Causal Discovery for Image Datasets via Feature Extraction

Atul Rawal[a], Adrienne Raglin[b], Qianlong Wang[a], and Ziying Tang[a]

[a,]Department of Computer & Information Sciences, Towson University, Towson, MD, USA.
[b]DEVCOM Army Research Laboratory, Adelphi, MD, USA

ABSTRACT

Artificial reasoning systems via Artificial Intelligence (AI) and Machine Learning (ML) have made tremendous progress within the past decade. AI/ML systems have been able to reach unprecedented new levels of autonomy for a multitude of applications ranging from autonomous vehicles to biomedical imaging. This new level of intelligence and freedom for AI/ML systems requires them to have a degree of human-like intelligence in terms of causation beyond the correlation. This, however, has remained a major challenge for investigators when combining causality with AI/ML systems. AI/ML systems that are capable of generating cause and effect relationships are still in their infancy, as the literature highlights. The lack of investigations for causal reasoning systems that are capable of using datasets other than tabular data is well highlighted within literature. Causal learning for image, audio, video, radio-frequency, and other modalities still remain a major challenge. While there are open-source tools available for causal learning with tabular data, there is a lack of tools for other modalities. To this extent, this study proposes a causal learning method with image datasets by using existing tools and methodologies. Specifically, we propose to use existing causal discovery toolboxes for investigating causal relations within image datasets by converting image datasets into tabular form with feature extraction using tools such as auto-encoders and deep neural networks. The converted dataset can then be used to generate causal graphs by using tools such as the Causal Discovery Toolbox to highlight the specific cause and effect relations within the data. For AI/ML systems using causal learning for image datasets via existing tools and methodologies can provide an extra layer of robustness to ensure fairness and trustworthiness.

Keywords: Causality, Causal Learning, Machine Learning, Artificial Intelligence,

1. INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) systems have made remarkable progress in the past decade. This progress has caused AI/ML systems to be effortlessly assimilated into our daily lives. Examples include the elementary recommendation systems used in Netflix/Amazon/Spotify to more complex classification systems used for tumor detection and other biomedical applications.[1–3] Due to the unprecedented integration of AI systems within our lives, challenges related to trust, fairness, and responsibility have taken center-stage for the utilization of these systems. Exciting new fields of research (both theoretical and experimental) have seen an increase interest to aid in addressing the challenges associated with the use of AI/ML systems. Research areas such as causal learning, which is based on the field of causality, have been a focal point for artificial reasoning systems that are capable of human-like thinking. The trivial task of generating cause and effect relationships for the human mind is not so trivial for artificial reasoning systems.

Causality is the cause-and-effect relation between variables/features in a dataset which has a been a core concept for science throughout the years. The *cause* between the variables describes the *why* while the *effect* describes the *what*[4,5] It is most commonly used incorrectly as a synonym for statistical correlation, even though it has been thoroughly defined that correlation does not imply causation. Even though statistical correlation has been the backbone of many scientific theories and discoveries, interpreting it as causation can have unwarranted effects. Causality when applied to machine learning can be referred to causal learning. It refers the investigation of cause-and-effect relations between different variables in datasets for AI/ML systems. Therefore, causal learning should be viewed as a critical core component of any artificial reasoning system, not just an ad-hoc feature.

