Causal Deep Learning: Encouraging Impact on Real-world Problems Through Causality

Other titles in Foundations and Trends® in Signal Processing

Energy-Based Models with Applications to Speech and Language Processing
Zhijian Ou
ISBN: 978-1-63828-306-5

Model-Based Deep Learning
Nir Shlezinger and Yonina C. Eldar
ISBN: 978-1-63828-264-8

Generalizing Graph Signal Processing: High Dimensional Spaces, Models and Structures
Xingchao Jian, Feng Ji and Wee Peng Tay
ISBN: 978-1-63828-150-4

Learning with Limited Samples: Meta-Learning and Applications to Communication Systems
Lisha Chen, Sharu Theresa Jose, Ivana Nikoloska, Sangwoo Park, Tianyi Chen and Osvaldo Simeone
ISBN: 978-1-63828-136-8

Causal Deep Learning: Encouraging Impact on Real-world Problems Through Causality

Jeroen Berrevoets
University of Cambridge
jeroen.berrevoets@maths.cam.ac.uk

Krzysztof Kacprzyk
University of Cambridge
kk751@cam.ac.uk

Zhaozhi Qian
University of Cambridge
zhaozhi.qian@maths.cam.ac.uk

Mihaela van der Schaar
University of Cambridge, and
The Alan Turing Institute
mv472@cam.ac.uk

the essence of knowledge
Boston — Delft

Foundations and Trends® in Signal Processing

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

J. Berrevoets *et al.*. *Causal Deep Learning: Encouraging Impact on Real-world Problems Through Causality.* Foundations and Trends® in Signal Processing, vol. 18, no. 3, pp. 200–309, 2024.

ISBN: 978-1-63828-401-7

Foundations and Trends® in Signal Processing
Volume 18, Issue 3, 2024

Editorial Scope

Foundations and Trends® in Signal Processing publishes survey and tutorial articles in the following topics:

- Adaptive signal processing
- Audio signal processing
- Biological and biomedical signal processing
- Complexity in signal processing
- Digital signal processing
- Distributed and network signal processing
- Image and video processing
- Linear and nonlinear filtering
- Multidimensional signal processing
- Multimodal signal processing
- Multirate signal processing
- Multiresolution signal processing
- Nonlinear signal processing
- Randomized algorithms in signal processing
- Sensor and multiple source signal processing, source separation
- Signal decompositions, subband and transform methods, sparse representations
- Signal processing for communications
- Signal processing for security and forensic analysis, biometric signal processing
- Signal quantization, sampling, analog-to-digital conversion, coding and compression
- Signal reconstruction, digital-to-analog conversion, enhancement, decoding and inverse problems
- Speech/audio/image/video compression
- Speech and spoken language processing
- Statistical/machine learning
- Statistical signal processing
  - Classification and detection
  - Estimation and regression
  - Tree-structured methods

Information for Librarians

Foundations and Trends® in Signal Processing, 2024, Volume 18, 4 issues. ISSN paper version 1932-8346. ISSN online version 1932-8354. Also available as a combined paper and online subscription.

Contents

Causal Deep Learning: Encouraging Impact on Real-world Problems Through Causality

Jeroen Berrevoets[1], Krzysztof Kacprzyk[1], Zhaozhi Qian[1] and Mihaela van der Schaar[1,2]

[1] *University of Cambridge, UK; jeroen.berrevoets@maths.cam.ac.uk, kk751@cam.ac.uk, zhaozhi.qian@maths.cam.ac.uk, mv472@cam.ac.uk*
[2] *The Alan Turing Institute, UK*

ABSTRACT

Causality has the potential to truly transform the way we solve a large number of real-world problems. Yet, so far, its potential largely remains to be unlocked as causality often requires crucial assumptions which cannot be tested in practice. To address this challenge, we propose a new way of thinking about causality-– we call this *causal deep learning*. Our causal deep learning framework spans three dimensions: (1) a structural dimension, which incorporates partial yet testable causal knowledge rather than assuming either complete or no causal knowledge among the variables of interest; (2) a parametric dimension, which encompasses parametric forms that capture the type of relationships among the variables of interest; and (3) a temporal dimension, which captures exposure times or how the variables of interest interact (possibly causally) over time. Our CDL framework enables us to precisely categorise and compare

Jeroen Berrevoets, Krzysztof Kacprzyk, Zhaozhi Qian and Mihaela van der Schaar (2024), "Causal Deep Learning: Encouraging Impact on Real-world Problems Through Causality", Foundations and Trends® in Signal Processing: Vol. 18, No. 3, pp 200–309. DOI: 10.1561/2000000123.

causal statistical learning methods. We use this categorisation to provide a comprehensive review of the CDL field. More importantly, CDL enables us to make progress on a variety of real-world problems by aiding us to leverage partial causal knowledge (including independencies among variables) and quantitatively characterising causal relationships among variables of interest (possibly over time). Our framework clearly identifies which assumptions are testable and which are not, so the resulting solutions can be judiciously adopted in practice. Our formulation helps us to combine or chain causal representations to solve specific problems without losing track of which assumptions are required to build these solutions, pushing real-world impact in healthcare, economics and business, environmental sciences and education, through causal deep learning.

1

Introduction

Causality holds the promise to transform the way we solve a large number of real-world problems [158], [162]. Unlocking this promise amounts to adapting causality to the real world with its highly complex, unstructured, and abundant data. Beyond rich datasets, causal inference methods typically rely on diverse and extensive prior knowledge about the causal nature of the underlying system, many of which are not available or even testable in the application domain [83]. This has led to two consequences. Firstly, practitioners may shy away from adopting causal inference in general because the available causal knowledge is not sufficient, let alone complete. Secondly, the emphasis on causal knowledge may overshadow other important considerations such as the statistical, functional, or temporal properties. These consequences may lead the practitioner to under-utilize these methods which ultimately results in sub-optimal solutions.

We want to encourage researchers and practitioners to take causality to its next step: real-world impact. For this, we recognise the need to integrate causal inference with the sophisticated modelling capabilities of deep learning. This integration is particularly vital in real-world domains where understanding causal relationships can lead to better decision-making and predictions.

In this monograph, we introduce a new framework which enables pragmatically adopting causality ideas to solve real-world problems: Causal Deep Learning (CDL). Informally, CDL methods can leverage *partial causal knowledge* (among some and not necessarily all variables of interest), and *quantitatively characterises the functional form* (among variables of interest) and the *time-evolution of the variables* to provide significant insights to researchers and decision-makers. The reason why the above properties are important is two-fold: (i) We need a good way to match a model with prior knowledge in a complex system, this should allow information of any type (be it no information, partial information, or full information); and (ii) We need a good way to *evaluate* a solution. This latter point is important in any real-world setting, especially where we rely on modelling solutions to support impactful decisions. Inherently being built on a strong set of assumptions, model validation is a particularly tricky endeavour in causality and one which CDL can help us solve [160].

With new methods being proposed at various machine learning and statistics venues, tracking solved problems across research efforts is hard. Our CDL framework allows this necessary comparison across these fields. Allowing: (i) accurate comparison of existing methods, (ii) identifying gaps in contemporary research, ultimately driving research forward, (iii) communicate methods to a practical audience, encouraging the adoption of causal deep learning in practice. Throughout this text, we will use “Causal deep learning” and its acronym “CDL” to refer to methods that leverage causal ideas in their models as presented in this work.

Illustrative examples. Let us explore four diverse domains where causal deep learning, characterised by its focus on causal structure, functional relationships among variables of interest, and time, to understand the need for this new way of thinking.

First, consider the medical and healthcare domain, where determining the effectiveness of medications is paramount. Traditional models from pharmacology or physiology often struggle to capture the complexity of drug interactions and patient responses. To make progress in studying the effects of medications on health outcomes, CDL might focus on the causal relationship between the drug dosage and patient recovery rate while controlling for other variables like age, gender, and pre-existing conditions. In such a case, a CDL method would not assert complete causality of overall health factors but rather isolate the drug's impact. By employing only a partial causal structure, CDL can model the nuanced relationship between drug dosage and patient recovery, considering both linear and non-linear effects. Moreover, by incorporating temporal dynamics, CDL may unravel how medication effectiveness evolves in time, providing critical insights for personalized medicine. Naturally, without first identifying methods that allow partial causal knowledge as input, we may make mistakes and employ a method requiring full (and correct) causal knowledge. Using our framework may avoid such critical mistakes.

