Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Engineering

Data-driven causal inference of process-structure relationships in nanocatalysis

Jonathan YC Ting and Amanda S Barnard

While the field of nanocatalysis has benefited from the application of conventional machine learning methods by leveraging the correlations between processing/structure/property variables, the outcomes from purely correlational studies lack actionability due to missing mechanistic insights. Statistical learning, particularly causal inference, can potentially provide access to more actionable insights by allowing the discovery and verification of deeply obscured causal relationships between variables, using strong correlations identified from interpretable machine learning models as starting points. Recent studies that exemplify the collaborative usage of correlational and causal analysis in catalysis are discussed, including studies potentially benefiting from this approach. Some challenges remaining in the application of inference techniques to the field are identified and suggestions of future directions are provided.

Address
School of Computing, Australian National University, 145 Science Road, Acton, ACT 2601, Australia

Corresponding author:
Barnard, Amanda S (Amanda.S.Barnard@anu.edu.au)

Current Opinion in Chemical Engineering 2022, **36**:xx-yy

This review comes from a themed issue on **Reaction engineering and catalysis**

Edited by **Lars Grabow** and **Ryan Hartman**

https://doi.org/10.1016/j.coche.2022.100818

Introduction

Historically, the discovery of new molecules and materials has mostly been reliant on empirical experiments until the development of theoretical models based on physical laws. These models facilitated the predictions related to chemical reactions, which guided research in chemical engineering and underpin the design of catalysts. However, the complexity of theoretical models grows exponentially when applied to nontrivial systems, rendering analytical solutions impossible. The bottleneck was overcome by the emergence of computers in the last century. Computational chemistry and materials science rose to popularity as the ability to conduct experiments *in silico* enables more complex phenomena to be studied with better precision and control. Nonetheless, a significant proportion of discovery efforts still employ the Edisonian approach, where different combinations of composition and synthesis parameters are systematically tested via trial-and-error in hope of identifying potential candidates with desirable properties [1]. This is an impediment to progress in the field.

Since complex molecules and materials are intrinsically multi-functional and are affected by multiple components, chemical engineering of catalysts typically involves many degrees of freedom, especially when industrial sustainability is required to be taken into account, making the trial-and-error Edisonian approach even more inefficient. Consequently, there is a rise in demand for techniques adaptable to highly complex systems while complementing domain expertise.

With the adoption of data-intensive science into the natural sciences, we are experiencing the fourth paradigm shift and the coming of new opportunities [2]. A fundamentally different path of using models to describe complex phenomena is now available. In contrast to conventional practices where models were based on the knowledge and intuition of researchers, it is now possible to construct models of which we could extract useful knowledge from directly from the data [3]. As more data becomes available, this approach becomes more feasible. The adoption of the data-driven approach by researchers for chemical and materials design has given rise to the field of cheminformatics [4] and materials informatics [5], which could be defined as the application of information sciences to solve chemical and materials problems. Nanoinformatics, focusing on the data-intensive study of nanomaterials with both chemical and materials characteristics, lies somewhere in between [6]. These highly interdisciplinary fields involve high-throughput analysis of complex data using multivariate statistics and artificial intelligence (AI) in conjunction with standard experimental and computational tools, and now serve as an alternative to explicit mathematical modelling for extraction of new insights from data [7,8].

A goal of researchers in informatics is to develop reliable, actionable predictions for which we can be accountable. The term 'actionability' refers to the ability of the model outcomes to inform researchers of potentially useful actions to take to achieve a certain goal. This review will

cover recent examples of informatics applications in reaction engineering, a discussion about a promising direction to increase the actionability of insights obtained from such studies, and the challenges that need to be tackled to do so.

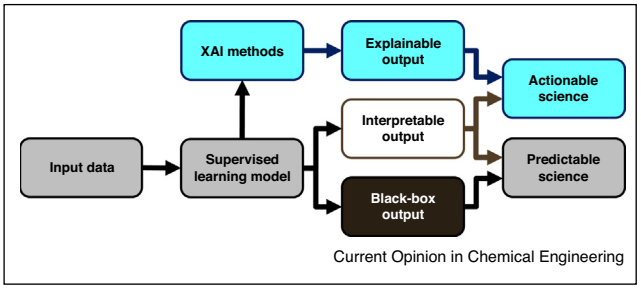
Machine learning

Machine learning (ML) is an essential component of modern data-driven scientific discovery [9] concerned with computer algorithms that learn to perform specific tasks based on historical data without being explicitly programmed [10]. ML problems are categorised into unsupervised learning (UL), supervised learning (SL), semi-supervised learning, and reinforcement learning [11]. Only the first two are discussed here.