Even though the field of causality has been around for quite some time and has been utilized in other topics such as economics and finance, causal learning for AI/ML systems is still a relatively new field. Due to the relatively young age of the field, advancements in causal learning have not kept pace with the progress that AI/ML systems have observed. This can partly be attributed to the challenges that hinder the field's progress such as the lack of robust observational datasets with ground truth and the limitations on the modality of the datasets that can be used.[6–8] In order to ensure that causal learning is on par with advanced AI/ML systems new methodologies and perspectives must be investigated that address the associated challenges. For example, in order to perform robust causal investigations, experimental studies via randomized control trials have been shown to be the golden standard as observational datasets lack the ground truth needed to establish the cause and effect relations within features. However, researchers often have to utilize observational datasets as obtaining experimental data is not always feasible and AI/ML systems are most commonly trained on observational datasets. Even though a few studies have tried to address this issue, existing tools and methods focus mostly on tabular datasets leaving a gap for causal studies of other modalities such as image and video.[9–11] Studies investigating causal relations within image datasets are scarce in literature. To address this challenge, we present a novel perspective for causal learning to generate causal relations from observational image data via causal discovery. The paper is organized as follows: Section 2 provides the overview of causal learning and causal discovery, Section 3 provides the proof of concept example for causal discovery for image datasets, while Section 4 presents the results and discussion. Section 5 includes concluding remarks.

2. OVERVIEW OF CAUSAL LEARNING

Causality refers to the relation between a cause and its effect.[12] In relation to artificial reasoning systems, this highlights the model's ability to reason with causation and surpass statistical correlation and association. Causal learning for AI/ML systems investigates the change in the model predictions when modifications/manipulations to a feature cause a change in another feature. The feature/variable being modified is referred to as the *treatment*, and the variable whose change is being investigated is called the *outcome*. Other features within the dataset that can cause a change in both the treatment and the outcome are called *confounders*, while background/noise variables are referred to as the *covariates*.

Judea Pearl classified causal relations between variables in a dataset into three separate categories called the causal hierarchy : *association, intervention,* and *counterfactuals*.[5,13,14] Association, the first level of hierarchy refers to statistical correlation between the variables in the dataset. Modern AI/ML systems rely on association to make informed data-driven predictions based on the correlations observed within the dataset. The second level of the causal hierarchy, *intervention* utilizes the causal structure between variables to investigate the effect of changes to the treatment on the outcome. Finally, the last level of the hierarchy, *counterfactuals,* combines both the association and intervention to generate causal relations underlying the two former levels to make predictions based on unknown outcomes.

For AI/ML causal learning is utilized to answer two basic questions within a plethora of applications:

1. How much effect does a change in the treatment have on the outcome?
2. Which treatment needs to be modified to observe a change in the outcome?

These two questions form the basis of causal learning with machine learning and are referred to as *causal discovery* and *causal inference*.[15,16] Causal discovery can be used to highlight the existing causal relations between the variables in a dataset for AI/ML systems. Causal inference can be utilized to study the extent to which a treatment can be manipulated to effect a change in the outcome. Two formal frameworks are available to investigate causal discovery and causal inference: structural causal models (SCMs) and *potential outcome framework*. Structural causal models consist of *causal graphs* and *structural equations* to describe a holistic theory for causality.[12,13,17,18] Causal graphs describe the causal relations between the variables in a dataset by utilizing a directed graph with different variables such as the treatment, outcome, confounders and covariates, represented by nodes.[12] The potential outcome framework relates the treatment applied to a specific *unit* , which is defined as an atomic research object in the treatment effect study.[19] Here the potential outcome for an event is the outcome of the instance if the treatment was applied. The framework categorizes the estimation of the treatment effects on the outcome at three population level; *individual treatment effect* (ITE), *average treatment effect* (ATE), and *conditional average treatment effect* (CATE).[12,13,17,18]

2.1 Causal Discovery for Image Data

As mentioned before, causal learning for AI/ML systems is still a relatively new field. Therefore most of the existing methodologies and techniques for both causal discovery and causal inference can be utilized for tabular data only. Even with a plethora of studies describing the use of causal discovery for different applications, there is a lack of studies highlighting the use of causal discovery with image datasets. While a few studies highlight causal discovery for images, extensive work is required to create robust datasets, methods and tools. Here we provide a brief overview of the some existing studies that utilize image datasets for causal discovery.