In economics and business, the relationship between interest rate changes and consumer spending is a classic example of a complex causal interaction. Traditional econometric models might fail to capture the intricacy of this relationship. CDL, with its ability to handle non-linear and high-dimensional data, can provide a more robust understanding. By adopting a pragmatic approach, CDL could enable the investigation of partial causality between interest rate changes and consumer spending. While acknowledging that other factors like education, employment rates, inflation, and economic policies also impact spending, the focus here is on understanding how variations in interest rates specifically influence consumer behaviour. By considering the temporal dimension, CDL may also uncover lag effects, where changes in interest rates take time to manifest in consumer behaviour, a crucial insight for policymakers.

Environmental science, especially the study of climate change, presents another compelling case for causal deep learning. An interesting research aspect might be to explore the causal relationship between carbon dioxide emissions and global temperature increase, acknowledging that other factors like deforestation and solar radiation also play roles in climate change. Using such a pragmatic CDL approach which requires only partial causal structure would allow researchers to isolate and understand the specific impact of CO2 emissions. In addition, the relationship between CO2 emissions and global temperature increase is not straightforward and likely non-linear. Here, causal deep learning framework can capture the complexity of this relationship beyond traditional linear models. Additionally, understanding how this relationship evolves over time is critical for predicting future climate patterns and for formulating effective environmental policies.

Lastly, in the education sector, the impact of classroom size on student performance is a topic of ongoing debate. While simple linear models might suggest a straightforward inverse relationship, the reality is likely more nuanced. Causal deep learning can help in identifying not just whether, but how class size impacts student performance, including potential threshold effects or non-linear dynamics. Furthermore, by examining how this relationship changes over an academic year or across different educational stages, deeper insights can be gained into effective educational planning and resource allocation. Solving this problem requires detailed descriptions of these dynamics are typically only possible with a set of fully known equations describing the data-generating process. These methods exist but often require a large set of modelling assumptions which are best charted with our framework.

In these and other real-world domains, causal deep learning can be used as a powerful framework to make progress. Unlike current causal models in machine learning, CDL allows for a nuanced understanding of causal relationships without requiring complete causal knowledge, learns in a data-driven manner the parametric form of the relationships among the variables of interest using powerful deep-learning models, and importantly, considers the dimension of time. This approach can uncover insights that traditional models might miss, leading to more effective interventions and policies across various sectors. The exploration of causal deep learning in these diverse real-world domains not only underscores its versatility but also highlights its potential to revolutionize how we understand and interact with the world around us.

With our detailed definitions, we can better align a problem with the (causal) solution. While research in causality has a long history of listing out the various assumptions necessary to identify causal effects, these assumptions often do not map easily into practice. Furthermore, deep learning has no such history, typically because most deep learning architectures are assumed to be non-parametric. However, not all problems require the enormous flexibility of deep neural networks, but may still be non-linear. Our proposal encompasses all the above in hopes of aiding the correct and useful adoption of causal deep learning methods.

Please refer to https://nowpublishers.com/article/Details/SIG-123 for the online version of this monograph.

References

[1] V. Aglietti, N. Dhir, J. González, and T. Damoulas, "Dynamic causal bayesian optimization," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, pp. 10 549–10 560, Curran Associates, Inc., 2021. URL: https://proceedings.neurips.cc/paper/2021/file/577bcc914f9e55d5e4e4f82f9f00e7d4-Paper.pdf.

[2] V. Aglietti, X. Lu, A. Paleyes, and J. González, "Causal bayesian optimization," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, S. Chiappa and R. Calandra, Eds., ser. Proceedings of Machine Learning Research, vol. 108, pp. 3155–3164, PMLR, 2020. URL: https://proceedings.mlr.press/v108/aglietti20a.html.

[3] C. Ahn *et al.*, "Global transplantation covid report march 2020," *Transplantation*, 2020.

[4] K. Ahuja, D. Mahajan, Y. Wang, and Y. Bengio, "Interventional causal representation learning," in *International conference on machine learning*, PMLR, pp. 372–407, 2023.

[5] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, 1974, pp. 716–723. DOI: 10.1109/TAC.1974.1100705.

[6] A. M. Alaa and M. van der Schaar, "Bayesian inference of individualized treatment effects using multi-task gaussian processes," *Advances in neural information processing systems*, vol. 30, 2017.

[7] A. M. Alshami, "Knee osteoarthritis related pain: A narrative review of diagnosis and treatment," *International journal of health sciences*, vol. 8, no. 1, 2014, pp. 85–104.

[8] A. Aussem, A. Tchernof, S. R. de Morais, and S. Rome, "Description of the recursive hybrid parents and children algorithm," 2010.

[9] J. Bannon, B. Windsor, W. Song, and T. Li, *Causality and batch reinforcement learning: Complementary approaches to planning in unknown domains*, 2020. DOI: 10.48550/ARXIV.2006.02579.

[10] E. Bareinboim, J. D. Correa, D. Ibeling, and T. Icard, "On pearl's hierarchy and the foundations of causal inference," in *Probabilistic and Causal Inference: The Works of Judea Pearl*, 2022, pp. 507–556.

[11] J. Basilico and Y. Raimond, "Déja vu: The importance of time and causality in recommender systems," in *Proceedings of the eleventh ACM conference on recommender systems*, pp. 342–342, 2017.

[12] A. Basu, "Estimating person-centered treatment (pet) effects using instrumental variables: An application to evaluating prostate cancer treatments," *Journal of Applied Econometrics*, vol. 29, no. 4, 2014, pp. 671–691.

[13] K. Bello, B. Aragam, and P. Ravikumar, "Dagma: Learning dags via m-matrices and a log-determinant acyclicity characterization," *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 8226–8239.

[14] A. Belyaeva, L. Cammarata, A. Radhakrishnan, C. Squires, K. D. Yang, G. Shivashankar, and C. Uhler, "Causal network models of sars-cov-2 expression and aging to identify candidates for drug repurposing," *Nature communications*, vol. 12, no. 1, 2021, p. 1024.

[15] J. Berrevoets, A. Alaa, Z. Qian, J. Jordon, A. E. S. Gimson, and M. van der Schaar, “Learning queueing policies for organ transplantation allocation using interpretable counterfactual survival analysis,” in *Proceedings of the 38th International Conference on Machine Learning*, M. Meila and T. Zhang, Eds., ser. Proceedings of Machine Learning Research, vol. 139, pp. 792–802, PMLR, 2021. URL: https://proceedings.mlr.press/v139/berrevoets21a.html.

[16] J. Berrevoets, A. Curth, I. Bica, E. McKinney, and M. van der Schaar, “Disentangled counterfactual recurrent networks for treatment effect inference over time,” *arXiv preprint arXiv:2112.03811*, 2021.

[17] J. Berrevoets, S. Holt, K. Kacprzyk, Z. Qian, and M. van der Schaar, “ODE discovery for longitudinal heterogeneous treatment effects inference,” in *The Twelfth International Conference on Learning Representations*, 2024. URL: https://openreview.net/forum?id=pxI5IPeWgW.

[18] J. Berrevoets, F. Imrie, T. Kyono, J. Jordon, and M. van der Schaar, “To impute or not to impute? missing data in treatment effect estimation,” in *International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 3568–3590, 2023.

[19] J. Berrevoets, D. Jarrett, A. Chan, and M. van der Schaar, “Allsim: Simulating and benchmarking resource allocation policies in multi-user systems,” *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[20] J. Berrevoets, J. Jordon, I. Bica, and M. van der Schaar, “Organite: Optimal transplant donor organ offering using an individual treatment effect,” *Advances in neural information processing systems*, vol. 33, 2020, pp. 20 037–20 050.

[21] J. Berrevoets, N. Seedat, F. Imrie, and M. van der Schaar, “Differentiable and transportable structure learning,” *arXiv preprint arXiv:2206.06354*, 2022.

[22] J. Berrevoets, S. Verboven, and W. Verbeke, “Treatment effect optimisation in dynamic environments,” *Journal of Causal Inference*, vol. 10, no. 1, 2022, pp. 106–122. DOI: doi:10.1515/jci-2020-0009.

[23] R. Bhattacharya, T. Nagarajan, D. Malinsky, and I. Shpitser, "Differentiable causal discovery under unmeasured confounding," in *International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 2314–2322, 2021.

[24] I. Bica, A. M. Alaa, J. Jordon, and M. van der Schaar, "Estimating counterfactual treatment outcomes over time through adversarially balanced representations," in *International Conference on Learning Representations*, 2020.

[25] I. Bica, J. Jordon, and M. van der Schaar, "Estimating the effects of continuous-valued interventions using generative adversarial networks," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33, pp. 16 434–16 445, Curran Associates, Inc., 2020. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/bea5955b308361a1b07bc55042e25e54-Paper.pdf.