The objective of UL is to discover inherent patterns in an unlabelled dataset, such as the similarities in chemical structures, in the absence of chemical properties. The inputs are 'fingerprints' or features that represent structural characteristics, which can be grouped into descriptors. Common UL tasks include cluster analysis [12], archetypal analysis (AA) [13], and dimensionality reduction (DR) [14]. Cluster analysis involves the grouping of data instances (chemicals or materials) based on their similarities or differences using distance metrics. Representative instances (prototypes) can be identified from each cluster centroid. AA aims to approximate all samples in a dataset as a linear combination of extremal points that reside on the convex hull. The 'pure type' archetypes can be identified even if they do not exist in the training set. Clustering and AA are useful for the engineering of chemical reactions as the prototypes and archetypes could be used to form representative yet robust subsets of data with a manageable size [15•]. DR involves obtaining lower-dimensional representations of data, which allows simplification and acceleration of model training and improvement in model generalisability. DR is typically used to reduce the number of features to a manageable size while also preserving as much information as possible.

SL involves constructing a mathematical model using a dataset labelled by target outputs to predict output values of unseen instances. The target labels are dependent variables representing properties of interest. An SL model learns an input–output mapping function that minimises a loss in the model prediction: the structure/property relationship. The training involves gradual adjustments of the mapping function parameters using optimisation algorithms. The two most common SL tasks are classification and regression, which predict categorical (discrete) and numerical (continuous) outputs, respectively. While the application of SL to the modelling of quantitative structure/property relationships (QSPR) are very common [16], SL applications that model processing/structure or processing/property relationships are rare [17]. This is likely due to the relatively lower availability of measurable or calculable processing parameters compared to structural and property variables, but the processing/structure and processing/property relationships are more informative or directly actionable for manufacturers, provided the relationships are not spurious.

Figure 1



A supervised learning workflow learns a model from input data to predict an output. A predictable outcome is not necessarily actionable. The actionability of the model outcomes relies on the ability to understand the rationale behind the model decisions and the context of the predictions, and adapt the original experimental strategy. An actionable outcome is promoted by the incorporation of explainable artificial intelligence (XAI) methods and interpretable models. Examples of a black-box model and an interpretable model are artificial neural networks and decision trees respectively [23], while the examples of explainable methods (applied to model outputs) include Shapley additive explanations (SHAP) [24] and locally interpretable model-agnostic (LIME) explanations [25].

The key to obtaining meaningful outcomes from SL lies in the interpretability (degree to which one can understand the reason behind model predictions [18]) and explainability (degree to which one can understand the model outcome in a given context [19]) of the model. The need for high performing yet interpretable models has reignited public interests in explainable artificial intelligence (XAI) research, which focuses on the understanding and interpretation of AI model behaviours [20]. A variety of model specific and agnostic ML interpretability techniques and XAI methods suited for different data and model types are available [21]. A common interpretable output from many classifiers and regressors is the feature importance profile, which ranks the input variables based on how often they are needed to determine model outputs [22]. As illustrated in Figure 1, interpretable models and XAI methods ensure our ability to understand model decisions, which is essential if an actionable outcome is needed, either for experimental design or establishment of a new theory.

Examples

Typical applications of ML techniques in chemical product engineering address a few major objectives, namely prediction of reaction outcomes, design of new products, sensorial analysis, and modelling of processes [17,26,27].

Some examples related to platinum nanocatalysts and porous carbon materials are discussed here for illustrative purposes.

Sun and Barnard compared multiple types of UL methods to visualise hidden structure/property relationships within silver and platinum nanoparticles [28•]. Kohonen networks were found to be especially suited for the nonlinear and imbalance materials data sets compared to other more commonly used DR methods. In another UL study, the prototypes and archetypes of a set of disordered platinum nanoparticles were identified via the iterative label spreading (ILS) clustering method and AA, resulting in a data subset that enables the structure/property relationships within disordered platinum nanoparticles to be investigated more rigorously [15•].

In a study involving a combination of UL and SL, two groups of platinum nanoparticles were identified via clustering which exhibited different structure/property relationships modelled using interpretable classification and regression [29•]. The clustered nanoparticles were classified by their degrees of surface disorder before class-dependent predictions of their catalytic performance were obtained. It was found that more disordered nanoparticles perform better for oxygen reduction reactions (ORRs) if their surfaces are less ordered and hydrogen oxidation reactions (HORs) if they are smaller in size, where edge-like atoms are maximised. The performance of ordered nanoparticles on ORRs and HORs are correlated to {111} and {110} surface areas, respectively. While these findings agree with experimental observations, no mechanistic insight could be obtained from the purely correlative study.

Supervised processing/structure and structure/property relationships of platinum nanoparticles have also been explored using extremely randomised regressor trees (ExtraTrees) and genetic algorithms [30]. The bulk and surface disorder parameters were predicted from the growth time, atom deposition rate, and temperature while various structural parameters were used to predict some catalytic activity indicators. Although reasonable predictive performance was achieved and the correlations identified among the structural and properties parameters were sensible according to established theory, it was pointed out that the lack of processing and property parameters in the study had limited the usefulness of the insights obtained for the design of more efficient nanoparticles.