Castro et al., emphasized the use of causal learning for medical imaging applications.[20] The article presented the use of causal graphs to highlight the causal relations between images and the annotations. *Causal* and *anti-causal* tasks, were introduced in the study with causal tasks described as tasks where an effect is predicted from the cause, and anti-causal tasks described as tasks where the cause is predicted from the effect. Using causal and anti-causal tasks, the author provided an example of a causal graph for cancer classification and accentuated the use of causal learning for addressing existing issues within the medical imaging field, such as data scarcity and mismatch.

Chalupka et al., presented a new framework for generalizing causal learning in settings by reconstructing the causal variables from micro-variables for visual causes in images.[21] Here the visual causes were defined as "a function/feature of raw image pixels that has a causal effect on a target behavior of a perceiving system of interest." The authors utilized causal learning to enhance the performance of correlation-based classifiers via the causal manipulator network. It is able to detect causal features and perform causal modifications based on those features.

Lopez-Paz et al., presented the Neural Causation Coefficient (NCC) to derive causation from image datasets.[22] The article presented the use of "causal dispositions" of different objects within images. A classifier is used to derive causal relations between random variables in the image datasets with samples from the joint distribution. A causal direction classifier was used to distinguish between object features and context features in images. Causal relations in images between objects and contexts are highlighted in image datasets by using statistical properties of the images.

Li et al., derived causal models from MRI images and clinical variables of patients with Alzheimer's via causal image synthesis.[23] Here the authors used structural causal models in combination with a styled generator to derive the causal relations and merge the images. Low dimensional latent feature representations of 3d images were used to generate causal relations between the images and the tabular data, which allowed for the derivations of counterfactual 3D brain MRI images.

While these studies provide a way forward to derive causal relations from image datasets they are still limited. The Chalupka et al., study needs synthetic scatter-plots for the images to train the classifier model and distinguish between the object and context in the images. The absence of the scatter-plots would be a major challenge in generating causal signals in the images. The causal features in the image are described as the features that cause the presence of the object in the scene. Other studies have featured an active learning scheme which would require SME validation and could be a potential challenge when domain knowledge isn't available.

3. PROOF OF CONCEPT EXAMPLE

3.1 Data Source

The dataset for this study, the RAREPLANES dataset was derived from CosmiQ Works open-source data repository via the AWS open-data program (https://www.cosmiqworks.org/rareplanes/). It includes a combination of real and synthetic satellite imagery of various airplanes in different settings such as airfields and airports. It consists of 253 satellite images of 14,700 annotated aircrafts for the real data portion and 50K synthetic satellite images of 630K annotated aircrafts. For this study we used the real dataset with 253 scenes and randomly selected annotated attributes for 1000 aircrafts based on two classes: military or civilian. The dataset included nine features/attributes: number of tail-fins, FAA class, wing position, wing shape, wingspan, length, canards, propulsion, and number of engines (Fig 1).