[26] B. Bilodeau, L. Wang, and D. M. Roy, "Adaptively exploiting d-separators with causal bandits," 2022. DOI: 10.48550/ARXIV.2202.05100.

[27] L. E. Bothwell, J. Avorn, N. F. Khan, and A. S. Kesselheim, "Adaptive design clinical trials: A review of the literature and clinicaltrials. gov," *BMJ open*, vol. 8, no. 2, 2018, e018320.

[28] J. Brehmer, P. De Haan, P. Lippe, and T. S. Cohen, "Weakly supervised causal representation learning," *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 38 319–38 331.

[29] P. Brouillard, S. Lachapelle, A. Lacoste, S. Lacoste-Julien, and A. Drouin, "Differentiable causal discovery from interventional data," *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 21 865–21 877.

[30] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Sparse identification of nonlinear dynamics with control (sindyc)," *IFAC-PapersOnLine*, vol. 49, no. 18, 2016, pp. 710–715.

[31] P. Bühlmann, "Invariance, causality and robustness," *Statistical Science*, vol. 35, no. 3, 2020, pp. 404–426.

[32] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on scientific computing*, vol. 16, no. 5, 1995, pp. 1190–1208.

[33] B. Calderhead, M. Girolami, and N. Lawrence, "Accelerating bayesian inference over nonlinear differential equations with gaussian processes," *Advances in neural information processing systems*, vol. 21, 2008.

[34] G. M. Caporale and N. Pittis, "Causality and forecasting in incomplete systems," *Journal of Forecasting*, vol. 16, no. 6, 1997, pp. 425–437. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291099-131X%28199711%2916%3A6%3C425%3A%3AAID-FOR657%3E3.0.CO%3B2-9.

[35] B. D. Carter, C. C. Abnet, D. Feskanich, N. D. Freedman, P. Hartge, C. E. Lewis, J. K. Ockene, R. L. Prentice, F. E. Speizer, M. J. Thun, and E. J. Jacobs, "Smoking and mortality — beyond established causes," *New England Journal of Medicine*, vol. 372, no. 7, 2015, pp. 631–640. DOI: 10.1056/NEJMsa1407211.

[36] A. Chan, R. Peck, M. Gibbs, and M. van der Schaar, "Synthetic model combination: A new machine learning method for pharmacometric model ensembling," *CPT: pharmacometrics & systems pharmacology*, 2023.

[37] M.-L. Charpignon, B. Vakulenko-Lagun, B. Zheng, C. Magdamo, B. Su, K. Evans, S. Rodriguez, A. Sokolov, S. Boswell, Y.-H. Sheu, M. Somai, L. Middleton, B. T. Hyman, R. A. Betensky, S. N. Finkelstein, R. E. Welsch, I. Tzoulaki, D. Blacker, S. Das, and M. W. Albers, "Drug repurposing of metformin for alzheimer's disease: Combining causal inference in medical records data and systems pharmacology for biomarker identification," *medRxiv*, 2021, pp. 2021–08.

[38] M.-L. Charpignon, B. Vakulenko-Lagun, B. Zheng, C. Magdamo, B. Su, K. Evans, S. Rodriguez, A. Sokolov, S. Boswell, Y.-H. Sheu, M. Somai, L. Middleton, B. T. Hyman, R. A. Betensky, S. N. Finkelstein, R. E. Welsch, I. Tzoulaki, D. Blacker, S. Das, and M. W. Albers, "Causal inference in medical records and complementary systems pharmacology for metformin drug repurposing towards dementia," *Nature communications*, vol. 13, no. 1, 2022, p. 7652.

[39] Y. Chen, L. Wu, and M. Zaki, "Iterative deep graph learning for graph neural networks: Better and robust node embeddings," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33, pp. 19 314–19 326, Curran Associates, Inc., 2020. URL: https://proceedings.neurips.cc/paper/2020/file/e05c7ba4e087beea9410929698dc41a6-Paper.pdf.

[40] M. Cinquini, F. Giannotti, and R. Guidotti, "Boosting synthetic data generation with effective nonlinear causal discovery," in *2021 IEEE Third International Conference on Cognitive Machine Intelligence (CogMI)*, pp. 54–63, 2021. DOI: 10.1109/CogMI52975.2021.00016.

[41] A. Craft, "The first randomised controlled trial," *Archives of Disease in Childhood*, vol. 79, no. 5, 1998, pp. 410–410.

[42] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Transactions on pattern analysis and machine intelligence*, no. 1, 1983, pp. 25–39.

[43] A. Curth, A. Hüyük, and M. Van Der Schaar, "Adaptive identification of populations with treatment benefit in clinical trials: Machine learning challenges and solutions," in *Proceedings of the 40th International Conference on Machine Learning*, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., ser. Proceedings of Machine Learning Research, vol. 202, pp. 6603–6622, PMLR, 23–29 Jul 2023. URL: https://proceedings.mlr.press/v202/curth23a.html.

[44] A. Curth, D. Svensson, J. Weatherall, and M. van der Schaar, "Really doing great at estimating cate? a critical look at ml benchmarking practices in treatment effect estimation," in *Thirty-fifth conference on neural information processing systems datasets and benchmarks track (round 2)*, 2021.

[45] A. Curth and M. van der Schaar, "Nonparametric estimation of heterogeneous treatment effects: From theory to learning algorithms," in *International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 1810–1818, 2021.

[46] A. Curth and M. van der Schaar, "On inductive biases for heterogeneous treatment effect estimation," *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 15 883–15 894.

[47] I. Dasgupta, J. Wang, S. Chiappa, J. Mitrovic, P. Ortega, D. Raposo, E. Hughes, P. Battaglia, M. Botvinick, and Z. Kurth-Nelson, "Causal reasoning from meta-reinforcement learning," *arXiv preprint arXiv:1901.08162*, 2019.

[48] I. Dattner and C. A. Klaassen, "Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters," *Electronic Journal of Statistics*, vol. 9, no. 2, 2015, pp. 1939–1973.

[49] P. De Haan, D. Jayaraman, and S. Levine, "Causal confusion in imitation learning," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[50] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, pp. 248–255, 2009.

[51] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*, vol. 3. Wiley New York, 1973.

[52] F. Eberhardt, "Introduction to the foundations of causal discovery," *International Journal of Data Science and Analytics*, vol. 3, no. 2, 2017, pp. 81–91.

[53] M. Eichler, R. Dahlhaus, and J. Dueck, "Graphical modeling for multivariate hawkes processes with nonparametric link functions," *Journal of Time Series Analysis*, vol. 38, no. 2, 2017, pp. 225–242.

[54] E. J. Emanuel, G. Persad, R. Upshur, B. Thome, M. Parker, A. Glickman, C. Zhang, C. Boyle, M. Smith, and J. P. Phillips, *Fair allocation of scarce medical resources in the time of covid-19*, 2020.

[55] A. Faragó and G. Lugosi, "Strong universal consistency of neural network classifiers," *IEEE Transactions on Information Theory*, vol. 39, no. 4, 1993, pp. 1146–1151.

[56] B. Fatemi, L. El Asri, and S. M. Kazemi, "Slaps: Self-supervision improves structure learning for graph neural networks," *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 22 667–22 681.

[57] J. R. Freeman, "Granger causality and the times series analysis of political relationships," *American Journal of Political Science*, 1983, pp. 327–358.

[58] M. Gasse, D. Grasset, G. Gaudron, and P.-Y. Oudeyer, "Causal reinforcement learning using observational and interventional data," *arXiv preprint arXiv:2106.14421*, 2021.

[59] D. Geiger and D. Heckerman, "Learning gaussian networks," in *Uncertainty Proceedings 1994*, Elsevier, 1994, pp. 235–243.

[60] D. Geiger and J. Pearl, "On the logic of causal models," in *Machine Intelligence and Pattern Recognition*, vol. 9, Elsevier, 1990, pp. 3–14.

[61] A. Ghazimatin, O. Balalau, R. Saha Roy, and G. Weikum, "Prince: Provider-side interpretability with counterfactual explanations in recommender systems," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, pp. 196–204, 2020.

[62] C. Glymour, K. Zhang, and P. Spirtes, "Review of causal discovery methods based on graphical models," *Frontiers in genetics*, vol. 10, 2019, p. 524.

[63] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[64] O. Goudet, D. Kalainathan, P. Caillou, I. Guyon, D. Lopez-Paz, and M. Sebag, "Causal generative neural networks," *arXiv preprint arXiv:1711.08936*, 2017.

[65] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: journal of the Econometric Society*, 1969, pp. 424–438.

