test approach, proposed by Fisher [40]. A famous example of a CB algorithm that operates on the four assumptions that constitute the algorithm is the Peter-Clark Algorithm (PC-algorithm for short). This algorithm is effective and consistent with datasets that are i.i.d, with a Gaussian or multinomial distribution; which are employed for generating the causal graph [24, 39, 41, 42]. Other forms of the constraint-based algorithms that exist are the Inductive Causation algorithm (IC-algorithm) [43] and their variations [36, 44]. Other families of algorithms that seek to overcome the restriction of the normal and multinomial distribution are covered in these references [7, 45-47]. Other CB algorithm such as the fast causal inference (FCI) and its variants [48, 49], are developed to cater for the issue of unobserved confounding in the causal graph, by dropping the causal sufficiency assumption. Also, other CB causal discovery algorithms such as the Cyclic Causal Discovery (CCD) algorithm are developed to cater for the feedback loop for variables that violates the acyclicity assumption [50]. Furthermore, another CB-algorithm called the SAT-based algorithm, was developed and it drops the faithfulness and the acyclicity assumptions [51].

5.2 Score-based (SB) Algorithms: It is similar to the CB-algorithm in terms of graph formation through the BNF, albeit it drops the faithfulness assumption and uses the goodness-of-fit method as a test instead of the conditional independence test method. The set of candidate graphs produced from the variable set V in the distribution, are each represented with a structural equation and are also allocated an important score by the algorithm via some measure of adjustment scores. The Bayesian information criteria (BIC)

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 06 Issue 11 November 2023
DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789
IJCSRR @ 2023

scores method [52] is mostly used and widely adopted. The BIC scores is employed to determine how well the dataset fits the DAG structure, while at the same time attempting to correct the complexity of the DAG that does not fit the data [53]. Allocating scores to all possible graphs in the distribution is infeasible and computational expensive. Hence, heuristic approach such as the greedy equivalence search (GES) [54] algorithm and its variant, the Fast greedy equivalence search (F-GES) [55] are employed to attain the desired graph optimization within the local context.

Some *hybrid algorithm* that seeks to combine the techniques of the CB and the SB algorithms exists and are gaining traction [56, 57]. Algorithm such as the Max-Min Hill-Climbing (MMHC) algorithm [57] is employed to scale-up variables set in their thousands, which overcome the computational constraint of SB approach, which cannot scale-up such magnitude of variables set. This algorithm is configured to first learn MEC skeleton of the DAG by employing an algorithm called the MMPC (Max-Min Parent and Child) [56], which is a technique used in CB algorithm. The edges are afterwards oriented using a SB technique called the Bayesian scoring hill climbing search (BS-HCS). A hybrid modification of the GES algorithm is developed [58] and is called AR-GES and its search space is dependent on the conditional independent graph estimation. Also, adaptive changes in the AR-GES search space is dependent on the present state the algorithm is at, and this modification is imperative for general consistency of the algorithm.

5.3 Functional Causal Model (FCM) Algorithms*:* Causal models are sometime built on the assumption of a set of semi-parametric ***structural equations***, called functional causal models (FCM), which can also be generated from the graph and the causal effects which are indicated by the arrow head edges can be computed and quantified using structural equations. Similar to the SB, a variable in FCM can be depicted mathematically as a function of its direct cause or parent Pa_i and some disturbance or noise term ϵ_Y. Written as: $Y = f(X, \epsilon_y)$. For instance, in the graph of figure 2 :1B, the structural equations that expresses the causal relations can be given as $A = f(\epsilon_A),\ B = f(A, \epsilon_B),\ C = f(A, B, \epsilon_C),\ D = f(C, \epsilon_D)$. where $\epsilon_A, \epsilon_B, \epsilon_{C,} \epsilon_D$ depict the noise or disturbance terms for each function respectively. Although all the noise terms are usually implicit and are not shown in the graph. Thus, with the FCM, the assumption of causal faithfulness is not required, albeit the Markov, sufficiency and acyclicity remained valid and essential assumptions. FCM are able to differentiate DAGs that share the same MEC. With the assumptions employed in FCM, the DAG configuration from data that is produced from a mechanism that is linear but with non-Gaussian noise can be recovered with a causal model called the Linear non-Gaussian Model (LinGaM) [59]. Also, the ICA-LinGaM model exist which estimates the model coefficient by the use of the independent component analysis (ICA) technique for signal analysis [60]. However, the issue with ICA-LinGaM is that ICA algorithms converges easily to a local prime; hence, a DirectLinGaM algorithm [61] is proposed to overcome the challenge and guarantee a global optimum. Similarly, with a nonlinear data generated structure, the additive noise model (ANM) assumption [62, 63] is proposed for the variables and the noise distribution. With this ANM condition, the identification of the true causal graph is guaranteed, as there is no backward causal model with ANM for a nonlinear structure in the non-causal direction [13]. Also, a post-nonlinear (PNL) model for non-LinGaM is proposed to handle the functional space increase amongst the variables and the noise term [64]. FCM such as ICA-LinGaM auto-regression are also adjusted and adapted for causal identification in time-series dataset [65]. Other FCM are adapted to handle cycles or feedback loop by dropping the acyclicity assumption [66], while other are adapted to cater for hidden or confounding variables [67] by dropping the causal sufficiency assumption.