Maulana Kudshany and Lyth also applied the tree-based random forest regressor to predict hydrogen uptake in porous carbon materials at 77 K based on their chemical and textural features, and analysed the relative importance of the features using SHAP [31•]. They identified pressure and Brunauer–Emmett–Teller surface area as the features that contributed the most towards successful model predictions, while correlations between oxygen and nitrogen content with hydrogen uptake contradicted claims made in previous studies. The authors took care to clarify that the study was not causal, pointing out that unconsidered confounder variables (correlated to both hydrogen uptake and other variables in the study) could be responsible, especially for variables that had previously received little attention from the research community such as oxygen content.

In cases where causal relationships are unclear, it is difficult to infer new scientific knowledge from the correlations identified by conventional SL. This highlights the importance of comprehensive features extraction, and the risk of misinterpreting the meaning of structure/property relationships based solely on correlations, which are nondeterministic and insufficient to achieve rational design of molecules and materials. This is highly relevant to the issue of actionability.

Statistical learning

Actionability can be addressed using statistical learning [32]. Like ML, statistical learning is based on data, but with some fundamental differences: statistical learning relies on rule-based programming, in the form of relationship between variables, whereas ML learns from data without explicitly programmed instructions; statistical learning is more mathematics intensive and focused on estimating coefficients, while ML is used to find trends and patterns; statistical learning is dependent on more stringent assumptions such as normality, no multicollinearity, and homoscedasticity, which are not necessarily required for ML (which have their own assumptions such as the i.i.d. assumption). The most commonly recognised difference is that statistical learning is concerned with inferences (of known facts), in comparison to ML which emphasises predictions (of an unknown).

While the application of statistical learning methods such as the information entropy concept has been foundational to the field of chemistry [33], this article will mainly discuss a specific area of statistical learning that aims to identify cause-and-effect relationships. There are currently three state-of-the-art languages for causal inference, namely structural equation models, potential outcomes models, and graphical models. While Pearl proved that they are interconnected [34], each of them has advantages for different purposes.

Potential outcome models are mainly used in genetics [35], economics [36], and epidemiology [37]. They enable assumption refinement, combination of experimental and observational data, bounding probabilities of necessary and sufficient causation, and offers great mathematical convenience, but the formulation of their causal

assumptions is typically not straightforward, especially when the system investigated is highly complex [38].

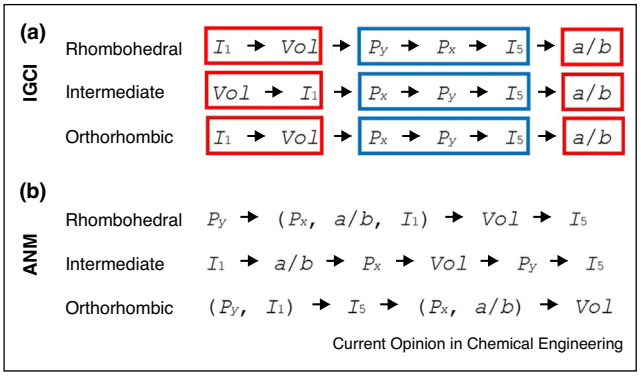
Structural equation models are mainly used in social sciences [39], philosophy [40], and ecology [41]. They are capable of handling complex systems, as they allow testing of the models by imposing and assessing certain structures and estimation of latent variables via observed variables [42]. This encourages the formulation of clearer hypotheses before the commencement of data analysis. Structural equation models also allow explicit assessment of measurement errors, but suffer from the problem of potentially omitted variables, negligence of lower-order model components, and propagation of specification errors.

Graphical models have mainly been used in genetics [43], computer science [44], and medicine [45]. They allow easy visualisation of causal assumptions and identification of bias sources that are difficult to be detected using other alternatives [46]. Graphical models are difficult for statistical inference because the model is nonparametric, but the majority of causal inference in chemical engineering uses this approach [47,48]; specifically Bayesian networks (BNs) as they are capable of capturing complex relationships between variables in an informative way.

BNs are probabilistic graphical models that represent variables dependency structure in the form of joint distributions [49]. They consist of directed acyclic graphs (DAGs), where the nodes and edges represent variables and the relationships between them, respectively. In a DAG, each edge is orientated from one node (parent node) to another (child node). A DAG does not contain any cyclic path, which is a sequence of edges that end at the starting node of its first edge. The learning of a BN from data includes structure learning, where the topology of the network is built, and parameter learning, where the conditional probabilities that govern the networks are identified [50]. The direction and strength of the inter-variable dependencies can then inform decisions.

BNs are reasonably interpretable and easily explainable, and multiple metrics can be used to measure the significance of the relationships and help to identify the effect of specific actions. The learnt graph provides a compact knowledge representation about a particular system, which enables intuitive visualisation of the relationships among the variables. Another incentive of using BNs is the possibility of encoding domain expertise into the process of learning statistically significant information from the data. A drawback of BNs is that the computational complexity of their learning process increases exponentially with the number of variables, so careful feature selection is necessary to ensure the relevance and sufficiency of variables included in the model.