[66] N. Hassanpour and R. Greiner, "Counterfactual regression with importance sampling weights.," in *IJCAI*, pp. 5880–5887, 2019.

[67] A. Hauser 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*, vol. 77, no. 1, 2015, pp. 291–318.

[68] T. He, J. Gajcin, and I. Dusparic, "Causal counterfactuals for improving the robustness of reinforcement learning," *arXiv preprint arXiv:2211.05551*, 2022.

[69] D. Heckerman, "A tutorial on learning with bayesian networks," *Innovations in Bayesian networks: Theory and applications*, 2008, pp. 33–82.

[70] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning bayesian networks: The combination of knowledge and statistical data," *Machine learning*, vol. 20, 1995, pp. 197–243.

[71] D. Heckerman, C. Meek, and G. Cooper, "A bayesian approach to causal discovery," in *Innovations in Machine Learning*, Springer, 2006, pp. 1–28.

[72] T. Helliwell, C. Mallen, G. Peat, and E. Hay, "Research into practice: Improving musculoskeletal care in general practice," *British Journal of General Practice*, vol. 64, no. 624, 2014, pp. 372–374. DOI: 10.3399/bjgp14X680749.

[73] R. Helwegen, C. Louizos, and P. Forré, "Improving fair predictions using variational inference in causal models," *arXiv preprint arXiv:2008.10880*, 2020.

[74] T. Herlau and R. Larsen, "Reinforcement learning of causal variables using mediation analysis," in *36th AAAI Conference on Artificial Intelligence*, Association for the Advancement of Artificial Intelligence, pp. 6910–6917, 2022.

[75] J. L. Hill, "Bayesian nonparametric modeling for causal inference," *Journal of Computational and Graphical Statistics*, vol. 20, no. 1, 2011, pp. 217–240.

[76] P. W. Holland, "Statistics and causal inference," *Journal of the American statistical Association*, vol. 81, no. 396, 1986, pp. 945–960.

[77] N. Hollmann, S. Müller, K. Eggensperger, and F. Hutter, "TabPFN: A transformer that solves small tabular classification problems in a second," *arXiv preprint arXiv:2207.01848*, 2022.

[78] P. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf, "Nonlinear causal discovery with additive noise models," *Advances in neural information processing systems*, vol. 21, 2008.

[79] N. Huynh, J. Berrevoets, N. Seedat, J. Crabbé, Z. Qian, and M. van der Schaar, "DAGnosis: Localized identification of data inconsistencies using structures.," in *International Conference on Artificial Intelligence and Statistics*, PMLR, 2024.

[80] A. Hyvärinen, "Independent component analysis: Recent advances," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, 2013, p. 20 110 534.

[81] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural networks*, vol. 13, no. 4-5, 2000, pp. 411–430.

[82] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: Existence and uniqueness results," *Neural networks*, vol. 12, no. 3, 1999, pp. 429–439.

[83] G. W. Imbens, "Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics," *Journal of Economic Literature*, vol. 58, no. 4, 2020, pp. 1129–79.

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

[85] A. Jain, H. Patel, L. Nagalapatti, N. Gupta, S. Mehta, S. Guttula, S. Mujumdar, S. Afzal, R. Sharma Mittal, and V. Munigala, "Overview and importance of data quality for machine learning tasks," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 3561–3562, 2020.

[86] C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics letters*, vol. 6, no. 3, 1980, pp. 255–259.

[87] L. Jiang, Z. Cai, D. Wang, and H. Zhang, "Improving tree augmented naive bayes for class probability estimation," *Knowledge-Based Systems*, vol. 26, 2012, pp. 239–245.

[88] L. Jiang, C. Li, S. Wang, and L. Zhang, "Deep feature weighting for naive bayes and its application to text classification," *Engineering Applications of Artificial Intelligence*, vol. 52, 2016, pp. 26–39.

[89] L. Jiang, H. Zhang, and Z. Cai, "A novel bayes model: Hidden naive bayes," *IEEE Transactions on knowledge and data engineering*, vol. 21, no. 10, 2008, pp. 1361–1371.

[90] M. Johnson, M. Patel, A. Phipps, M. Van der Schaar, D. Boulton, and M. Gibbs, "The potential and pitfalls of artificial intelligence in clinical pharmacology," *CPT: Pharmacometrics & Systems Pharmacology*, vol. 12, no. 3, 2023, pp. 279–284.

[91] P. S. Kamath and R. W. Kim, "The model for end-stage liver disease (meld)," *Hepatology*, vol. 45, no. 3, 2007, pp. 797–805.

[92] Y. Kan-Tor, N. Srebnik, M. Gavish, U. Shalit, and A. Buxboim, "Evaluating the heterogeneous effect of extended incubation to blastocyst transfer on the implantation outcome via causal inference," *bioRxiv*, 2021, pp. 2021–11.

[93] S. S. Kancheti, A. G. Reddy, V. N. Balasubramanian, and A. Sharma, "Matching learned causal effects of neural networks with domain priors," in *Proceedings of the 39th International Conference on Machine Learning*, K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, Eds., ser. Proceedings of Machine Learning Research, vol. 162, pp. 10 676–10 696, PMLR, 2022. URL: https://proceedings.mlr.press/v162/kancheti22a.html.

[94] C. Kim and O. Bastani, "Learning interpretable models with causal guarantees," *arXiv preprint arXiv:1901.08576*, 2019.

[95] M. Kocaoglu, C. Snyder, A. G. Dimakis, and S. Vishwanath, "CausalGAN: Learning causal implicit generative models with adversarial training," in *International Conference on Learning Representations*, 2018. URL: https://openreview.net/forum?id=BJE-4xW0W.

[96] R. Kohavi, "Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid," in *Kdd*, vol. 96, pp. 202–207, 1996.

[97] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[98] A. D. Kroon, J. Mooij, and D. Belgrave, "Causal bandits without prior knowledge using separating sets," in *Proceedings of the First Conference on Causal Learning and Reasoning*, B. Schölkopf, C. Uhler, and K. Zhang, Eds., ser. Proceedings of Machine Learning Research, vol. 177, pp. 407–427, PMLR, 2022. URL: https://proceedings.mlr.press/v177/kroon22a.html.

[99] A. Kunft, A. Katsifodimos, S. Schelter, S. Breß, T. Rabl, and V. Markl, "An intermediate representation for optimizing machine learning pipelines," *Proceedings of the VLDB Endowment*, vol. 12, no. 11, 2019, pp. 1553–1567.

[100] S. R. Künzel, J. S. Sekhon, P. J. Bickel, and B. Yu, "Metalearners for estimating heterogeneous treatment effects using machine learning," *Proceedings of the national academy of sciences*, vol. 116, no. 10, 2019, pp. 4156–4165.

[101] M. J. Kusner, J. Loftus, C. Russell, and R. Silva, "Counterfactual fairness," *Advances in neural information processing systems*, vol. 30, 2017.

[102] T. Kyono, I. Bica, Z. Qian, and M. van der Schaar, "Selecting treatment effects models for domain adaptation using causal knowledge," *arXiv preprint arXiv:2102.06271*, 2021.

[103] T. Kyono and M. van der Schaar, "Improving model robustness using causal knowledge," *arXiv preprint arXiv:1911.12441*, 2019.

[104] T. Kyono and M. van der Schaar, "Exploiting causal structure for robust model selection in unsupervised domain adaptation," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 6, 2021, pp. 494–507.

[105] T. Kyono, Y. Zhang, A. Bellot, and M. van der Schaar, "Miracle: Causally-aware imputation via learning missing data mechanisms," *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 23 806–23 817.

[106] T. Kyono, Y. Zhang, and M. van der Schaar, "Castle: Regularization via auxiliary causal graph discovery," *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 1501–1512.

[107] P. Langley, W. Iba, and K. Thompson, "An analysis of bayesian classifiers," in *Aaai*, vol. 90, pp. 223–228, 1992.

[108] F. Lattimore, T. Lattimore, and M. D. Reid, "Causal bandits: Learning good interventions via causal inference," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29, Curran Associates, Inc., 2016. URL: https://proceedings.neurips.cc/paper/2016/file/b4288d9c0ec0a1841b3b3728321e7088-Paper.pdf.

[109] F. Lattimore, T. Lattimore, and M. D. Reid, "Causal bandits: Learning good interventions via causal inference," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[110] S. Lee and E. Bareinboim, "Structural causal bandits: Where to intervene?" In *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31, Curran Associates, Inc., 2018. URL: https://proceedings.neurips.cc/paper/2018/file/c0a271bc0ecb776a094786474322cb82-Paper.pdf.