6. CAUSAL IDENTIFIABILITY TECHNIQUES IN DIFFERENT DATA SETTINGS

In this section, we elicit and discuss the different models employed for causal discovery in different data settings, cutting across the three classes of causal models discussed in section 5, which are the CB, SB and the FCM algorithms. The section identifies and elucidates 10 data types settings, alongside the assumptions that drives the process; which most cases are a deviation from the i.i.d, the Gaussian or multinomial structures which works best on the four major assumptions enumerated in section 4. These includes:

6.1 Linear Non-Gaussian Settings: Causal discovery and identifiability in a dataset of bivariate distribution of random variables such as (x, y) in a linear and Gaussian setting, is difficult to determine. It does not suffice to tell whether x is the cause of y or vice-visa (i.e., $x \longrightarrow y$: called the forward model or y is the cause of x, $x \longleftarrow y$: called the backward model) from the SCM using the MEC assumption and the concomitant assumptions of acyclicy and causal sufficiency; as the symmetric relations and Markov conditions holds in both directions. The Non-parametric structural equations models (NPSEM) generated from both the forward and backward SCMs are given as: $y = f_y(x, \varepsilon_y)$, where ε_y denote the noise term in the f_y function and the variable x is independent of

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 06 Issue 11 November 2023
DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789
IJCSRR @ 2023

the noise term. Written as: $x \coprod \varepsilon_y$ in the forward model: $x \longrightarrow y$ and the backward model is given as $x = f_x(y, \varepsilon_x)$, where ε_x also denote the noise term in the function f_x and the variable y is independent of the noise term. Written as: $y \coprod \varepsilon_x$ in the backward model: $x \longleftarrow y$. If you carry out a parametric linear regression for the two random variables, using the noise as an additive term for the forward and backward models, and using the regression equations: $y = ax + \varepsilon_y$ and $x = ay + \varepsilon_x$ respectively; and by plotting the predictor (effects) and its cause alongside the cause and each noise (the residuals) term, the relationships that would ensue, would make it impossible for causal identifiability in both cases for a bivariate linear Gaussian distribution [13, 19, 21].

Albeit to overcome this issue, some assumptions about the parametric form will suffice. Hence, it would suffice to assume that the structural equations of the bivariate distribution are linear, non-Gaussian and acyclicity assumptions [59, 62, 68] aside the MEC condition (not requiring the faithfulness and sufficiency assumptions). Thus, with these assumptions, the causal identifiability of the bivariate distribution' real-valued would become feasible. The structural equation that generate the data is assumed to take the form: $y = f(x) + \varepsilon$ where the function f is said to be linear and $x \coprod \varepsilon$ (x is independent of ε) while ε is said to be a distribution of non-Gaussian. And all these are based on the Theorem of Shimizu et al.,[59]. Accordingly, there does not exist an SCM in the backward model or direction such that: $x = f(y) + \xi$ and where $y \coprod \xi$ that can generate data consistent with the same bivariate random distribution $p(x, y)$. As a matter of fact, when you fit a regression model in the opposite direction using equation $= f(y) + \xi$, you will realized that the variable y is dependent on the noise (residuals) ξ : ($y \perp\!\!\!\perp \xi$: y is dependent on ξ). Thus, the production of a causal asymmetry between the two variables x and y, based on the theorem of [59], which is non-Gaussian assumption is only in the forward model and not the converse or backward direction, and the general proof of this concept is elicited and explicated in this reference [59]. In fact, in any such distribution when any one of the causal variable x and the noise term ε are Gaussian, it is possible to identify the causal direction, due to a theory by Hyvärinen et al.,[60, 69] called Independent component analysis (ICA) or more directly as a result of the Darmois-Skitovich theorem [70]. This method is broadly known as Linear non-Gaussian Model (LinGaM) [59]. Several extension of the LinGaM model exists. For instance, Shimizu et al.[59] employed the multivariate distribution apart from the bivariate form in their work. Hoyer et al.,[62] included the causal sufficiency assumption in their work to help handle confounders and latent variables; while Lacerda et al.,[71] and S-Romero et al., [72] works permitted graphs with cycle or feedbacks under other assumptions, thus, violating or dropping the assumption of acyclicity in order to uncover causal knowledge in the dataset model of LinGaM. Also, studies that demonstrates the process of generating data, which satisfies the LinGaM proposition but the actual models are termed Post-nonlinear (PNL) due to the fact that the real data were inverted through a nonlinear concept and process exists in this references [64, 73].

6.2 Nonlinear Additive Noise Dataset Setting: During the process of data generation, a nonlinear process is something involved at the end transformation phase. Thus, causal identifiability of such a data setting, should be able to consider the functional class of such nonlinear process in order to bring about causal identifiability in the random variable distribution of such nature. Hoyer et al.[62] and Mooij et al.,[74] have suggested a direct extrapolation of the LinGaM model for the nonlinear setting by expressing the effect Y as a nonlinear function of the cause (or parent variable) X and an additive noise or an independent error term ε. Written as: $Y = f_y(X) + \varepsilon_y$, where the causal variable $X \coprod \varepsilon_y$ (X is independent of the noise term ε_y), while f_y is a non-linear function. Thus, in this case, the causal identifiability of the dataset become easy to determine as there is no backward model in the nonlinear case where a similar function such as: $X = f_x(Y) + \varepsilon_x$ that exist in the backward direction ($X \longleftarrow Y$), but only in the forward ($X \longrightarrow Y$). Albeit in the nonlinear case, the assumptions of acyclicity and the Markov conditions, holds true alongside the additive noise assumption [21]. The PNL functions also exists in the additive noise setting, where the function f_y is superimposed with another function g_y to get ride of the noise term. Given as: $Y = g_y(f_y(X) + \varepsilon_y)$ in order to bring about causal identifiability as explicated and proofed in works of [64, 73].

