

ORIGINAL RESEARCH

Open access

Causal Inference in Microstructure–Property Mapping: A Framework for Learning under Unobserved Confounding

George Brown¹, Michael Taylor^{1*}, Sarah Wilson², Olivia Harris¹

Abstract

Microstructure–property mapping remains a foundational challenge in materials science and engineering. Microstructural features such as grain size, phase distribution, and defect density are known to govern macroscopic properties including strength, conductivity, and toughness. Yet the relationship between microstructure and property is rarely a simple direct link. Processing history, trace impurities, measurement conditions, and other hidden factors act as confounders that simultaneously influence both microstructure and observed properties. Standard machine learning approaches, which dominate current computational materials engineering, learn statistical correlations from historical datasets rather than causal mechanisms. As a result, these models frequently fail when applied to new microstructures produced under different processing conditions or in novel material systems. Unobserved confounding introduces systematic bias that cannot be corrected through conventional statistical adjustment. Historical processing data are often incomplete or proprietary, impurities fall below detection limits, and measurement artifacts vary across laboratories. Consequently, correlation-based predictions break down under distribution shifts, limiting their utility for materials design and extrapolation. This paper presents a conceptual causal inference framework tailored to microstructure–property mapping in the presence of unobserved confounders. The framework rests on four core components: causal graph specification to formalize directed relationships among microstructure features (M), properties (Y), observed confounders, and unobserved confounders; confounder identification to separate what can and cannot be measured; instrumental variable estimation to recover unbiased causal effects; and counterfactual reasoning to answer interventional questions essential for design. By integrating these elements, the framework enables predictions that remain robust when processing histories change and supports counterfactual inquiries of the form “what would the property be if the microstructure were altered while holding all else fixed?” Recent advances in causal analysis of microstructural data and domain-adapted representation learning illustrate the practical relevance of such an approach. The framework also connects to causal discovery techniques suitable for high-dimensional microstructural images and time-series processing records. Ultimately, it shifts computational materials engineering from correlational prediction toward causal understanding, offering a pathway to reliable extrapolation, improved transfer across material classes, and more effective rational design of next-generation materials with targeted performance.

Keywords Causal inference, Counterfactual reasoning, Microstructure–property relationships, Unobserved confounding, Instrumental variables, Causal discovery

*Correspondence:

Michael Taylor

michael.taylor@outlook.com

¹ Department of Data-Driven Materials Engineering, Faculty of Engineering, University of Auckland, Auckland, New Zealand

² Department of Smart Materials Systems, Faculty of Science and Technology, University of Otago, Dunedin, New Zealand

Introduction

Microstructure determines properties. But the mapping from microstructure to property is not a simple correlation. Processing history, impurities, measurement conditions—all are confounders that affect both microstructure and property. Standard machine learning models learn correlations, not causal relationships. When deployed on new microstructures with different processing histories, they fail. This paper proposes a causal inference framework for microstructure–property mapping that explicitly handles unobserved confounding, enabling predictions that are robust to distribution shift and useful for counterfactual design.

In computational and data-driven materials engineering, microstructure–property relationships form the cornerstone of both scientific understanding and technological advancement. Grain boundaries, precipitates, dislocations, and phase morphologies dictate mechanical strength, thermal conductivity, corrosion resistance, and countless other functional attributes. For decades, researchers have relied on empirical observation and physics-based modeling to link these features to performance. More recently, machine learning has accelerated discovery by fitting high-dimensional microstructural descriptors to property values extracted from large datasets [1, 2]. Yet these models remain fundamentally correlational. They excel at interpolation within the training distribution but collapse when the underlying data-generating process changes—even slightly.

The root cause of this fragility is unobserved confounding. Variables such as prior thermal history, unintentional impurities at parts-per-million levels, or subtle variations in characterization protocols influence both the observed microstructure and the measured property without being recorded in the dataset. Because these confounders are not measured, standard regression or neural-network approaches absorb their effects into spurious associations. A model may correctly predict that finer grain size correlates with higher yield strength in a particular alloy batch, yet the correlation arises partly from an unrecorded cooling rate that simultaneously refines grains and strengthens the matrix. When the same model is applied to a microstructure produced under a different cooling schedule, the prediction deviates sharply [3, 4].

Such failures are not merely academic. In industrial materials design, engineers routinely need to extrapolate: altering a processing parameter to achieve a desired microstructural change and reliably forecasting the resulting property. Correlation-based models cannot guarantee that the predicted change reflects a true causal effect. Counterfactual questions—“If we refine the grain size while keeping all other factors fixed, what happens to toughness?”—remain unanswered. Recent work on causal discovery in microstructural datasets underscores the growing recognition that purely predictive models are insufficient for forward design [5–7]. Similarly, studies in ferroelectric and catalytic materials have begun to separate competing atomistic mechanisms through causal lenses rather than black-box fitting [3, 8].

The present conceptual framework addresses this gap by embedding causal inference principles directly into microstructure–property modeling. It begins with explicit causal graph specification, proceeds through confounder identification and sensitivity analysis, employs instrumental variable strategies where feasible, and culminates in counterfactual reasoning for design. By doing so, the framework distinguishes itself from domain adaptation techniques that merely reweight distributions without explaining the source of shift [9, 10]. It also extends beyond explainable machine learning by providing actionable interventional predictions rather than post-hoc feature importance [2].