[111] S. Lee and E. Bareinboim, "Structural causal bandits with non-manipulable variables," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 4164–4172. DOI: 10.1609/aaai.v33i01.33014164.

[112] S. Lee and E. Bareinboim, "Characterizing optimal mixed policies: Where to intervene and what to observe," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33, pp. 8565–8576, Curran Associates, Inc., 2020. URL: https://proceedings.neurips.cc/paper/2020/file/61a10e6abb1149ad9d08f303267f9bc4-Paper.pdf.

[113] D. Li, J. Sun, H. Yang, and X. Wang, "An enhanced naive bayes model for dissolved oxygen forecasting in shellfish aquaculture," *IEEE Access*, vol. 8, 2020, pp. 217 917–217 927. DOI: 10.1109/ACCESS.2020.3042180.

[114] W. Li, Y. Li, S. Zhu, Y. Shao, J. Hao, and Y. Pang, "Gflowcausal: Generative flow networks for causal discovery," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2022. URL: arXiv%20preprint%20arXiv:2210.08185.

[115] B. Lim, "Forecasting treatment responses over time using recurrent marginal structural networks," *Advances in neural information processing systems*, vol. 31, 2018.

[116] P. Lippe, S. Magliacane, S. Löwe, Y. M. Asano, T. Cohen, and E. Gavves, "Causal representation learning for instantaneous and temporal effects in interactive systems," in *The Eleventh International Conference on Learning Representations*, 2022.

[117] R. Liu, L. Wei, and P. Zhang, "When deep learning meets causal inference: A computational framework for drug repurposing from real-world data," *arXiv preprint arXiv:2007.10152*, 2020.

[118] R. Liu, L. Wei, and P. Zhang, "A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data," *Nature machine intelligence*, vol. 3, no. 1, 2021, pp. 68–75.

[119] T. Liu, Z. Qian, J. Berrevoets, and M. van der Schaar, "GOGGLE: Generative modelling for tabular data by learning relational structure," in *International Conference on Learning Representations*, 2023. URL: https://openreview.net/forum?id=fPVRcJqspu.

[120] Y. Liu, R. Cadei, J. Schweizer, S. Bahmani, and A. Alahi, "Towards robust and adaptive motion forecasting: A causal representation perspective," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17 081–17 092, 2022.

[121] J. R. Loftus, C. Russell, M. J. Kusner, and R. Silva, "Causal reasoning for algorithmic fairness," *arXiv preprint arXiv:1805.05859*, 2018.

[122] C. Lu, J. M. Hernández-Lobato, and B. Schölkopf, "Invariant causal representation learning for generalization in imitation and reinforcement learning," in *ICLR2022 Workshop on the Elements of Reasoning: Objects, Structure and Causality*, 2022.

[123] Y. Lu, A. Meisami, and A. Tewari, "Causal bandits with unknown graph structure," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, pp. 24 817–24 828, Curran Associates, Inc., 2021. URL: https://proceedings.neurips.cc/paper/2021/file/d010396ca8abf6ead8cacc2c2f2f26c7-Paper.pdf.

[124] Y. Lu, A. Meisami, and A. Tewari, "Efficient reinforcement learning with prior causal knowledge," in *Proceedings of the First Conference on Causal Learning and Reasoning*, B. Schölkopf, C. Uhler, and K. Zhang, Eds., ser. Proceedings of Machine Learning Research, vol. 177, pp. 526–541, PMLR, 2022. URL: https://proceedings.mlr.press/v177/lu22a.html.

[125] Y. Lu, A. Meisami, A. Tewari, and W. Yan, "Regret analysis of bandit problems with causal background knowledge," in *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, J. Peters and D. Sontag, Eds., ser. Proceedings of Machine Learning Research, vol. 124, pp. 141–150, PMLR, 2020. URL: https://proceedings.mlr.press/v124/lu20a.html.

[126] C. Lyle, A. Zhang, M. Jiang, J. Pineau, and Y. Gal, "Resolving causal confusion in reinforcement learning via robust exploration," in *Self-Supervision for Reinforcement Learning Workshop-ICLR 2021*, 2021.

[127] B. Macdonald and D. Husmeier, "Gradient matching methods for computational inference in mechanistic models for systems biology: A review and comparative analysis," *Frontiers in Bioengineering and Biotechnology*, vol. 3, 2015. DOI: 10.3389/fbioe.2015.00180.

[128] S. Magliacane, T. van Ommen, T. Claassen, S. Bongers, P. Versteeg, and J. M. Mooij, "Causal transfer learning," *arXiv preprint arXiv:1707.06422*, 2017.

[129] A. Maiti, V. Nair, and G. Sinha, "A causal bandit approach to learning good atomic interventions in presence of unobserved confounders," in *The 38th Conference on Uncertainty in Artificial Intelligence*, 2022.

[130] C. Mao, A. Cha, A. Gupta, H. Wang, J. Yang, and C. Vondrick, "Generative interventions for causal learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3947–3956, 2021.

[131] I. Mayer, E. Sverdrup, T. Gauss, J.-D. Moyer, S. Wager, and J. Josse, "Doubly robust treatment effect estimation with missing attributes," 2020.

[132] J. Medlock and A. P. Galvani, "Optimizing influenza vaccine distribution," *Science*, vol. 325, no. 5948, 2009, pp. 1705–1708.

[133] C. Meek, *Complete orientation rules for patterns.* Carnegie Mellon [Department of Philosophy], 1995.

[134] C. Meek, "Strong completeness and faithfulness in bayesian networks," *arXiv preprint arXiv:1302.4973*, 2013.

[135] A. Méndez-Molina, E. F.Morales, and L. E. Sucar, "Causal discovery and reinforcement learning: A synergistic integration," in *Proceedings of The 11th International Conference on Probabilistic Graphical Models*, A. Salmerón and R. Rumi, Eds., ser. Proceedings of Machine Learning Research, vol. 186, pp. 421–432, PMLR, 2022. URL: https://proceedings.mlr.press/v186/mendez-molina22a.html.

[136] F. V. Mikkelsen and N. R. Hansen, "Learning large scale ordinary differential equation systems," *arXiv preprint arXiv:1710.09308*, 2017.

[137] J. Mitrovic, B. McWilliams, J. Walker, L. Buesing, and C. Blundell, "Representation learning via invariant causal mechanisms," *arXiv preprint arXiv:2010.07922*, 2020.

[138] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, 2015, pp. 529–533.

[139] K. Mohan and J. Pearl, "Graphical models for processing missing data," *Journal of the American Statistical Association*, vol. 116, no. 534, 2021, pp. 1023–1037.

[140] K. Mohan, J. Pearl, and J. Tian, "Graphical models for inference with missing data," *Advances in neural information processing systems*, vol. 26, 2013.

[141] R. Moraffah, M. Karami, R. Guo, A. Raglin, and H. Liu, "Causal interpretability for machine learning-problems, methods and evaluation," *ACM SIGKDD Explorations Newsletter*, vol. 22, no. 1, 2020, pp. 18–33.

[142] P. Morales-Alvarez, W. Gong, A. Lamb, S. Woodhead, S. P. Jones, N. Pawlowski, M. Allamanis, and C. Zhang, "Simultaneous missing value imputation and structure learning with groups," in *Advances in Neural Information Processing Systems*, A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, Eds., 2022. URL: https://openreview.net/forum?id=4rm6tzBjChe.

[143] P. Morales-Alvarez, A. Lamb, S. Woodhead, S. P. Jones, M. Allamanis, and C. Zhang, "Vicause: Simultaneous missing value imputation and causal discovery with groups," *arXiv preprint arXiv:2110.08223*, 2021.

[144] Y. Mounk, "The extraordinary decisions facing italian doctors," *The Atlantic*, vol. 11, 2020, p. 2020.

[145] J. Müller, R. Schmier, L. Ardizzone, C. Rother, and U. Köthe, "Learning robust models using the principle of independent causal mechanisms," in *DAGM German Conference on Pattern Recognition*, Springer, pp. 79–110, 2021.

[146] K. P. Murphy, *Machine learning: a probabilistic perspective.* MIT press, 2012.

[147] M. Mutti, R. D. Santi, E. Rossi, J. F. Calderon, M. M. Bronstein, and M. Restelli, "Provably efficient causal model-based reinforcement learning for environment-agnostic generalization," in *A causal view on dynamical systems, NeurIPS 2022 workshop*, 2022. URL: https://openreview.net/forum?id=VDunku3bDk.