6.3 Time-series Causality Issue: The collection of data on a single event by the observance of a sequence of changes multiple times on an item of data indexed with time order is called a time-series data [5]. Time-series observational dataset can be univariate (involving one variable) or multivariate (involving two or more variables), which may take different forms such as discrete, binary, continuous, text form etc. Causal search in time-series observation data is no doubt a daunting task, albeit there is an exponential escalation in the causal discovery and identifiability in time-series observational data recently by researchers. Many techniques are designed specifically for solving tasks involving causal discovery in sequential or time-series observational data. A popular

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 06 Issue 11 November 2023
DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789
IJCSRR @ 2023

framework known as Ganger causality, that was proposed by Ganger [75] exist for this purpose. Mathematically, this framework model can be expressed as an auto-regression task such as: $Y = \sum_{i=1}^{n} a_i Y_{t-i} + \sum_{i=1}^{n} b_i X_{t-i} + \epsilon_t$ where n is the order of the model or the maximum number of lags to be employed and a_i and b_i are the coefficients of Y and X respectively, and they contribute to the delay in observations of the variables in the framework. The coefficients b_i of the Ganger causal variable in particular are considered to be statistically significant. Intuitively, the Ganger Variable X Causes Y if the prediction of Y is based on the previous observations, and the previous observation of X does better in performance than the prediction of Y occasioned by its previous performance only [5]. Other less similar methods of causal identifiability techniques in time-series data includes: The Windows approach: which is an adaption of the FCI algorithm for time-series data, and it is known formally as Time-series-FCI (Ts-FCI), proposed by Entner & Hoyer [76]. It involves partitioning the data into disjoint windows, and measuring each unit as a distinction analytic unit. Also another strategy called "*Timestamp*" involves the handling of measurements at different times setting as independent and separate timestamp from each other. The algorithm used is called the PC-MCI (Peter Clark, Momentary Conditional Independence) algorithm and it was proposed by Runge et al.,[77]. Furthermore, this algorithm is a graphical method designed to handle linear and nonlinear time-series observational datasets. The PC-MCI algorithm is a two-phased algorithm (i.e., PC1 and MCI) with each phase representing conditional independence of different timestamp. In the PC1 phase, it employs the conditional independence technique of the PC skeleton structure to discovery possible variables that have dependent relations in each and every timestamp, both current and previous. In the next phase which is the MCI, the algorithm employs the momentary conditional independence as suggestive of the abbreviation MCI as a test to ascertain the causal relations amongst the variables in each timestamp. Extensions of the PCMCI algorithm exists, for instance the PCMCI+ proposed by Runge [78] for handling of concurrent links within timestamp. I.e., a causal relations that exist amongst variables within the same timestamp [5]. Also another extension called the LPCMCI which was proposed by Gerhardus & Runge [79], is designed for handling of latent variables in time-series observational data. Albeit all these methods of causal discovery in time-series data as discussed above are not without their downsides. The Granger and the Timestamp methods have the issue of not having all units of measurements independent of each other (although most are). While the windows technique most times excludes relations athwart from the partition of windows and this may concomitantly affect result across each window size selection [8]. Causal assumption associated with time-series dataset include; Markov condition, faithfulness, ancestral graphs assumption with their extensions (dynamic, partial, maximum etc.,)[80].

6.4 Deterministic Data setting: Causal identifiability in a deterministic bivariate dataset, where the dataset does not contain noise exist. The mathematical model for a deterministic case is given as: $Y = f_y(X)$. Hence, since there is no noise term ε_y in the dataset, the technique of using the independence between the additive noise term ε_y alongside the cause variable X written as: $(X \perp\!\!\!\perp \varepsilon_y)$ to determine the causal direction in the datasets would not be applicable. Albeit Janzing et al.[81], proposed a method for deterministic cases, where the transformation function f_y and the causal variable's marginal distribution P_x can be exploited, based on some contrived perspective assumption known as information-geometric condition, aside the Markov and causal sufficiency, to describe the causal asymmetry in order to ascertain the causal direction in the dataset.

6.5 Heterogeneous and Nonstationary Datasets*:* Heterogeneous and nonstationary dataset are those sets of datasets where the process of generating the dataset is not uniform or identical, but rather changes across the dataset over time [8]. In cases such as these, even if the parameters and other mechanism vary but the qualitative domain knowledge which forms the causal structure is fixed, the causal discovery process may be feasible even if there is a shift in the distribution; as distribution shifts and causal modeling are closely related. Inspired by this distribution shift and nonstationary datasets, Huang et al.,[82] proposed a method called the Nonstationary Driving Force Estimation (NoDFEs), a Kernel Embedding (KE) technique [83]. While Zhang et al., [84] proposed frameworks called Constraint-based causal Discovery from Nonstationary/heterogeneous Data (CD-NOD), for these kinds of dataset. Both frameworks by [82] and [84] have capacities to do the following: (i) detects changes in the mechanism of the distribution, (ii) estimate causal skeletons in the distribution, (iii) identify the causal directions in the distribution and (iv) estimate the driving force behind nonstationary dataset. The causal assumptions employed in these frameworks includes; the Markov condition, a kind of pseudo- sufficiency, faithfulness, acyclicity and other assumptions that considers the independent changes in changing modules of the variables.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 06 Issue 11 November 2023
DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789
IJCSRR @ 2023