Figure 2



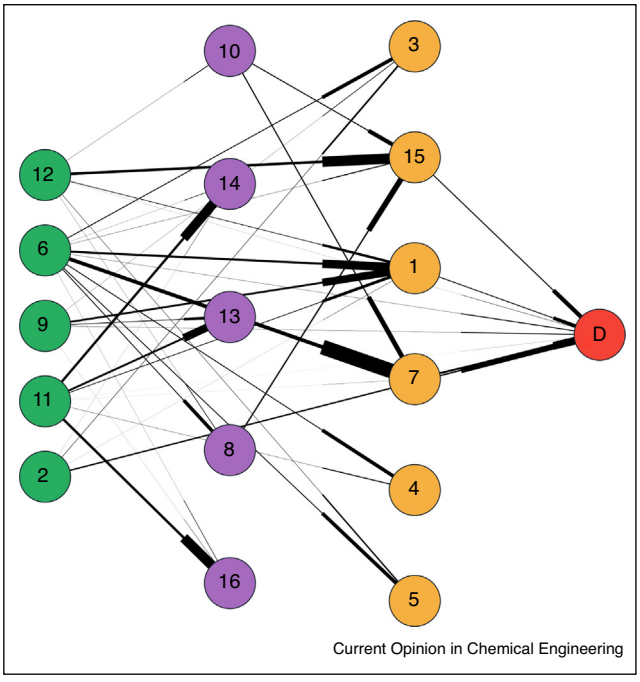
Causal chains for 0% samarium (rhombohedral), 7% samarium (intermediate), and 20% samarium (orthorhombic) doped bismuth ferrite compositions based on **(a)** information-geometric causal inference (IGCI) and **(b)** additive noise model (ANM) methods. The position of the descriptor in the chain indicates likely cause-and-effect relationship, which can be sensitive to the presence of confounders or observational bias. I_1, *Vol*, P_x, P_y, I_5, and a/b represent chemical composition, molar volume, polarisation component in x and y directions, differential chemical composition, and tetragonality, respectively. The red and blue boxes highlight the general trend in IGCI causal chains, where chemical effects (I_1 and *Vol*) were found to reside higher in the causal chains, I_5 always followed after the polarisation components (P_x and P_y), and a/b always ranked last. Reproduced 'Causal chain analysis' from reference [51••] licensed under CC BY 4.0 with permission from npj Computational Materials.

Examples of causal inference in reaction engineering and catalysis are still relatively rare.

Applications

One recent application was reported by Ziatdinov, who performed a mechanistic analysis on the emergence of unique functionalities within ferroelectric metal oxides [51••]. Information-geometric causal inference and additive noise model analysis were used to establish pairwise causal relationships between compositional, structural, and polarisation field descriptors, and order the dataset in the identified causal direction. On the basis of the causal chains obtained as displayed in Figure 2, the authors suggested that a common causal mechanism that govern the properties of ferroelectric perovskites exist across the composition series. Some interesting insights obtained include the greater tendency for features related to chemical effects such as local chemical composition and molar volume to be the primary causes of other features, and the consistent position of the differential contrast between A and B site cation intensity (I_5) after the polarisation components, which indicates that I_5 is more related to physical distortion rather than chemical composition. It was concluded that causal analysis of such multi-modal data can enable the exploration of causal links between competing mechanisms in non-equilibrium non-ergodic materials, improving our ability for materials optimisation.

Figure 3



Directed acyclic graph illustrating a Bayesian network causation model. Each node is a feature and each edge represents the causal relationship directed from the left (parent) node to right (child) node. The node numbers represent the order of the feature importance of the variables as listed in the original publication. The thickness of the edges of the graph is weighted based on the strength of the causal relationship between the connected pairs. The nodes are arranged in layers, based on whether the feature is a primary cause (green), secondary cause (purple), associative (yellow), or target (red). Primary causes are features characterised directly from initial graphene structures and represent sampling criteria that reflect potential design decisions made in laboratories, secondary causes are features resulted from input sampling decisions but cannot necessarily be controlled in laboratories, associatives are features resulted from simulation outputs, while the target is the observation feature of interest, in this case being defect concentration, indicated by D. Reprinted with permission from citation [52••]. Copyright 2020 American Chemical Society.

Another work by Motevalli involves an analysis of factors contributing to defects in graphene oxide nanoflakes [52••]. Feature importance profiles were first obtained from logistic and Ridge regressions of labels characterising defects formation and distributions based on 220 structural features. To decrease the complexity and facilitate the use of a BN to identify which correlations were causal, recursive feature elimination was applied to choose 16 features that are most highly correlated with the labels. The learnt BN was illustrated in Figure 3. It was found, for the first time, that the concentration of broken bond defects in graphene oxide nanostructures is largely determined by the presence and distribution of hydrogen (rather than oxygen) via several well-defined pathways, such as hydrogen concentration (node 2) causing graphene oxide defects (node D) directly, and C–H coordination (node 6) causing C–C coordination (node 7), which is closely associated with node D.