[148] M. Naaman, "On the tight constant in the multivariate Dvoretzky-Kiefer-Wolfowitz inequality," *Statistics & Probability Letters*, vol. 173, 2021, p. 109 088.

[149] V. Nair, V. Patil, and G. Sinha, "Budgeted and non-budgeted causal bandits," in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, A. Banerjee and K. Fukumizu, Eds., ser. Proceedings of Machine Learning Research, vol. 130, pp. 2017–2025, PMLR, 2021. URL: https://proceedings.mlr.press/v130/nair21a.html.

[150] A. Nazaret and G. Sapiro, "A large-scale observational study of the causal effects of a behavioral health nudge," in *NeurIPS 2022 Workshop on Causality for Real-world Impact*, 2022. URL: https://openreview.net/forum?id=4NocV3GQRG9.

[151] J. Neyman, "Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes," *Roczniki Nauk Rolniczych*, vol. 10, 1923, pp. 1–51.

[152] A. Ng, L. Aroyo, C. Coleman, G. Diamos, V. J. Reddi, J. Vanschoren, C.-J. Wu, and S. Zhou, *NeurIPS workshop on data-centric ai*, virtual, 2021. URL: https://datacentricai.org/neurips21.

[153] I. Ng, S. Lachapelle, N. R. Ke, S. Lacoste-Julien, and K. Zhang, "On the convergence of continuous constrained optimization for structure learning," in *International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 8176–8198, 2022.

[154] F. Noor, M. Asif, U. A. Ashfaq, M. Qasim, and M. Tahir ul Qamar, "Machine learning for synergistic network pharmacology: A comprehensive overview," *Briefings in Bioinformatics*, 2023, bbad120.

[155] C. J. Oates, F. Dondelinger, N. Bayani, J. Korkola, J. W. Gray, and S. Mukherjee, "Causal network inference using biochemical kinetics," *Bioinformatics*, vol. 30, no. 17, Aug. 2014, pp. i468–i474. DOI: 10.1093/bioinformatics/btu452.

[156] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, 1954, pp. 100–115.

[157] R. Pamfil, N. Sriwattanaworachai, S. Desai, P. Pilgerstorfer, K. Georgatzis, P. Beaumont, and B. Aragam, "Dynotears: Structure learning from time-series data," in *International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 1595–1605, 2020.

[158] N. Pawlowski, J. Berrevoets, C. Uhler, K. Zhang, M. van der Schaar, and C. Zhang, *NeurIPS workshop on causal machine learning for real-world impact (CML4Impact)*, New Orleans, LA, USA, 2022. URL: https://www.cml-4-impact.vanderschaar-lab.com/.

[159] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, 1988.

[160] J. Pearl, "Bayesianism and causality, or, why i am only a half-bayesian," in *Foundations of bayesianism*, Springer, 2001, pp. 19–36.

[161] J. Pearl, *Causality*. Cambridge university press, 2009.

[162] J. Pearl and D. Mackenzie, *The book of why: the new science of cause and effect*. Basic books, 2018.

[163] J. Pearl and T. S. Verma, "A theory of inferred causation," in *Studies in Logic and the Foundations of Mathematics*, vol. 134, Elsevier, 1995, pp. 789–811.

[164] J. R. Peters, *Machine learning of motor skills for robotics*. University of Southern California, 2007.

[165] J. Peters, S. Bauer, and N. Pfister, "Causal models for dynamical systems," in *Probabilistic and Causal Inference: The Works of Judea Pearl*, 2022, pp. 671–690.

[166] J. Peters, D. Janzing, and B. Schölkopf, *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

[167] J. Peters, J. M. Mooij, D. Janzing, and B. Schölkopf, "Causal discovery with continuous additive noise models," *Journal of Machine Learning Research*, vol. 15, no. 58, 2014, pp. 2009–2053. URL: http://jmlr.org/papers/v15/peters14a.html.

[168] N. Pfister, S. Bauer, and J. Peters, "Identifying causal structure in large-scale kinetic systems," *arXiv*, 2018, pp. 1810–11 776.

[169] N. Pfister, P. Bühlmann, and J. Peters, "Invariant causal prediction for sequential data," *Journal of the American Statistical Association*, vol. 114, no. 527, 2019, pp. 1264–1276.

[170] Z. Qian, W. Zame, L. Fleuren, P. Elbers, and M. van der Schaar, "Integrating expert odes into neural odes: Pharmacology and disease progression," *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 11 364–11 383.

[171] C. M. Queen and C. J. Albers, "Intervention and causality: Forecasting traffic flows using a dynamic bayesian network," *Journal of the American Statistical Association*, vol. 104, no. 486, 2009, pp. 669–681. DOI: 10.1198/jasa.2009.0042.

[172] J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao, "Parameter estimation for differential equations: A generalized smoothing approach," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 5, 2007, pp. 741–796.

[173] A. Reisach, C. Seiler, and S. Weichwald, "Beware of the simulated dag! causal discovery benchmarks may be easy to game," *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 27 772–27 784.

[174] T. S. Richardson and J. M. Robins, "Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality," *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, vol. 128, no. 30, 2013, p. 2013.

[175] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, 1978, pp. 465–471.

[176] J. M. Robins, M. A. Hernan, and B. Brumback, "Marginal structural models and causal inference in epidemiology," *Epidemiology*, 2000, pp. 550–560.

[177] M. Rojas-Carulla, B. Schölkopf, R. Turner, and J. Peters, "Invariant models for causal transfer learning," *The Journal of Machine Learning Research*, vol. 19, no. 1, 2018, pp. 1309–1342.

[178] B. P. Roßbach, M. F. Pietschmann, M. F. Gülecyüz, T. R. Niethammer, A. Ficklscherer, S. Wild, V. Jansson, and P. E. Müller, "Indications requiring preoperative magnetic resonance imaging before knee arthroscopy," *Archives of medical science : AMS*, vol. 10, no. 6, 2014, pp. 1147–1152. DOI: 10.5114/aoms.2014.47825.

[179] D. B. Rubin, "Comment: Neyman (1923) and causal inference in experiments and observational studies," *Statistical Science*, vol. 5, no. 4, 1990, pp. 472–480.

[180] D. B. Rubin, "Comment on "randomization analysis of experimental data: The fisher randomization test"," *Journal of the American Statistical Association*, vol. 75, no. 371, 1980, p. 591. DOI: 10.2307/2287653.

[181] F. Russo and F. Toni, "Causal discovery and injection for feed-forward neural networks," *arXiv preprint arXiv:2205.09787*, 2022.

[182] K. Sadeghi, "Faithfulness of probability distributions and graphs," *Journal of Machine Learning Research*, vol. 18, no. 148, 2017, pp. 1–29.

[183] N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo, ""everyone wants to do the model work, not the data work": Data cascades in high-stakes ai," in *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2021.

[184] A. Sauer and A. Geiger, "Counterfactual generative networks," in *International Conference on Learning Representations*, 2021. URL: https://openreview.net/forum?id=BXewfAYMmJw.

[185] A. W. Sauter, E. Acar, and V. Francois-Lavet, "A meta-reinforcement learning algorithm for causal discovery," in *UAI 2022 Workshop on Causal Representation Learning*, 2022. URL: https://openreview.net/forum?id=2Y2HwBqORqz.

[186] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical chemistry*, vol. 36, no. 8, 1964, pp. 1627–1639.

[187] J. Schmidt-Hieber, "Nonparametric regression using deep neural networks with relu activation function," *The Annals of Statistics*, vol. 48, no. 4, 2020, pp. 1875–1897.

[188] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio, "Toward causal representation learning," *Proceedings of the IEEE*, vol. 109, no. 5, 2021, pp. 612–634.

[189] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics*, 1978, pp. 461–464.

[190] M. Scutari, "Dirichlet bayesian network scores and the maximum entropy principle," in *Advanced Methodologies for Bayesian Networks*, PMLR, pp. 8–20, 2017.

[191] M. Scutari and K. Strimmer, "Introduction to graphical modelling," *arXiv preprint arXiv:1005.1036*, 2010.

[192] N. Seedat, J. Crabbé, and M. van der Schaar, "Data-SUITE: Data-centric identification of in-distribution incongruous examples," in *Proceedings of the 39th International Conference on Machine Learning*, K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, Eds., ser. Proceedings of Machine Learning Research, vol. 162, pp. 19 467–19 496, PMLR, 2022. URL: https://proceedings.mlr.press/v162/seedat22a.html.

[193] N. Seedat, F. Imrie, A. Bellot, Z. Qian, and M. van der Schaar, "Continuous-time modeling of counterfactual outcomes using neural controlled differential equations," *arXiv preprint arXiv:2206.08311*, 2022.