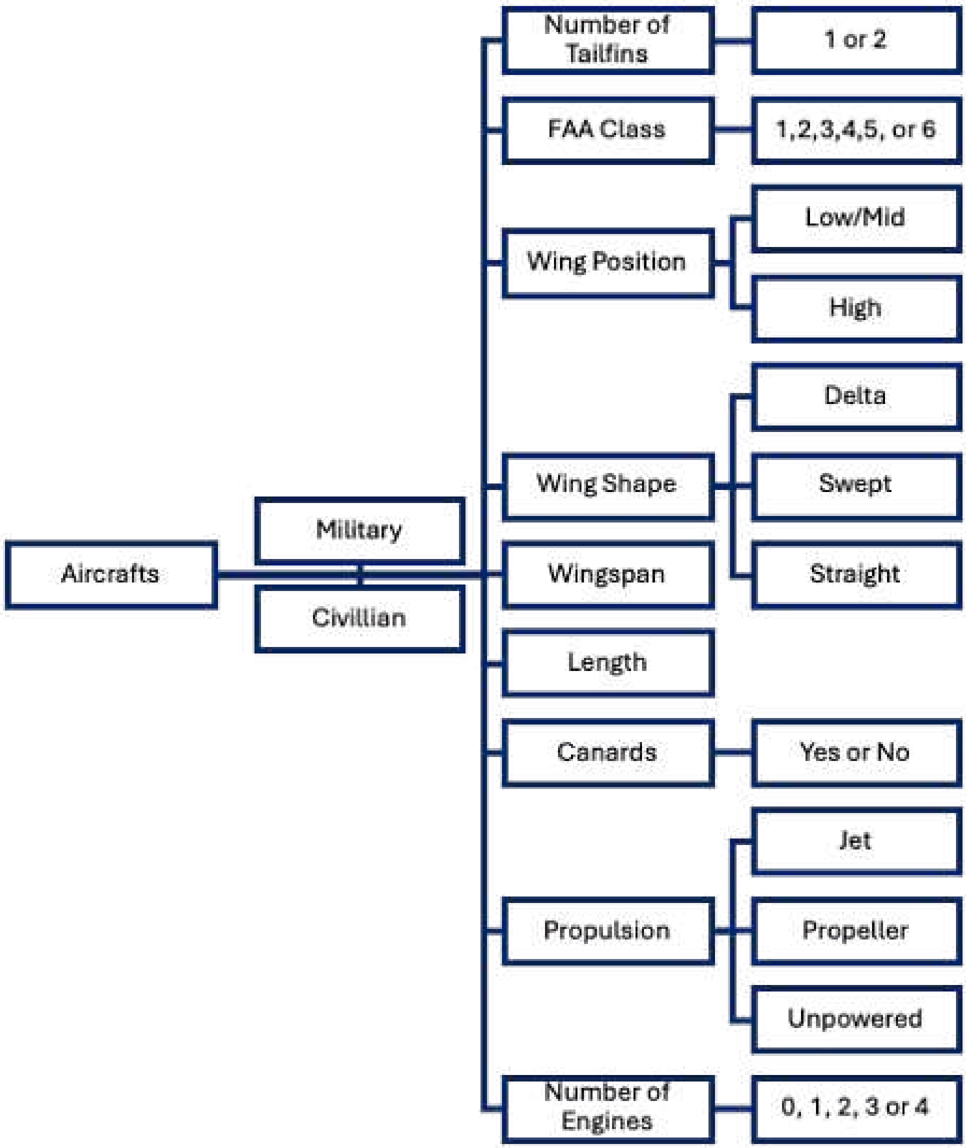

Figure 1: Attributes for each aircraft in the dataset used for this study

3.2 Methods

As mentioned before there is still a lack of a proper framework for generating causal graphs directly from image datasets. Therefore we propose a novel framework of existing methodologies to generate causal relations via causal graphs from observational image datasets (Fig 2). Mainly we propose to use existing tools/techniques into a workflow to generate causal graphs from observational image datasets. This consists of the following three components: attribute extraction from images, conversion of extracted features to tabular data, and causal discovery on the tabular data.

The first step of the proposed framework is attribute/annotation extraction from images. Feature representations in the form of annotated attributes can be extracted from raw image files (.json in this case) and saved in a tabular format to be utilized for causal discovery. These extracted attribute annotations can be used to perform causal discovery and generate causal graphs. A few studies presented in the previous section have highlighted the extraction of causal relations from image datasets such as.[22] The current study here is the first to extract causal features by generating causal graphs directly from observational image datasets.