6.6 Missing Data Challenge: Datasets are sometimes plagued with the issue of missing values in the variables distribution. This issue is common and ubiquitous in virtually all domains. Thus with this predicament, it is futile to apply existing causal techniques to ascertain the causal structure in the dataset as the end result would be error bound; as the conditional (in)dependencies from such a dataset with the missing values is likely to generate a different causal structure from the same dataset without missing values. A proposition to mitigate or overcome the issue of missing data is given by Tu et al.,[85] where they employed the modified version of the PC algorithm called missing values PC (MVPC) to determine the causal direction in such a dataset. The MVPC algorithm considered dataset that were either: missing at random (MAR), missing completely at random (MCAR), and missing not at random (MNAR). Thus, the output of this technique may not always guarantee accuracy, albeit it is said to be asymptotically accurate based on the following assumptions: non-causality in missing-ness, Markov, faithfulness, causal sufficiency for confounding and selection bias assumption and other sundry assumptions [85].

6.7 Measurement Errors Issue: Causal discovery output can greatly be tainted by the presence of measurement errors and thus, make useless the causal technique employed in such a process. These measurement errors occur as a result of the instruments or proxies employed in the process of measurement. Since this issue has become ubiquitous, a great deal of attention has been given to it with little success. Albeit a recent study by Zhang et al.,[86] has shown that under sufficient conditions (sets of assumptions related to the Markov condition, causal sufficiency, linearity, and the non-deterministic faithfulness condition and others), the causal structure or direction in the measurement error prone dataset that is partially or completely free of the measurement errors with undetermined variance can be generated. The measurement error causal algorithm which they called, causal model with measurement error (CAMME) is expressed mathematically as: $X = \hat{X} + \varepsilon$. Where X is a set of observed variables, and $\hat{X}$ is a set of variables that are free of measurement-noise, and ε is the measurement errors set. It is assumed that ε is independent of $\hat{X}$: $(\varepsilon \coprod \hat{X})$, and $\hat{X}$ possess a variance of non-zero. With this study carving a niche in the measurement error scenario, it is hoped that it will spur researchers to build on the work and come up with better approaches of causal discovery in regards to the issue of measurement error.

6.8 Data Selection Bias Issue: when the data is selected in such a way that during the statistical analysis process, it become difficult to achieve a proper randomization of data sample obtained, which also translate to indicate that the sampled data does not reflect the true population of the study. When this happened, the sampled dataset is said to suffer a selection bias or selection effect [87, 88]. It is sometime referred simply as the distortion in the statistical analysis process which is an outcome of the data collection process[89, 90]. Selection bias can taint the statistical and causal significance of a study and thus, falsify the outcome. However, a proposition by Zhang et al., [91] designed two FCM algorithms for a bivariate dataset with selection bias. The first one is based on the post nonlinear model which they called post nonlinear for outcome-dependent selection bias (PNL-OSB) and the second model was built using the additive nonlinear model with the same outcome-dependent selection bias which they called (ANM-OSB) for bivariate dataset, for causal identifiability and causal direction ascertainment in outcome-dependent selection bias dataset. Assumption used in the study are similar to the assumptions in the LinGaM and PNL frameworks apart from the outcome-dependent assumption. However, in spite of their specific approach on this issue, a more general approach regarding selection bias in datasets still needs to be studied.

6.9 Data setting with Confounding and Latent Variables: Causal discovery or inference in a dataset that is beclouded with confounding, latent, hidden or unmeasured variables can pose a major challenge to causal identifiability in such as dataset setting. In the case where there are latent or unmeasured confounding variable, the effects of these latent variables can be felt by the observable variables during the process of causal identifiability. Many causal techniques have been developed to overcome this confounding and latent variable issue. In the case of measured confounding variables in an i.i.d data, the causal identifiability can easily be determined by using the do-operator of the do-calculus as an intervention operation and the d-separation criteria proposed by Pearl [22, 92], under the Markov, sufficiency acyclicity and causal faithfulness assumptions, alongside algorithms such as the PC etc. However, if the latent variables are unmeasured or unobserved, then the sufficiency assumption is dropped, while the rest assumption holds sway. By using the causal identifiability techniques captured in the Fast Causal Inference (FCI) algorithm [39, 49] and its variants (RFCI, FCI+) [48, 93], the issue of latent and unmeasured variables are handled and the causal direction in such datasets are determined.