[194] M. Seitzer, B. Schölkopf, and G. Martius, "Causal influence detection for improving efficiency in reinforcement learning," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, pp. 22 905–22 918, Curran Associates, Inc., 2021. URL: https://proceedings.neurips.cc/paper/2021/file/c1722a7941d61aad6e651a35b65a9c3e-Paper.pdf.

[195] R. Sen, K. Shanmugam, M. Kocaoglu, A. Dimakis, and S. Shakkottai, "Contextual Bandits with Latent Confounders: An NMF Approach," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, A. Singh and J. Zhu, Eds., ser. Proceedings of Machine Learning Research, vol. 54, pp. 518–527, PMLR, 2017. URL: https://proceedings.mlr.press/v54/sen17a.html.

[196] N. V. Shah, M. Solow, J. J. Kelly, A. Aylyarov, J. P. Doran, L. R. Bloom, S. Akil, B. Siddiqui, J. M. Newman, D. Chatterjee, N. Pancholi, A. Dixit, B. Kavousi, S. E. Barbash, W. P. Urban, and D. T. Neuman, "Demographics and rates of surgical arthroscopy and postoperative rehabilitative preferences of arthroscopists from the arthroscopy association of north america (aana)," eng, *Journal of orthopaedics*, vol. 15, no. 2, 2018, pp. 591–595. DOI: 10.1016/j.jor.2018.05.033.

[197] U. Shalit, F. D. Johansson, and D. Sontag, "Estimating individual treatment effect: Generalization bounds and algorithms," in *International Conference on Machine Learning*, PMLR, pp. 3076–3085, 2017.

[198] A. Sharma, J. M. Hofman, and D. J. Watts, "Estimating the causal impact of recommendation systems from observational data," in *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pp. 453–470, 2015.

[199] K. Sharma, Y. K. Dwivedi, and B. Metri, "Incorporating causality in energy consumption forecasting using deep neural networks," *Annals of Operations Research*, 2022, pp. 1–36.

[200] N. Sharma, S. Basu, K. Shanmugam, and S. Shakkottai, "Warm starting bandits with side information from confounded data," *arXiv preprint arXiv:2002.08405*, 2020.

[201] M. Sheikh, A. Mukeriya, O. Shangina, P. Brennan, and D. Zaridze, "Postdiagnosis smoking cessation and reduced risk for lung cancer progression and mortality: A prospective cohort study," *Annals of internal medicine*, vol. 174, no. 9, 2021, pp. 1232–1239.

[202] X. Shen, F. Liu, H. Dong, Q. Lian, Z. Chen, and T. Zhang, "Weakly supervised disentangled generative causal representation learning," *Journal of Machine Learning Research*, vol. 23, 2022, pp. 1–55.

[203] C. Shi, X. Wang, S. Luo, H. Zhu, J. Ye, and R. Song, "Dynamic causal effects evaluation in a/b testing with a reinforcement learning framework," *Journal of the American Statistical Association*, 2022, pp. 1–13.

[204] C. Shi, V. Veitch, and D. M. Blei, "Invariant representation learning for treatment effect estimation," in *Uncertainty in Artificial Intelligence*, PMLR, pp. 1546–1555, 2021.

[205] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvarinen, Y. Kawahara, T. Washio, P. O. Hoyer, K. Bollen, and P. Hoyer, "Directlingam: A direct method for learning a linear non-gaussian structural equation model," *Journal of Machine Learning Research-JMLR*, vol. 12, no. Apr, 2011, pp. 1225–1248.

[206] J. Smith, M. Lipsitch, and J. W. Almond, "Vaccine production, distribution, access, and uptake," *The Lancet*, vol. 378, no. 9789, 2011, pp. 428–438.

[207] P. Spirtes and C. Glymour, "An algorithm for fast recovery of sparse causal graphs," *Social science computer review*, vol. 9, no. 1, 1991, pp. 62–72.

[208] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman, *Causation, prediction, and search*. MIT press, 2000.

[209] C. Squires, D. Shen, A. Agarwal, D. Shah, and C. Uhler, "Causal imputation via synthetic interventions," in *Conference on Causal Learning and Reasoning*, PMLR, pp. 688–711, 2022.

[210] K. Stankevičiūtė, J.-B. Woillard, R. W. Peck, P. Marquet, and M. van der Schaar, "Bridging the worlds of pharmacometrics and machine learning," *Clinical Pharmacokinetics*, vol. 62, no. 11, 2023, pp. 1551–1565.

[211] C. Subramanian and B. Ravindran, "Causal contextual bandits with targeted interventions," in *International Conference on Learning Representations*, 2022. URL: https://openreview.net/forum?id=F5Em8ASCosV.

[212] X. Sun, D. Janzing, and B. Schölkopf, "Causal inference by choosing graphs with most plausible markov kernels," in *Ninth International Symposium on Artificial Intelligence and Mathematics (AIMath 2006)*, pp. 1–11, 2006.

[213] S. Tabib and D. Larocque, "Non-parametric individual treatment effect estimation for survival data with random forests," *Bioinformatics*, vol. 36, no. 2, 2020, pp. 629–636.

[214] G. Tennenholtz, U. Shalit, S. Mannor, and Y. Efroni, "Bandits with partially observable confounded data," in *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, C. de Campos and M. H. Maathuis, Eds., ser. Proceedings of Machine Learning Research, vol. 161, pp. 430–439, PMLR, 2021. URL: https://proceedings.mlr.press/v161/tennenholtz21a.html.

[215] T. Teshima and M. Sugiyama, "Incorporating causal graphical prior knowledge into predictive modeling via simple data augmentation," in *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, C. de Campos and M. H. Maathuis, Eds., ser. Proceedings of Machine Learning Research, vol. 161, pp. 86–96, PMLR, 2021. URL: https://proceedings.mlr.press/v161/teshima21a.html.

[216] I. Tsamardinos, C. F. Aliferis, A. R. Statnikov, and E. Statnikov, "Algorithms for large scale markov blanket discovery.," in *FLAIRS*, vol. 2, pp. 376–81, 2003.

[217] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, "The max-min hill-climbing bayesian network structure learning algorithm," *Machine learning*, vol. 65, 2006, pp. 31–78.

[218] C. A. Umscheid, D. J. Margolis, and C. E. Grossman, "Key concepts of clinical trials: A narrative review," *Postgraduate medicine*, vol. 123, no. 5, 2011, pp. 194–204.

[219] W. A. van Amsterdam, J. J. Verhoeff, N. I. Harlianto, G. A. Bartholomeus, A. M. Puli, P. A. de Jong, T. Leiner, A. S. van Lindert, M. J. Eijkemans, and R. Ranganath, "Individual treatment effect estimation in the presence of unobserved confounding using proxies: A cohort study in stage iii non-small cell lung cancer," *Scientific reports*, vol. 12, no. 1, 2022, p. 5848.

[220] B. van Breugel, T. Kyono, J. Berrevoets, and M. van der Schaar, "Decaf: Generating fair synthetic data using causally-aware generative networks," *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 22 221–22 233.

[221] T. Vanderschueren, J. Berrevoets, and W. Verbeke, "NOFLITE: Learning to predict individual treatment effect distributions," *Transactions on Machine Learning Research*, 2023. URL: https://openreview.net/forum?id=EjqopDxLbG.

[222] T. Vanderschueren, A. Curth, W. Verbeke, and M. van der Schaar, "Accounting for informative sampling when learning to forecast treatment outcomes over time," in *Proceedings of the 40th International Conference on Machine Learning*, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., ser. Proceedings of Machine Learning Research, vol. 202, pp. 34 855–34 874, PMLR, 2023. URL: https://proceedings.mlr.press/v202/vanderschueren23a.html.

[223] V. N. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," in *Measures of complexity: festschrift for alexey chervonenkis*, Springer, 2015, pp. 11–30.

[224] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[225] M. Vergano, G. Bertolini, A. Giannini, G. R. Gristina, S. Livigni, G. Mistraletti, L. Riccioni, and F. Petrini, *Clinical ethics recommendations for the allocation of intensive care treatments in exceptional, resource-limited circumstances: The italian perspective during the covid-19 epidemic*, 2020.

[226] T. S. Verma and J. Pearl, "Equivalence and synthesis of causal models," in *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, 1990.

[227] M. J. Vowels, N. C. Camgoz, and R. Bowden, "D'ya like dags? a survey on structure learning and causal discovery," *ACM Computing Surveys (CSUR)*, 2021.