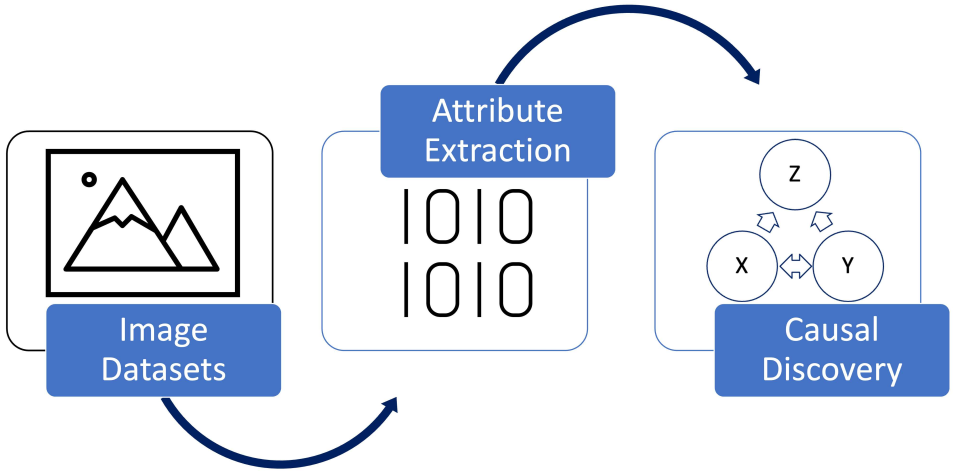

Figure 2: Proposed workflow for generating causal relations form image datasets

Once the pertinent features were finalized for the current study, all the attribute annotations were exported from multiple .json files into a single .csv file in tabular format such as for the the number of engines for the aircrafts, the presence/lack of canards, wing shape, and number of tail-fins. The finalized .csv file with the attribute data in a tabular form was then imported into a Python notebook using the pandas library. Causal relations between features were investigated for the dataset using the open-source causal discovery library Causal Discovery Toolbox (CDT). A skeleton graph from the raw data was generated to perform pairwise independence test. After that causal discovery is performed using the Peter Clark (PC) model within CDT. This generates the directed causal graphs with causal relations between the different variables. The PC algorithm is one of the most commonly used models for causal discovery in observational datasets. It consists of identifying the undirected causal graph, and then generating the directed edges between the nodes. To identify the causal relations the algorithm checks for existence of conditioned independent variables within a condition set. Then colliders are identified to direct the edges between the variables.

4. RESULTS & DISCUSSION

The generated causal graph yielded six features with a direct one-way causal impacts to the aircraft label; a) number of tail-fins, b) FAA class, c) canards, d) wing position, e) wing shape and f) number of engines. It also highlighted the causal relations between propulsion/wingspan/wing-shape/number of engines, canards/length/wing-shape/wing-position/number of tail-fins - FAA Class, which has a direct impact on the aircraft labels. As an example, the presence of canards is one of the easiest way of classifying a military and civilian aircraft as a most, if not all of these don't have canards. This is highlighted by the direct impact of canards for the classification in the causal graph. From the graph we can deduce that out of the nine features the six with the directed edge had a higher impact on the aircraft classification while wingspan, propulsion and length had an indirect causal relation. While this study is intended only as a proof-of-concept, it highlights the need of causal learning for artificial reasoning systems to obtain a human-like thinking. The addition of causal learning provides an extra-layer of robustness to the explanations generated from correlation based classifiers to ensure the relations generated from explanations are not only correlated, but also causally relevant. Therefore providing causal explanations based on causal learning and XAI can provide the AI/ML systems with the robustness and trustworthiness needed for them to achieve accountability.

In this paper we have provided a framework for extracting causal relations from observational image datasets. The addition of causal learning for artificial reasoning provides a more robust AI/ML model where trustworthiness is achieved. While traditional AI/ML models are progressing towards achieving robustness and trustworthiness, sometimes correlation based explanations are insufficient. Therefore, the extra layer of robustness from the causally relevant features can be a simple yet effective method to ensure trustworthiness. We presented a simple and efficient yet novel method to ensure the use of observational data for trustworthy AI/ML systems. Due to