The graphene oxide study serves as a demonstration of how causal models can be appended to conventional ML workflows involving classification and/or regression, in a way generic enough to be applied to other materials. For example, the work of Parker (discussed in Section 'Example') could provide greater insights into the mechanisms behind the catalytic activities of platinum nanoparticles based on their processing and structural parameters if causal relationships between the variables investigated could be found. While only two processing parameters were included in the study, a total of 179 structural (atomic, crystallographic, and topological) features were extracted. Fitting a causal model to a selected set of high-ranking features from the feature importance profiles could identify opportunities for tuning the nanoparticles based on processing conditions and structural features, to inform new experimental strategies for achieving desired catalytic activities.

Similarly, the Maulana Kudshany study of carbon materials could be transformed into a causal study by establishing a few assumptions, namely covariation, temporal precedence, and control for third variables [53]. This is in accordance with the definition of causal inference by Shaughnessy [54]: identification of cause(s) of a phenomenon by establishing cause-effect covariation, a time-order relationship with the cause preceding the effect, and eliminating plausible alternative causes. The responsibility of ensuring the validity of these assumptions nonetheless falls upon the researchers. This highlights the importance of interdisciplinary collaborations in future research as the most critical first step often requires significant domain expertise.

Summary

The field of catalysis and reaction engineering has benefited immensely from the application of ML. By leveraging correlational relationships, ML methods have been successfully used to predict the properties of chemical products based on their structures or synthesis conditions. However, purely correlational studies are vulnerable to erroneous conclusions due to spurious correlations and the ignorance of succession (inputs not preceding outputs). Mechanistic insights and answers to causal queries can potentially be provided by statistical learning. Strong correlations identified between variables of interest can serve as starting points to discover and verify deeply obscured causal relationships, but caution should be exercised when using them in isolation. The collaborative usage of correlational and causal analysis improves the credibility of conclusions and makes predictions actionable. However, some challenges in the application of inference techniques to catalysis and reaction engineering remain, beyond the urgent need for more

measured or calculated processing parameters to describe synthesis conditions and latent environmental conditions.

Being the most common method causal inference is conducted in this area, graphical models are limited in several ways. The usage of probabilistic graph models to conduct causal inference can easily result in combinatorial explosions of the search space as the time required to conduct structure learning increases exponentially with network size. Having multiple causal links with similar probabilities that are connected to the same node makes the identification of the optimal path to follow potentially challenging. Some paths may not even terminate on the target label, resulting in redundancy.

The integration of domain knowledge into probabilistic graphs is essential because reverse links that are approximately equal in probability as the forward links might be formed if automatic learning is employed. Domain knowledge needs to be encoded to ensure that the causal directions are intuitive and sequentially correct. Currently, the only way to incorporate domain knowledge into BNs is via direct participation of researchers (interactive learning), where causal links with directions known to be impossible are manually omitted during structure learning. As this can be time-consuming, any breakthrough on the subject of knowledge integration will benefit the field.

The greatest weakness of conducting causal inference is the validity of the assumptions made. While domain expertise is required for causal inference, the resultant models and outcomes obtained are at the same time susceptible to the mistakes in our understanding of the phenomenon studied and flaws in our knowledge-encoding process. The causal links in probabilistic graphs could also involve reaction barriers that are not captured in the network. Missing transition states could potentially cause gaps in researchers' understanding of the mechanism behind the reaction. The presence of transition states needs to be determined in such cases by exploring the possibility of intermediary variables between two nodes, or imposing some penalties to the path. Path optimisation is encountered in other areas of AI such as air traffic control (which also experience barriers between nodes, such as mountains) [55], but their applicability to processing/structure or structure/property relationships has yet to be explored.

While the majority of the causal inference applications in reaction engineering involved probabilistic graphical models due to their intuitive interpretability, there is still plenty of space for exploration of other causal inference frameworks in the field. The challenge lies within the proper formulation of the problem following the assumptions held by each framework (a common one is the inclusion of all possible confounders) and the availability of suitable data. One has to understand the limitations and advantages each of them offers before choosing the appropriate methods, and remain open to the insights they can provide based on relationships that are too complex for conventional analysis.

Conflict of interest statement

Nothing declared.