[228] D. Wang, Y. Yang, C. Tao, Z. Gan, L. Chen, F. Kong, R. Henao, and L. Carin, "Proactive pseudo-intervention: Causally informed contrastive learning for interpretable vision models," *arXiv preprint arXiv:2012.03369*, 2020.

[229] L. Wang, Z. Yang, and Z. Wang, "Provably efficient causal reinforcement learning with confounded observational data," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, pp. 21 164–21 175, Curran Associates, Inc., 2021. URL: https://proceedings.neurips.cc/paper/2021/file/b0b79da57b95837f14be95aaa4d54cf8-Paper.pdf.

[230] T. Wang, J. Huang, H. Zhang, and Q. Sun, "Visual commonsense representation learning via causal inference," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 378–379, 2020.

[231] X. Wang, Y. Du, S. Zhu, L. Ke, Z. Chen, J. Hao, and J. Wang, "Ordering-based causal discovery with reinforcement learning," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, IJCAI International Joint Conferences on Artificial Intelligence Organization, pp. 3566–3573, 2021.

[232] Y. Wang, D. Liang, L. Charlin, and D. M. Blei, "Causal inference for recommender systems," in *Proceedings of the 14th ACM Conference on Recommender Systems*, pp. 426–431, 2020.

[233] D. Wei, T. Gao, and Y. Yu, "Dags with no fears: A closer look at continuous optimization for learning bayesian networks," *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 3895–3906.

[234] B. Wen, L. O. Colon, K. Subbalakshmi, and R. Chandramouli, "Causal-tgan: Generating tabular data using causal generative adversarial networks," *arXiv preprint arXiv:2104.10680*, 2021.

[235] P. Wenk, A. Gotovos, S. Bauer, N. S. Gorbach, A. Krause, and J. M. Buhmann, "Fast gaussian process based gradient matching for parameter identification in systems of nonlinear odes," in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, K. Chaudhuri and M. Sugiyama, Eds., ser. Proceedings of Machine Learning Research, vol. 89, pp. 1351–1360, PMLR, 2019. URL: https://proceedings.mlr.press/v89/wenk19a.html.

[236] D. Westreich, J. K. Edwards, S. R. Cole, R. W. Platt, S. L. Mumford, and E. F. Schisterman, "Imputation approaches for potential outcomes in causal inference," *International journal of epidemiology*, vol. 44, no. 5, 2015, pp. 1731–1737.

[237] J. Witte, L. Henckel, M. H. Maathuis, and V. Didelez, "On efficient adjustment in causal graphs," *The Journal of Machine Learning Research*, vol. 21, no. 1, 2020, pp. 9956–10 000.

[238] F. Wozny, "The impact of covid-19 on airfares—a machine learning counterfactual analysis," *Econometrics*, vol. 10, no. 1, 2022, p. 8.

[239] P. Wu, H. Li, Y. Deng, W. Hu, Q. Dai, Z. Dong, J. Sun, R. Zhang, and X.-H. Zhou, "On the opportunity of causal learning in recommendation systems: Foundation, estimation, prediction and challenges," in *Proceedings of the International Joint Conference on Artificial Intelligence, Vienna, Austria*, pp. 23–29, 2022.

[240] Y. Wu, L. Zhang, X. Wu, and H. Tong, "Pc-fairness: A unified framework for measuring causality-based fairness," *Advances in neural information processing systems*, vol. 32, 2019.

[241] D. Xu, Y. Wu, S. Yuan, L. Zhang, and X. Wu, "Achieving causal fairness through generative adversarial networks," *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, DOI: 10.24963/ijcai.2019/201.

[242] G. Xu, T. D. Duong, Q. Li, S. Liu, and X. Wang, "Causality learning: A new perspective for interpretable machine learning," *arXiv preprint arXiv:2006.16789*, 2020.

[243] C.-H. H. Yang, I. Hung, T. Danny, Y. Ouyang, and P.-Y. Chen, "Causal inference q-network: Toward resilient reinforcement learning," *arXiv preprint arXiv:2102.09677*, 2021.

[244] F. Yang, Q. Zhang, X. Ji, Y. Zhang, W. Li, S. Peng, and F. Xue, "Machine learning applications in drug repurposing," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 14, no. 1, 2022, pp. 15–21.

[245] L. Yao, Y. Li, B. Ding, J. Zhou, J. Liu, M. Huai, and J. Gao, "Path-specific causal fair prediction via auxiliary graph structure learning," 2021.

[246] J. Yoon, A. M. Alaa, M. Cadeiras, and M. Van Der Schaar, "Personalized donor-recipient matching for organ transplantation," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[247] J. Yoon, W. R. Zame, A. Banerjee, M. Cadeiras, A. M. Alaa, and M. van der Schaar, "Personalized survival predictions via trees of predictors: An application to cardiac transplantation," *PloS one*, vol. 13, no. 3, 2018, e0194985.

[248] Y. Yu, J. Chen, T. Gao, and M. Yu, "DAG-GNN: DAG structure learning with graph neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, K. Chaudhuri and R. Salakhutdinov, Eds., ser. Proceedings of Machine Learning Research, vol. 97, pp. 7154–7163, PMLR, 2019. URL: https://proceedings.mlr.press/v97/yu19a.html.

[249] Y. Yu, T. Gao, N. Yin, and Q. Ji, "Dags with no curl: An efficient dag structure learning approach," in *Proceedings of the 38th International Conference on Machine Learning*, M. Meila and T. Zhang, Eds., ser. Proceedings of Machine Learning Research, vol. 139, pp. 12 156–12 166, PMLR, 2021. URL: https://proceedings.mlr.press/v139/yu21a.html.

[250] W. R. Zame, I. Bica, C. Shen, A. Curth, H.-S. Lee, S. Bailey, J. Weatherall, D. Wright, F. Bretz, and M. van der Schaar, "Machine learning for clinical trials in the era of covid-19," *Statistics in Biopharmaceutical Research*, vol. 12, no. 4, 2020, pp. 506–517.

[251] J. Zhang and P. Spirtes, "Detection of unfaithfulness and robust causal inference," *Minds and Machines*, vol. 18, no. 2, 2008, pp. 239–271.

[252] J. Zhang, Y. Chen, and A. Singh, "Causal bandits: Online decision-making in endogenous settings," *arXiv preprint arXiv: 2211.08649*, 2022.

[253] J. Zhang, "Designing optimal dynamic treatment regimes: A causal reinforcement learning approach," in *Proceedings of the 37th International Conference on Machine Learning*, H. D. III and A. Singh, Eds., ser. Proceedings of Machine Learning Research, vol. 119, pp. 11 012–11 022, PMLR, 2020. URL: https://proceedings.mlr.press/v119/zhang20a.html.

[254] K. Zhang and A. Hyvarinen, "On the identifiability of the post-nonlinear causal model," *arXiv preprint arXiv:1205.2599*, 2012.

[255] W. Zhang, J. Liao, Y. Zhang, and L. Liu, "Cmgan: A generative adversarial network embedded with causal matrix," *Applied Intelligence*, 2022, pp. 1–13.

[256] X. Zhang, Y. Wong, X. Wu, J. Lu, M. Kankanhalli, X. Li, and W. Geng, "Learning causal representation for training cross-domain pose estimator via generative interventions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11 270–11 280, 2021.

[257] Y. Zhang, J. Berrevoets, and M. van der Schaar, "Identifiable energy-based representations: An application to estimating heterogeneous causal effects," in *International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 4158–4177, 2022.

[258] A. Zhavoronkov, Q. Vanhaelen, and T. I. Oprea, "Will artificial intelligence for drug discovery impact clinical pharmacology?" *Clinical Pharmacology & Therapeutics*, vol. 107, no. 4, 2020, pp. 780–785.

[259] X. Zheng, B. Aragam, P. K. Ravikumar, and E. P. Xing, "Dags with no tears: Continuous optimization for structure learning," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[260] X. Zheng, C. Dan, B. Aragam, P. Ravikumar, and E. Xing, "Learning sparse nonparametric dags," in *International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 3414–3425, 2020.

[261] G. Zhou, L. Yao, X. Xu, C. Wang, L. Zhu, and K. Zhang, "On the opportunity of causal deep generative models: A survey and future directions," *arXiv preprint arXiv:2301.12351*, 2023.

[262] S. Zhu, I. Ng, and Z. Chen, "Causal discovery with reinforcement learning," in *International Conference on Learning Representations*, 2020. URL: https://openreview.net/forum?id=S1g2skStPB.

[263] W. Zhu, C. Yu, and Q. Zhang, "Causal deep reinforcement learning using observational data," *arXiv preprint arXiv:2211.15355*, 2022.