International Journal of Current Science Research and Review

ISSN: 2581-8341

Volume 06 Issue 11 November 2023

DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789

IJCSRR @ 2023

6.10 Datasets with Cycles or Feedback Setting: Virtually all the aforementioned challenges of causal discovery or inference discussed above are based on the full acyclicity or partial acyclicity conditions in the datasets. Dataset with variables having feedbacks or cycle is not a common phenomenon at all in the literature of causal community especially in computer science and statistics, albeit there exists important feedback tasks and applications in the social science and other disciplines like the laws of demand and supply in economics [94, 95] that involves graph with cycles. A number of causal techniques exists in extant literatures that tackles variables that form cycles or feedback in a sampled dataset. Prominent amongst them is the Cyclic Causal Discovery (CCD) algorithm proposed by Thomas [49]. This algorithm is applicable in dataset with variables with cycles but without latent variables. The CCD algorithms uses the partial ancestral graph (PAG) as output to describe a MEC of maximum ancestral graph (MAG) to determine the causal direction in the sampled dataset. However, there is a limitation of the CCD algorithm in most cases, the PAG that describe the data would not fit well on the dataset. A recent framework from Forri & Mooij [96] called modular structural causal model (*m*SCM) that handles dataset with cycles, latent variables and nonlinear structures exist. Their work introduces a set of mixed graphs which they called sigma connection graph (σ-CG) alongside other additional structure, with a concept called the sigma separation (σ-separation), which is an extrapolation and adaptation of the d-separation criteria to determine causal direction in nonlinear datasets with cycles and latent variables. A similar technique for handling causality in dataset with cycles, latent variables and selection bias (CLS) called Cycles Causal Inference (CCI), proposed by Strobl, [97] is also available in extant literature. This framework represents the cyclic graph involved in the causal process as a non-recursive linear structural equation model that has independent error terms. Empirical evidence shows that the CCI algorithm seems to outperform the CCD, FCI and RFCI algorithms regarding datasets with cycles [97]. Other approaches which adapts the LinGaM model of additive noise for causal discovery in cycles exists in this reference [98-100]. Table 1 below presented an aggregated summary of the data types settings with the algorithms, and assumptions employed alongside the authors pioneering them.

Table 1. A Summary of the different data types settings, with the algorithms employed alongside the assumptions and the authors of such initiatives

S/No.	Data Type Setting	Causal Algorithm Employed	Assumption(s)	Author(s)
1.	Linear Non-Gaussian	LinGaM and its extensions e.g., PNL	Markov, linearity, non-Gaussianity, acyclicity and (sufficiency only with author [67])	[59],[67] [60, 69] , [70], [64, 73]
2.	Nonlinear with Additive Noise	Nonlinear LinGaM and its extensions e.g., PNL	Markov, acyclicity, nonlinearity additive noise	[62], [74], [64, 73]
3.	Time series	Ganger Auto-regression, Ts-FCI, PCMCI (and its extensions; PCMCI+ LPCMCI)	Markov, faithfulness, ancestral graphs	[65],[75], [76], [77], [78], [79]
4.	Deterministic Data	FCM or LinGaM without noise	Markov, sufficiency and information-geometric	[81]
5.	Heterogeneous/Nonstationary	NoDFE, KE,	Markov, sufficiency (pseudo), faithfulness, acyclicity and independent changes in changing modules	[82], [83], [84]

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 06 Issue 11 November 2023
DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789
IJCSRR @ 2023

6.	Missing Data	MVPC	Markov, faithfulness, sufficiency for confounding and selection bias, non-causality in missing-ness and others.	[85]
7.	Measurement Errors	CAMME	Markov, sufficiency, linearity, puny acyclicity, non-deterministic faithfulness, and other sundry conditions	[86]
8.	Selection Bias	PNL-OSB, ANM-OSB	Markov, sufficiency, linearity, non-Gaussianity, acyclicity, linearity additive noise, and outcome-dependent	[91]
9.	Confounding and Latent Variables	PC, FCI, RFCI, FCI+	Markov, sufficiency (dropped with the FCIs), acyclicity, and faithfulness	[22, 92], [39, 49], [48, 93]
10.	Cycles or feedback data	CCD, *m*SCM, CCI, LinGaM	Markov, Cyclic, sufficiency (dropped with the FCIs), (non)linearity, Additive noise, and faithfulness	[49], [66], [96], [97], [98-100]

7. CONCLUSION

Having a good working knowing of causation is crucial and germane, as it helps man to articulate and predict his environment and also impact it positively; knowing the appropriate intervention to execute and the expected positive outcome from it, and avoiding experiments or interventions that are inimical and concomitant with negative consequences or outcome. Causality may not be a piece of cake that can easily be devour even with the evolution of modern scientific and advanced technological techniques, as the standard procedure for it is still the RCT, credit to Fisher [1], which involving separating the study population into treated and controlled groups in order to avoid any confounding issues with the study population, that may negate the outcome of the experiment. Hence, with this method, situations often arise when the RCT method becomes unethical, infeasible or too expensive to perform. Plagued with these anomalies and inadequacies, researchers have resorted to the study of causality in observational data settings. With the huge successes recorded in causality in observational settings especially with i.i.d, datasets that are mostly multinomial and Gaussian in nature that utilizes assumptions such as the Markov condition, causal sufficiency, faithfulness and acyclicity assumptions abound in extant literatures. However, various data settings exist which do not confirm to the general standard of i.i.d. and the normal distribution. Different data types settings such as; linear and non-Gaussian, Nonlinear & non-Gaussian, datasets with missing values, datasets beclouded with selection bias, datasets with cycles variables, datasets with heterogeneous/nonstationary variables, datasets with confounding or latent variables, time-series datasets etc., exists but are given little attention by researchers of causality in observational settings. Thus, this study aggregates the solutions techniques and the different assumptions advanced

International Journal of Current Science Research and Review

ISSN: 2581-8341

Volume 06 Issue 11 November 2023

DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789

IJCSRR @ 2023

for the aforementioned data types, in order to present researchers in observational data causality with a basic understanding of how causality techniques are performed and identified across these types of datasets, with the view to opening understanding and encourage new and innovative works in this areas.

Conflict of Interest Declaration

All authors have no financial or proprietary interests in any material discussed in this work.