Acknowledgments

Ting acknowledges the financial support from the Australian National University under the University Research Scholarship.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Cao B, Adutwum LA, Oliynyk AO, Luber EJ, Olsen BC, Mar A, Buriak JM: **How to optimize materials and devices via design of experiments and machine learning: demonstration using organic photovoltaics**. *ACS Nano* 2018, **12**:7434-7444 http://dx.doi.org/10.1021/acsnano.8b04726.
2. Hey T, Tansley S, Tolle K: *The Fourth Paradigm: Data-Intensive Science*. 2009 http://dx.doi.org/10.1007/978-3-642-33299-9_1.
3. Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L, Zdeborová L: **Machine learning and the physical sciences**. *Rev Mod Phys* 2019, **91**:45002 http://dx.doi.org/10.1103/RevModPhys.91.045002.
4. Martinez-Mayorga K, Madariaga-Mazon A, Medina-Franco JL, Maggiora G: **The impact of chemoinformatics on drug discovery in the pharmaceutical industry**. *Expert Opin Drug Discov* 2020, **15**:293-306 http://dx.doi.org/10.1080/17460441.2020.1696307.
5. Yosipof A, Khalemsky A, Gelbard R, Senderowitz H: **Dynamic classification for materials-informatics: mining the solar cell space**. *Mol Inform* 2020, **39**:2000173 http://dx.doi.org/10.1002/minf.202000173.
6. Barnard AS, Motevalli B, Parker AJ, Fischer M, Feigl CA, Opletal G: **Nanoinformatics, and the big challenges for the science of small things**. *Nanoscale* 2019, **11**:19190-19201 http://dx.doi.org/10.1039/c9nr05912a.
7. Medford AJ, Kunz MR, Ewing SM, Borders T, Fushimi R: **Extracting knowledge from data through catalysis informatics**. *ACS Catal* 2018, **8**:7403-7429 http://dx.doi.org/10.1021/acscatal.8b01708.
8. Wang M, Wang T, Cai P, Chen X: **Nanomaterials discovery and design through machine learning**. *Small Methods* 2019, **3**:1900025 http://dx.doi.org/10.1002/smtd.201900025.
9. Saal JE, Oliynyk AO, Meredig B: **Machine learning in materials discovery: confirmed predictions and their underlying approaches**. *Annu Rev Mater Res* 2020, **50**:49-69 http://dx.doi.org/10.1146/annurev-matsci-090319-010954.
10. Mitchell TM: *Machine Learning*. New York: McGraw-Hill; 1997 https://www.springer.com/gp/book/9780898382143.
11. Sarker IH: **Machine learning: algorithms, real-world applications and research directions**. *SN Comput Sci* 2021, **2**:160 http://dx.doi.org/10.1007/s42979-021-00592-x.
12. **Clustering**. *Encycl. Mach. Learn.*. Edited by Sammut C, Webb GI. *Springer US, Boston, MA: 2010:180 http://dx.doi.org/10.1007/978-0-387-30164-8_124*.
13. Cutler A, Breiman L: **Archetypal analysis**. *Technometrics* 1994, **36**:338-347 http://dx.doi.org/10.1080/00401706.1994.10485840.

14. Velliangiri S, Alagumuthukrishnan S, Thankumar Joseph SI: **A review of dimensionality reduction techniques for efficient computation**. *Proc Comput Sci* 2019, **165**:104-111 http://dx.doi.org/10.1016/j.procs.2020.01.079.

15. • Parker AJ, Motevalli B, Opletal G, Barnard AS: **The pure and representative types of disordered platinum nanoparticles from machine learning**. *Nanotechnology* 2021, **32**:95404 http://dx.doi.org/10.1088/1361-6528/abcc23
This work involved the use of unsupervised learning methods including iterative label spreading clustering and archetypal analysis to identify the prototypes and archetypes of a set of disordered platinum nanoparticles structural data. A robust subset of data that describe the variance of all structures collected well was formed as a result. It demonstrated some useful applications of these unsupervised learning techniques in high dimensional data.

16. Muratov EN, Bajorath J, Sheridan RPI, Tetko V, Filimonov D, Poroikov V, Oprea TI, Baskin II, Varnek A, Roitberg A, Isayev O, Curtalolo S, Fourches D, Cohen Y, Aspuru-Guzik A, Winkler DA, Agrafiotis D, Cherkasov A, Tropsha A: **QSAR without borders**. *Chem Soc Rev* 2020, **49**:3525-3564 http://dx.doi.org/10.1039/D0CS00098A.

17. Trinh C, Meimaroglou D, Hoppe S: **Machine learning in chemical product engineering: the state of the art and a guide for newcomers**. *Processes* 2021, **9**:1456 http://dx.doi.org/10.3390/pr9081456.

18. Miller T: **Explanation in artificial intelligence: insights from the social sciences**. *Artif Intell* 2019, **267**:1-38 http://dx.doi.org/10.1016/j.artint.2018.07.007.

19. Roscher R, Bohn B, Duarte MF, Garcke J: **Explainable machine learning for scientific insights and discoveries**. *IEEE Access* 2020, **8**:42200-42216 http://dx.doi.org/10.1109/ACCESS.2020.2976199.

20. Linardatos P, Papastefanopoulos V, Kotsiantis S: **Explainable AI: a review of machine learning interpretability methods**. *Entropy* 2021, **23** http://dx.doi.org/10.3390/e23010018.

21. Gunning D, Aha D: **DARPA's explainable artificial intelligence (XAI) program**. *AI Mag* 2019, **40**:44-58 http://dx.doi.org/10.1609/aimag.v40i2.2850.

22. Barnard AS, Opletal G: **Selecting machine learning models for metallic nanoparticles**. *Nano Futur* 2020, **4**:35003 http://dx.doi.org/10.1088/2399-1984/ab9c3b.