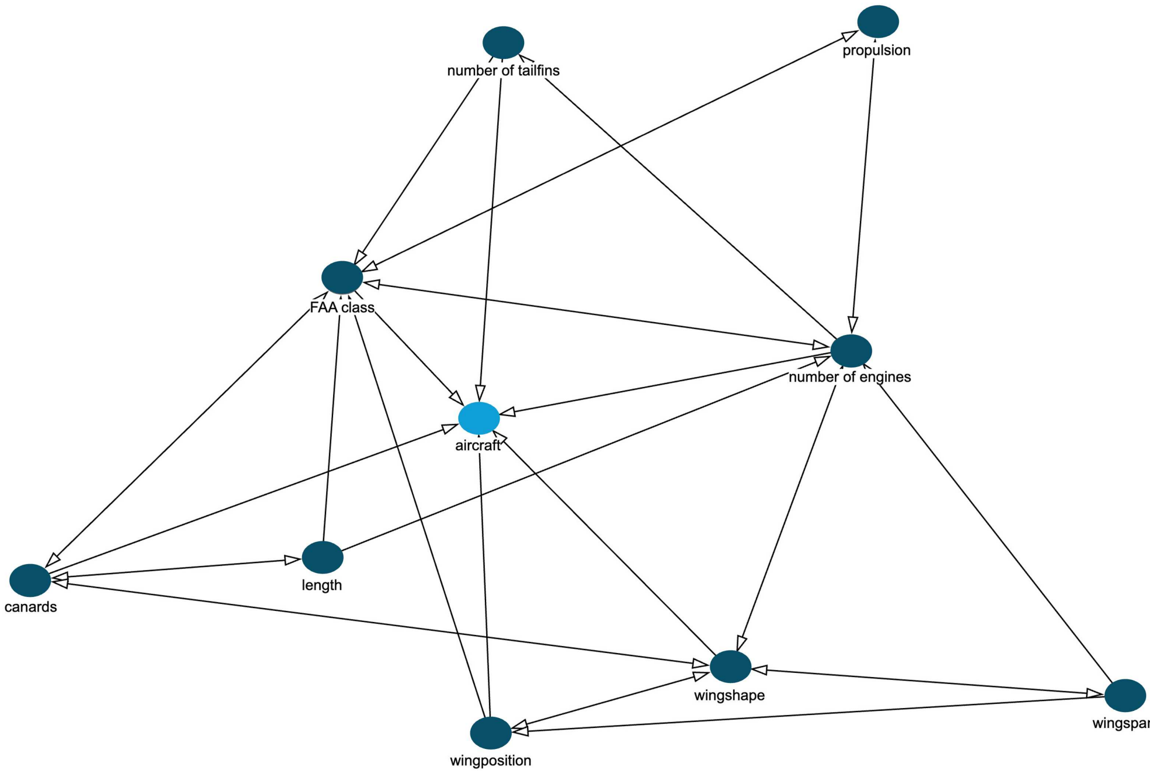

Figure 3: Causal graph for the RAREPLANES dataset

the lack of large observational datasets with available ground-truth, causal graphs generated via causal discover can play a vital role in generating robust and trustworthy explanations.

5. CONCLUSION

Artificial reasoning systems have made tremendous advancements in recent years. Yet, they lack the robustness and trust to achieve human-like thinking. Toward these goals of robustness and trustworthiness, causality must be included as a vital part of AI/ML systems. It can aide in achieving fairness, bias detection, & mitigation to provide a robust and trustworthy system . In this paper we have presented a novel perspective of applying causality to observational image data. For AI/ML systems using causal learning for image datasets via existing tools and methodologies can provide an extra layer of robustness to ensure fairness and trustworthiness.

ACKNOWLEDGMENTS

This research was conducted in collaboration with the Army Research Laboratory under CRADA Number 23-025 with the U.S. Army Research Laboratory. However, any opinion, finding, and conclusions or recommendations expressed in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the CRADA agency.