REFERENCES

1. Box, J.F., *RA Fisher, the Life of a Scientist.* Revue Philosophique de la France Et de l, 1980. 170(4).
2. Benson, K. and A.J. Hartz, *A comparison of observational studies and randomized, controlled trials.* New England Journal of Medicine, 2000. 342(25): p. 1878-1886.
3. Silverman, S.L., *From randomized controlled trials to observational studies.* The American journal of medicine, 2009. 122(2): p. 114-120.
4. Pearl, J., *Causal inference in statistics: An overview.* Statistics surveys, 2009. 3: p. 96-146.
5. Nogueira, A.R., et al., *Methods and tools for causal discovery and causal inference.* Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2022: p. e1449.
6. Morgan, S. and C. Winship, *Counterfactuals and causal inference. Cambridge University Press*. 2007, Cambridge University Press New York, NY.
7. Guo, R., et al., *A survey of learning causality with data: Problems and methods.* ACM Computing Surveys (CSUR), 2020. 53(4): p. 1-37.
8. Glymour, C., K. Zhang, and P. Spirtes, *Review of causal discovery methods based on graphical models.* Frontiers in genetics, 2019. 10: p. 524.
9. Yao, L., et al., *A survey on causal inference.* ACM Transactions on Knowledge Discovery from Data (TKDD), 2021. 15(5): p. 1-46.
10. Pearl, J. and D. Mackenzie, *The book of why: the new science of cause and effect*. 2018: Basic books.
11. Hitchcock, C. and M. Rédei, *Reichenbach's common cause principle.* 2020.
12. Pearl, J., *Causality*. 2009: Cambridge university press.
13. Peters, J., D. Janzing, and B. Schölkopf, *Elements of causal inference: foundations and learning algorithms*. 2017: The MIT Press.
14. Neyman, J., *Sur les applications de la theorie des probabilites aux experiences agricoles: essai des principes (Masters Thesis); Justification of applications of the calculus of probabilities to the solutions of certain questions in agricultural experimentation. Excerpts English translation (Reprinted).* Stat Sci, 1923. 5: p. 463-472.
15. Rubin, D.B., *Estimating causal effects of treatments in randomized and nonrandomized studies.* Journal of educational Psychology, 1974. 66(5): p. 688.
16. Spirtes, P., C. Glymour, and R. Scheines, *Discovery algorithms for causally sufficient structures*, in *Causation, prediction, and search*. 1993, Springer. p. 103-162.
17. Greenland, S., J. Pearl, and J.M. Robins, *Causal diagrams for epidemiologic research.* Epidemiology, 1999: p. 37-48.
18. Lauritzen, S.L., *Causal Inference from.* Complex stochastic systems, 2000: p. 63.
19. Neal, B., *Introduction to causal inference from a machine learning perspective.* Course Lecture Notes (draft), 2020.
20. Holland, P.W., *Statistics and causal inference.* Journal of the American statistical Association, 1986. 81(396): p. 945-960.
21. Eberhardt, F., *Introduction to the foundations of causal discovery.* International Journal of Data Science and Analytics, 2017. 3(2): p. 81-91.
22. Halpern, J.Y., *The Book of Why, Judea Pearl, Basic Books (2018)*. 2019, Elsevier.
23. Elwert, F., *Graphical causal models*, in *Handbook of causal analysis for social research*. 2013, Springer. p. 245-273.
24. Pearl, J., *Probabilistic reasoning in intelligent systems: networks of plausible inference*. 1988: Morgan kaufmann.

International Journal of Current Science Research and Review

ISSN: 2581-8341

Volume 06 Issue 11 November 2023

DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789

IJCSRR @ 2023

25. Gultchin, L., et al. *Differentiable causal backdoor discovery*. in *International Conference on Artificial Intelligence and Statistics*. 2020. PMLR.
26. Correa, J. and E. Bareinboim. *Causal effect identification by adjustment under confounding and selection biases*. in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2017.
27. Tian, J. and J. Pearl, *A general identification condition for causal effects*. 2002: eScholarship, University of California.
28. Pearl, J., *Theoretical impediments to machine learning with seven sparks from the causal revolution.* arXiv preprint arXiv:1801.04016, 2018.
29. Richardson, T.S. and J.M. Robins. *Single world intervention graphs: a primer*. in *Second UAI workshop on causal structure learning, Bellevue, Washington*. 2013. Citeseer.
30. Zhang, J. and P. Spirtes, *Intervention, determinism, and the causal minimality condition.* Synthese, 2011. 182(3): p. 335-347.
31. Hauser, A. and P. Bühlmann, *Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs.* The Journal of Machine Learning Research, 2012. 13(1): p. 2409-2464.
32. Hauser, A. and P. Bühlmann, *Jointly interventional and observational data: estimation of interventional Markov equivalence classes of directed acyclic graphs.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2015. 77(1): p. 291-318.
33. Sharma, S., et al., *Monitoring protein conformation along the pathway of chaperonin-assisted folding.* Cell, 2008. 133(1): p. 142-153.
34. Uhler, C., et al., *Geometry of the faithfulness assumption in causal inference.* The Annals of Statistics, 2013: p. 436-463.
35. Mayrhofer, R. and M.R. Waldmann, *Sufficiency and necessity assumptions in causal structure induction.* Cognitive science, 2016. 40(8): p. 2137-2150.
36. Zhang, J. and W. Mayer, *Weakening faithfulness: some heuristic causal discovery algorithms.* International journal of data science and analytics, 2017. 3(2): p. 93-104.
37. Zhang, J. and P.L. Spirtes, *Strong faithfulness and uniform consistency in causal inference.* arXiv preprint arXiv:1212.2506, 2012.
38. Agresti, A., *Two Bayesian/frequentist challenges for categorical data analyses.* Metron, 2014. 72(2): p. 125-132.
39. Spirtes, P., et al., *Causation, prediction, and search*. 2000: MIT press.
40. Pena, J.M. *Learning gaussian graphical models of gene networks with false discovery rate control*. in *European conference on evolutionary computation, machine learning and data mining in bioinformatics*. 2008. Springer.
41. Nyberg, E. and K. Korb, *Informative interventions.* Causality and probability in the sciences. College Publications, London, 2006.
42. Meek, C., *Strong completeness and faithfulness in Bayesian Networks. In uncertainty in artificial intelligence: Proceedings of the eleventh conference*. 1995, San Francisco, CA: Morgan Kaufmann.
43. Pearl, J. and T.S. Verma, *A theory of inferred causation*, in *Studies in Logic and the Foundations of Mathematics*. 1995, Elsevier. p. 789-811.
44. Kalisch, M. and P. Bühlman, *Estimating high-dimensional directed acyclic graphs with the PC-algorithm.* Journal of Machine Learning Research, 2007. 8(3).
45. Ramsey, J.D., *A scalable conditional independence test for nonlinear, non-Gaussian data.* arXiv preprint arXiv:1401.5031, 2014.
46. Sejdinovic, D., et al., *Equivalence of distance-based and RKHS-based statistics in hypothesis testing.* The Annals of Statistics, 2013: p. 2263-2291.
47. Zhang, K., et al., *Kernel-based conditional independence test and application in causal discovery.* arXiv preprint arXiv:1202.3775, 2012.
48. Colombo, D., et al., *Learning high-dimensional directed acyclic graphs with latent and selection variables.* The Annals of Statistics, 2012: p. 294-321.