23. Loyola-González O: **Black-box vs. white-box: understanding their advantages and weaknesses from a practical point of view**. *IEEE Access* 2019, **7**:154096-154113 http://dx.doi.org/10.1109/ACCESS.2019.2949286.

24. Lundberg SM, Lee S-I: **A unified approach to interpreting model predictions**. *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*. Red Hook, NY, USA: Curran Associates Inc.; 2017, 4768-4777 http://dx.doi.org/10.5555/3295222.3295230.

25. Ribeiro MT, Singh S, Guestrin C: **'Why should i trust you?': explaining the predictions of any classifier**. *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*. New York, NY, USA: Association for Computing Machinery; 2016, 1135-1144 http://dx.doi.org/10.1145/2939672.2939778.

26. Back S, Tran K, Ulissi ZW: **Toward a design of active oxygen evolution catalysts: insights from automated density functional theory calculations and machine learning**. *ACS Catal* 2019, **9**:7651-7659 http://dx.doi.org/10.1021/acscatal.9b02416.

27. Nandy A, Zhou J, Janet JP, Duan C, Getman RB, Kulik HJ: **Machine learning accelerates the discovery of design rules and exceptions in stable metal-oxo intermediate formation**. *ACS Catal* 2019, **9**:8243-8255 http://dx.doi.org/10.1021/acscatal.9b02165.

28. • Sun B, Barnard AS: **Visualising multi-dimensional structure/property relationships with machine learning**. *J Phys Mater* 2019, **2**:34003 http://dx.doi.org/10.1088/2515-7639/ab0faa
This work compared the performance of multiple types of unsupervised learning techniques in the visualisation of multi-structure/multi-property relationships in silver and platinum nanoparticles structural data sets. The Kohonen networks were found to be superior over other more popular dimensionality reduction methods. It demonstrated the usefulness of unsupervised learning techniques in extracting patterns from high dimensional data.

29. • Parker AJ, Opletal G, Barnard AS: **Classification of platinum nanoparticle catalysts using machine learning**. *J Appl Phys* 2020, **128**:1-11 http://dx.doi.org/10.1063/5.0009129
This work employed unsupervised learning to classify platinum nanoparticle catalysts based on their degrees of surface disorder. Subsequent application of supervised machine learning techniques revealed that each class of nanoparticles exhibit different structure-property relationships, in agreement with experimental observations. It demonstrated the feasibility of combining unsupervised and supervised learning in knowledge extraction from complex data.

30. Sun B, Barron H, Opletal G, Barnard AS: **From process to properties: correlating synthesis conditions and structural disorder of platinum nanocatalysts**. *J Phys Chem C* 2018, **122**:28085-28093 http://dx.doi.org/10.1021/acs.jpcc.8b08386.

31. • Maulana Kusdhany MI, Lyth SM: **New insights into hydrogen uptake on porous carbon materials via explainable machine learning**. *Carbon N Y* 2021, **179**:190-201 http://dx.doi.org/10.1016/j.carbon.2021.04.036
This work predicted hydrogen uptake in porous carbon materials based on their chemical and textural features and found conclusions that are in contrast to claims made in previous studies in their analysis of the relative importance of the features. It highlighted the capability of obtaining interpretable machine learning models by employing post hoc explainable artificial intelligence method such as Shapley Additive Explanations.

32. Hastie T, Tibshirani R, Friedman J: *The Elements of Statistical Learning*. New York, NY, USA: Springer New York Inc; 2001 http://dx.doi.org/10.1007/978-0-387-84858-7.

33. Rajan K: **Materials informatics: the materials "gene" and big data**. *Annu Rev Mater Res* 2015, **45**:153-169 http://dx.doi.org/10.1146/annurev-matsci-070214-021132.

34. Pearl J: **The causal foundations of structural equation modeling**. *Handb. Struct. Equ. Model*. New York, NY, US: The Guilford Press; 2012, 68-91 https://psycnet.apa.org/record/2012-16551-005.

35. Pingault J-B, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F: **Using genetic data to strengthen causal inference in observational research**. *Nat Rev Genet* 2018, **19**:566-580 http://dx.doi.org/10.1038/s41576-018-0020-3.

36. Imbens GW: **Potential outcome and directed acyclic graph approaches to causality: relevance for empirical practice in economics**. *J Econ Lit* 2020, **58**:1129-1179 http://dx.doi.org/10.1257/jel.20191597.

37. Sommer AJ, Leray E, Lee Y, Bind M-AC: **Assessing environmental epidemiology questions in practice with a causal inference pipeline: an investigation of the air pollution-multiple sclerosis relapses relationship**. *Stat Med* 2021, **40**:1321-1335 http://dx.doi.org/10.1002/sim.8843.

38. Pearl J: **Statistics and causal inference: a review**. *Test* 2003, **12**:281-345 http://dx.doi.org/10.1007/BF02595718.