Throughout the following sections, the framework is developed conceptually, without reference to specific datasets or performance benchmarks. Its components are illustrated through representative materials scenarios drawn from alloys, ceramics, and composites. The goal is to equip computational materials scientists with a structured way to reason about causation under the pervasive reality of unobserved confounding. In doing so, the framework lays groundwork for a causal materials science capable of supporting robust extrapolation and rational microstructural design.

Correlation vs. Causation in Microstructure–Property Mapping

Correlation — A statistical association between microstructure features and properties. Does not imply that changing microstructure will change property.

Causation — A relationship where changing microstructure directly changes property, holding all else fixed.

Why correlation is insufficient: Spurious correlations arise when two variables appear linked only because a third factor influences both. In microstructure–property mapping, grain size and yield strength are often correlated, but both may be driven by an unrecorded processing temperature that simultaneously controls grain growth and precipitation hardening. A machine learning model trained on such data will attribute strength changes to grain size alone, yet the apparent relationship dissolves when temperature varies independently. Similar spurious associations appear across materials classes: precipitate volume fraction and fatigue life may correlate through cooling rate; defect density and electrical resistivity may correlate through measurement humidity. Standard models capture these associations faithfully within the training domain but cannot distinguish them from true causal pathways [1, 2].

Machine learning algorithms optimize for predictive accuracy on held-out samples drawn from the same distribution. They therefore encode any statistical dependence, whether causal or confounded. When the confounder distribution shifts—as it inevitably does in new processing facilities, scaled-up production, or novel alloy compositions—the learned mapping no longer holds. Recent literature on machine learning for concrete and high-entropy alloys illustrates how correlation-based models achieve impressive in-sample performance yet degrade rapidly under even modest changes in processing variables [1, 11, 12].

Why causation is needed for design: Materials design is inherently interventional. Engineers ask what will happen if they modify a processing step to alter grain size, phase fraction, or texture. Only a causal relationship supports reliable answers to such questions because it isolates the direct effect while holding confounders fixed. Correlation-based predictions, by contrast, mix the desired intervention with uncontrolled changes in hidden variables, leading to optimistic or pessimistic bias. Counterfactual reasoning—“what would the property have been under a different microstructure but identical processing history?”—becomes possible only within a causal framework [3, 5].

The distinction carries practical consequences for transferability. A model trained on laboratory-scale microstructures may fail when applied to industrially processed material because the unobserved processing confounders differ. Causal models, by explicitly modeling these confounders, remain stable across such shifts. Moreover, causation clarifies which microstructural features are actionable levers for property optimization and which are merely correlated byproducts of upstream decisions.

Table 1 contrasts correlational and causal paradigms, highlighting their fundamentally different assumptions, objectives, and implications for materials design.

Table 1. Structural Comparison of Correlational and Causal Paradigms in Microstructure–Property Modeling

Dimension	Correlational Paradigm	Causal Paradigm
Objective	Minimize predictive error on observed data	Estimate invariant causal effect of M on Y
Treatment of Confounders	Implicitly absorbed into model	Explicitly modeled (C_obs, C_unobs)
Handling of Unobserved Confounding	Ignored	Addressed via IVs and sensitivity analysis
Interpretation of M–Y Relationship	Statistical association	Interventional effect (do(M))
Response to Distribution Shift	Performance degradation	Structural robustness if causal model

		valid
Feature Importance	Reflects correlation strength	Reflects causal influence
Generalization Across Processing Conditions	Limited	High if causal assumptions hold
Role in Materials Design	Descriptive / predictive	Prescriptive / decision-support
Failure Mode	Spurious correlation collapse	Instrument invalidity or weak identification
Epistemic Status	Data-dependent	Mechanism-oriented

In summary, while correlation provides a useful starting point for hypothesis generation, it is insufficient for the interventional demands of materials design. The transition from correlation to causation requires deliberate handling of both observed and unobserved confounders—the focus of the remainder of this framework.

Unobserved Confounders in Materials Science

Unobserved Confounder — A variable that affects both microstructure and property but is not measured or included in the dataset.

Unobserved confounders are ubiquitous in materials datasets. Historical processing records are frequently incomplete; laboratories rarely log every thermal cycle, atmosphere exposure, or contamination event. Trace impurities below the sensitivity of routine characterization—such as oxygen or carbon at parts-per-billion levels—can dramatically alter grain boundary mobility or phase stability yet remain invisible in standard microstructural descriptors. Measurement conditions introduce further confounding: slight variations in indentation load, electron-beam dosage, or environmental humidity can change both apparent microstructure (via imaging artifacts) and measured property (via testing artifacts). Proprietary or batch-specific details in industrial data add another layer of hidden influence.