International Journal of Current Science Research and Review

ISSN: 2581-8341

Volume 06 Issue 11 November 2023

DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789

IJCSRR @ 2023

49. Spirtes, P.L., C. Meek, and T.S. Richardson, *Causal inference in the presence of latent variables and selection bias.* arXiv preprint arXiv:1302.4983, 2013.
50. Richardson, T., *Feedback models: Interpretation and discovery*. 1996, Ph. D. thesis, Carnegie Mellon.
51. Zhang, J. and P. Spirtes, *The three faces of faithfulness.* Synthese, 2016. 193(4): p. 1011-1027.
52. Schwarz, G., *Estimating the dimension of a model.* The annals of statistics, 1978: p. 461-464.
53. Maathuis, M.H. and P. Nandy, *A Review of Some Recent Advances in Causal Inference.* Handbook of big data, 2016: p. 387-407.
54. Chickering, D.M., *Optimal structure identification with greedy search.* Journal of machine learning research, 2002. 3(Nov): p. 507-554.
55. Chickering, D.M., D. Geiger, and D. Heckerman, *Learning Bayesian networks is NP-hard*. 1994, Citeseer.
56. Tsamardinos, I., L.E. Brown, and C.F. Aliferis, *The max-min hill-climbing Bayesian network structure learning algorithm.* Machine learning, 2006. 65(1): p. 31-78.
57. Wong, M.L., S.Y. Lee, and K.S. Leung. *A hybrid approach to discover Bayesian networks from databases using evolutionary programming*. in *2002 IEEE International Conference on Data Mining, 2002. Proceedings.* 2002. IEEE.
58. Heinze-Deml, C., M.H. Maathuis, and N. Meinshausen, *Causal structure learning.* Annual Review of Statistics and Its Application, 2018. 5: p. 371-391.
59. Shimizu, S., et al., *A linear non-Gaussian acyclic model for causal discovery.* Journal of Machine Learning Research, 2006. 7(10).
60. Hyvärinen, A., J. Karhunen, and E. Oja, *Independent component analysis, adaptive and learning systems for signal processing, communications, and control.* John Wiley & Sons, Inc, 2001. 1: p. 11-14.
61. Shimizu, S., et al., *DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model.* The Journal of Machine Learning Research, 2011. 12: p. 1225-1248.
62. Hoyer, P., et al., *Nonlinear causal discovery with additive noise models.* Advances in neural information processing systems, 2008. 21.
63. Hoyer, P.O., et al., *Causal discovery of linear acyclic models with arbitrary distributions.* arXiv preprint arXiv:1206.3260, 2012.
64. Zhang, K. and A. Hyvarinen, *On the identifiability of the post-nonlinear causal model.* arXiv preprint arXiv:1205.2599, 2012.
65. Hyvärinen, A., et al., *Estimation of a structural vector autoregression model using non-gaussianity.* Journal of Machine Learning Research, 2010. 11(5).
66. Lacerda, G., et al., *Discovering cyclic causal models by independent components analysis.* arXiv preprint arXiv:1206.3273, 2012.
67. Hoyer, P.O., et al., *Estimation of causal effects using linear non-Gaussian causal models with hidden variables.* International Journal of Approximate Reasoning, 2008. 49(2): p. 362-378.
68. Kano, Y. and S. Shimizu. *Causal inference using nonnormality*. in *Proceedings of the international symposium on science of modeling, the 30th anniversary of the information criterion*. 2003.
69. Kiviniemi, V., et al., *Independent component analysis of nondeterministic fMRI signal sources.* Neuroimage, 2003. 19(2): p. 253-260.
70. Dillon, W.R. and M. Goldstein, *Multivariate analysis: Methods and applications*. 1984: New York (NY): Wiley, 1984.
71. Humeniuk, R., et al., *Validation of the alcohol, smoking and substance involvement screening test (ASSIST).* Addiction, 2008. 103(6): p. 1039-1047.
72. Sanchez-Romero, R., et al., *Estimating feedforward and feedback effective connections from fMRI time series: Assessments of statistical methods.* Network Neuroscience, 2019. 3(2): p. 274-306.
73. Zhang, K. and L.-W. Chan. *Extensions of ICA for causality discovery in the hong kong stock market*. in *International Conference on Neural Information Processing*. 2006. Springer.