39. Pósch K: **Testing complex social theories with causal mediation analysis and G-computation: toward a better way to do causal structural equation modeling**. *Sociol Methods Res* 2019, **50**:1376-1406 http://dx.doi.org/10.1177/0049124119826159.

40. Bochman A: **Actual causality in a logical setting**. *Proc. Twenty-Seventh Int. Jt. Conf. Artif. Intell. IJCAI-18*; *International Joint Conferences on Artificial Intelligence Organization: 2018:1730-1736 http://dx.doi.org/10.24963/ijcai.2018/239*.

41. Das Gupta S, Pinno BD: **Drivers of understory species richness in reconstructed boreal ecosystems: a structural equation modeling analysis**. *Sci Rep* 2020, **10**:11555 http://dx.doi.org/10.1038/s41598-020-68353-z.

42. Tarka P: **An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences**. *Qual Quant* 2018, **52**:313-354 http://dx.doi.org/10.1007/s11135-017-0469-8.

43. Hines O, Diaz-Ordaz K, Vansteelandt S, Jamshidi Y: **Causal graphs for the analysis of genetic cohort data**. *Physiol*

Genomics 2020, **52**:369-378 http://dx.doi.org/10.1152/physiolgenomics.00115.2019.

44. Luo Y, Peng J, Ma J: **When causal inference meets deep learning**. *Nat Mach Intell* 2020, **2**:426-427 http://dx.doi.org/10.1038/s42256-020-0218-x.

45. Castro DC, Walker I, Glocker B: **Causality matters in medical imaging**. *Nat Commun* 2020, **11**:3673 http://dx.doi.org/10.1038/s41467-020-17478-w.

46. Greenland S, Brumback B: **An overview of relations among causal modelling methods**. *Int J Epidemiol* 2002, **31**:1030-1037 http://dx.doi.org/10.1093/ije/31.5.1030.

47. Sizochenko N, Gajewicz A, Leszczynski J, Puzyn T: **Causation or only correlation? Application of causal inference graphs for evaluating causality in nano-QSAR models**. *Nanoscale* 2016, **8**:7203-7208 http://dx.doi.org/10.1039/c5nr08279j.

48. Jeong J, Song T, Chatterjee N, Choi I, Cha YK, Choi J: **Developing adverse outcome pathways on silver nanoparticle-induced reproductive toxicity via oxidative stress in the nematode *Caenorhabditis elegans* using a Bayesian network model**. *Nanotoxicology* 2018, **12**:1182-1197 http://dx.doi.org/10.1080/17435390.2018.1529835.

49. Zhang J, Yue H, Wu X, Chen W: **A brief review of Bayesian belief network**. *2019 Chinese Control Decis. Conf.* 2019:3910-3914 http://dx.doi.org/10.1109/CCDC.2019.8832649.

50. Scanagatta M, Salmerón A, Stella F: **A survey on Bayesian network structure learning from data**. *Prog Artif Intell* 2019, **8**:425-439 http://dx.doi.org/10.1007/s13748-019-00194-y.

51. •• Ziatdinov M, Nelson CT, Zhang X, Vasudevan RK, Eliseev E, Morozovska AN, Takeuchi I, Kalinin SV: **Causal analysis of competing atomistic mechanisms in ferroelectric materials from high-resolution scanning transmission electron microscopy data**. *NPJ Comput Mater* 2020, **6**:1-9 http://dx.doi.org/10.1038/s41524-020-00396-2

This work performed a causal analysis on the competing atomistic mechanisms in ferroelectric materials from high-resolution scanning transmission electron microscopic data, demonstrating the usefulness of causal inference methods in extraction of mechanistic insights.

52. •• Motevalli B, Sun B, Barnard AS: **Understanding and predicting the cause of defects in graphene oxide nanostructures using machine learning**. *J Phys Chem C* 2020, **124**:7404-7413 http://dx.doi.org/10.1021/acs.jpcc.9b10615

This work employed a Bayesian network to conduct causal inference on defective graphene oxide nanoflakes, identifying for the first time that the frequency of bond breakage is largely determined by the presence and distribution of hydrogen instead of oxygen, providing a demonstrative workflow that could be generalised to other nanomaterials.

53. Oppewal H: **Concept of causality and conditions for causality**. *Wiley Int. Encycl. Mark* 2010 http://dx.doi.org/10.1002/9781444316568.wiem02059.

54. Shaughnessy JJ, Zechmeister EB, Zechmeister JS: **The scientific method**. In *Res. Methods Psychol.,* edn 9. Edited by Ryan M. New York, NY, US: McGraw-Hill; 2012:27-56 https://www.worldcat.org/title/research-methods-in-psychology/oclc/664665878.

55. Geiβer F, Povéda G, Trevizan F, Bondouy M, Teichteil-Königsbuch F, Thiébaux S: **Optimal and heuristic approaches for constrained flight planning under weather uncertainty**. *Proc Int Conf Autom Plan Sched* 2020, **30**:384-393 https://ojs.aaai.org/index.php/ICAPS/article/view/6684.