REFERENCES

[1] Athanasopoulou, K., Daneva, G. N., Adamopoulos, P. G., and Scorilas, A., "Artificial intelligence: the milestone in modern biomedical research," *BioMedInformatics* **2**(4), 727–744 (2022).

[2] Dixit, V. V. and Gulame, M. B., "Artificial intelligence and machine learning in biomedical applications," in [*Artificial Intelligence, Internet of Things (IoT) and Smart Materials for Energy Applications*], 101–116, CRC Press (2022).

[3] Petrick, L. M. and Shomron, N., "Ai/ml-driven advances in untargeted metabolomics and exposomics for biomedical applications," *Cell Reports Physical Science* **3**(7) (2022).

[4] Morgan, S. L. and Winship, C., [*Counterfactuals and causal inference*], Cambridge University Press (2015).

[5] Pearl, J., "Theoretical impediments to machine learning with seven sparks from the causal revolution," *arXiv preprint arXiv:1801.04016* (2018).

[6] Rawal, A., Mccoy, J., Rawat, D. B., Sadler, B., and Amant, R., "Recent advances in trustworthy explainable artificial intelligence: Status, challenges and perspectives," *IEEE Transactions on Artificial Intelligence* **1**(01), 1–1 (2021).

[7] Rawal, A., Raglin, A., Rawat, D. B., and Sadler, B. M., "Causality and Machine Learning Review," tech. rep., DEVCOM Army Research Laboratory (07 2022).

[8] Cheng, L., Guo, R., Moraffah, R., Sheth, P., Candan, K. S., and Liu, H., "Evaluation methods and measures for causal learning algorithms," *IEEE Transactions on Artificial Intelligence* **3**(6), 924–943 (2022).

[9] Rohrer, J. M., "Thinking clearly about correlations and causation: Graphical causal models for observational data," *Advances in methods and practices in psychological science* **1**(1), 27–42 (2018).

[10] Nichols, A., "Causal inference with observational data," *The Stata Journal* **7**(4), 507–541 (2007).

[11] Maathuis, M. H., Colombo, D., Kalisch, M., and Bühlmann, P., "Predicting causal effects in large-scale systems from observational data," *Nature methods* **7**(4), 247–248 (2010).

[12] Guo, R., Cheng, L., Li, J., Hahn, P. R., and Liu, H., "A survey of learning causality with data: Problems and methods," *ACM Computing Surveys (CSUR)* **53**(4), 1–37 (2020).

[13] Pearl, J., "Causal inference in statistics: An overview," *Statistics surveys* **3**, 96–146 (2009).

[14] Pearl, J., "The seven tools of causal inference, with reflections on machine learning," *Communications of the ACM* **62**(3), 54–60 (2019).

[15] Gelman, A., "Causality and statistical learning," (2011).

[16] Peters, J., Janzing, D., and Schölkopf, B., [*Elements of causal inference: foundations and learning algorithms*] (2017).

[17] Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., and Zhang, A., "A survey on causal inference," *ACM Transactions on Knowledge Discovery from Data (TKDD)* **15**(5), 1–46 (2021).

[18] Pearl, J., [*Causality*], Cambridge university press (2009).

[19] Imbens, G. W. and Rubin, D. B., [*Causal inference in statistics, social, and biomedical sciences*], Cambridge University Press (2015).

[20] Castro, D. C., Walker, I., and Glocker, B., "Causality matters in medical imaging," *Nature Communications* **11**(1), 3673 (2020).

[21] Chalupka, K., Perona, P., and Eberhardt, F., "Visual causal feature learning," *arXiv preprint arXiv:1412.2309* (2014).

[22] Lopez-Paz, D., Nishihara, R., Chintala, S., Scholkopf, B., and Bottou, L., "Discovering causal signals in images," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 6979–6987 (2017).

[23] Li, Y., Shi, J., and Zhou, S. K., "Causal image synthesis of brain mr in 3d," *arXiv preprint arXiv:2303.14349* (2023).