International Journal of Current Science Research and Review

ISSN: 2581-8341

Volume 06 Issue 11 November 2023

DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789

IJCSRR @ 2023

74. Mooij, J., et al. *Regression by dependence minimization and its application to causal inference in additive noise models*. in *Proceedings of the 26th annual international conference on machine learning*. 2009.
75. Granger, C.W., *Investigating causal relations by econometric models and cross-spectral methods.* Econometrica: journal of the Econometric Society, 1969: p. 424-438.
76. Entner, D. and P.O. Hoyer, *On causal discovery from time series data using FCI.* Probabilistic graphical models, 2010: p. 121-128.
77. Runge, J., et al., *Inferring causation from time series in Earth system sciences.* Nature communications, 2019. 10(1): p. 1-13.
78. Runge, J. *Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets*. in *Conference on Uncertainty in Artificial Intelligence*. 2020. PMLR.
79. Gerhardus, A. and J. Runge, *High-recall causal discovery for autocorrelated time series with latent confounders.* Advances in Neural Information Processing Systems, 2020. 33: p. 12615-12625.
80. Eichler, M. *Causal inference from time series: What can be learned from granger causality*. in *Proceedings of the 13th International Congress of Logic, Methodology and Philosophy of Science*. 2007. Citeseer.
81. Janzing, D., et al., *Information-geometric approach to inferring causal directions.* Artificial Intelligence, 2012. 182: p. 1-31.
82. Huang, B., et al. *Behind distribution shift: Mining driving forces of changes and causal arrows*. in *2017 IEEE International Conference on Data Mining (ICDM)*. 2017. IEEE.
83. Schölkopf, B., A.J. Smola, and F. Bach, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. 2002: MIT press.
84. Zhang, K., et al. *Causal discovery from nonstationary/heterogeneous data: Skeleton estimation and orientation determination*. in *IJCAI: Proceedings of the Conference*. 2017. NIH Public Access.
85. Tu, R., et al. *Causal discovery in the presence of missing data*. in *The 22nd International Conference on Artificial Intelligence and Statistics*. 2019. PMLR.
86. Zhang, K., et al., *Causal discovery in the presence of measurement error: Identifiability conditions.* arXiv preprint arXiv:1706.03768, 2017.
87. Institute, N.C., *Cancer*, in *Dictionary of Cancer Terms*. 2009, cancer.gov: online.
88. Westreich, D., *Berkson's bias, selection bias, and missing data.* Epidemiology (Cambridge, Mass.), 2012. 23(1): p. 159.
89. Hernán, M.A., S. Hernández-Díaz, and J.M. Robins, *A structural approach to selection bias.* Epidemiology, 2004: p. 615-625.
90. Kopec, J.A. and J.M. Esdaile, *Bias in case-control studies. A review.* Journal of epidemiology and community health, 1990. 44(3): p. 179.
91. Zhang, K., et al. *On the Identifiability and Estimation of Functional Causal Models in the Presence of Outcome-Dependent Selection*. in *UAI*. 2016.
92. Pearl, J., *Causal inference.* Causality: objectives and assessment, 2010: p. 39-58.
93. Claassen, T., J. Mooij, and T. Heskes, *Learning sparse causal models is not NP-hard.* arXiv preprint arXiv:1309.6824, 2013.
94. Agarwal, A. and R. Shankar, *Modeling supply chain performance variables.* Asian Academy of Management Journal, 2005. 10(2): p. 47-68.
95. Spirtes, P., *Building causal graphs from statistical data in the presence of latent variables*, in *Studies in Logic and the Foundations of Mathematics*. 1995, Elsevier. p. 813-829.
96. Forré, P. and J.M. Mooij, *Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders.* arXiv preprint arXiv:1807.03024, 2018.
97. Strobl, E.V., *A constraint-based algorithm for causal discovery with cycles, latent variables and selection bias.* International Journal of Data Science and Analytics, 2019. 8(1): p. 33-56.

International Journal of Current Science Research and Review

ISSN: 2581-8341

Volume 06 Issue 11 November 2023

DOI: 10.47191/ijcsrr/V6-i11-10, Impact Factor: 6.789

IJCSRR @ 2023

98. Richardson, T. *A discovery algorithm for directed cyclic graphs. Uncertainty in Artificial Intelligence*. in *Proceedings, 12th Conference, Morgan Kaufman, CA*. 1996.
99. Mooij, J. and T. Heskes, *Cyclic causal discovery from continuous equilibrium data.* arXiv preprint arXiv:1309.6849, 2013.
100. Mooij, J.M., et al., *On causal discovery with cyclic additive noise models.* Advances in neural information processing systems, 2011. 24.

Cite this Article: Gabriel Terna Ayem, Salu George Thandekkattu, Augustine Shey Nsang (2023). A Review of Causal Identifiability Techniques across Different Observational Datasets. International Journal of Current Science Research and Review, 6(11), 7054-7072