These unobserved factors create backdoor paths in the causal graph: arrows from the confounder into both microstructure and property. Standard machine learning adjusts only for observed covariates. When the confounder is omitted, the estimated relationship between microstructure and property absorbs the confounder's effect, producing biased coefficients or feature importances. A model may conclude that increasing precipitate density raises hardness, yet the true driver is an unrecorded aging temperature that simultaneously promotes precipitation and matrix strengthening. Deploying the model on material processed at a different temperature yields erroneous predictions [13, 14].

The consequences compound in high-stakes applications. In aerospace alloys, an unobserved impurity level can shift fatigue life by orders of magnitude; correlation-based models trained without this variable will misattribute performance differences to microstructural features alone. In additive manufacturing, layer-by-layer thermal history is rarely fully captured, yet it governs residual stress, porosity, and final properties. When models ignore these unobserved histories, they cannot generalize from one build plate to another [4, 15].

Domain knowledge helps identify candidate confounders, but many remain unmeasurable with current instrumentation. Sensitivity analysis therefore becomes essential: one must ask how strong an unobserved confounder would need to be to overturn a claimed causal effect. Without such analysis, materials scientists risk overconfidence in correlational findings. Recent causal inference literature emphasizes that even modest unobserved confounding can reverse policy-relevant conclusions; the same principle applies to microstructure–property decisions [13, 16, 17].

In short, unobserved confounding is not an edge case but the default condition in real-world materials data. Acknowledging and addressing it is prerequisite to moving beyond correlational machine learning toward causal, actionable microstructure–property mapping.

Instrumental Variables for Materials

Instrumental Variable (IV) — A variable that affects microstructure but does not directly affect property, except through microstructure.

A valid instrument satisfies three conditions. First, relevance: it must strongly predict variation in the microstructure feature of interest. Second, exclusion: it must influence the property only via its effect on microstructure, with no direct path to the outcome. Third, independence: it must be uncorrelated with unobserved confounders. When these assumptions hold, the instrument isolates exogenous variation in microstructure that is free of confounding bias.

Potential instruments in materials science include randomized or as-if-random fluctuations in processing parameters. For example, small uncontrolled variations in furnace temperature during a controlled experiment may affect grain size without directly altering final strength except through grain refinement. Geological source variations in naturally sourced minerals can serve as instruments for impurity levels in ceramics. Time of production batch may act as an instrument when confounders are known to be time-invariant. In some cases, measurement instrument settings deliberately randomized across samples can provide exogenous variation in apparent microstructural descriptors [18, 19].

Instrumental variable estimation proceeds in two stages. In the first stage, microstructure is regressed on the instrument (and any observed covariates). The predicted microstructure values capture only the variation attributable to the instrument. In the second stage, property is regressed on these predicted values. The resulting coefficient estimates the causal effect of microstructure on property, purged of unobserved confounding. This approach has proven powerful in economics and epidemiology for recovering causal effects under confounding and is directly transferable to materials contexts where true randomization is rare but natural experiments exist [18, 20].

Limitations remain. True random assignment of processing parameters is uncommon outside designed experiments. The exclusion restriction is often plausible but difficult to verify; an instrument may have subtle direct effects on property through unmodeled pathways. Weak instruments—those only marginally predictive of microstructure—amplify bias and require large samples. Nonlinear relationships between instrument, microstructure, and property further complicate standard linear IV methods, calling for extensions suited to high-dimensional microstructural data [19].

Despite these challenges, instrumental variables offer a practical route to causal identification when full confounder measurement is impossible. By leveraging domain knowledge to propose and justify candidate instruments, materials scientists can begin to disentangle causation from correlation even in the presence of substantial unobserved confounding.

A Causal Framework for Microstructure–Property Mapping

The specification of a causal graph constitutes the foundational step, wherein microstructure features (M), target property (Y), observed confounders, unobserved confounders, and candidate instruments (Z) are defined and arranged in a directed acyclic graph according to hypothesized causal directions grounded in physical understanding. This representation explicitly delineates all backdoor paths from M to Y that require blocking for causal identification.

Domain expertise then guides the identification of potential confounders, distinguishing those amenable to measurement and adjustment from those that remain unobservable. Sensitivity analysis further quantifies the magnitude an unobserved confounder would need to exert to overturn qualitative conclusions, thereby anchoring the approach in materials-specific insight while transparently acknowledging inherent data limitations [13].

When a valid instrument is available, instrumental variable estimation isolates the causal pathway through two-stage procedures: the first stage predicts microstructure from the instrument, and the second recovers the effect on the

property, thereby mitigating bias from unobserved confounders and yielding an estimate of the direct causal influence of microstructure on property [18, 19].

The resulting causal model enables counterfactual queries of the form “what would property Y be if microstructure M were set to a new value while holding confounders fixed,” directly informing design decisions by distinguishing actionable microstructural interventions from mere associations.

Validation proceeds through sensitivity checks, comparisons of causal versus correlational predictions on held-out data subject to known confounder shifts, and qualitative alignment with physical principles; where interventional data emerge, they provide gold-standard benchmarks for assessing the framework’s assumptions.

Figure 1 illustrates the causal identification architecture showing how unobserved confounders create spurious microstructure–property correlations and how instrumental variable–based causal inference isolates the true effect of microstructure on property.

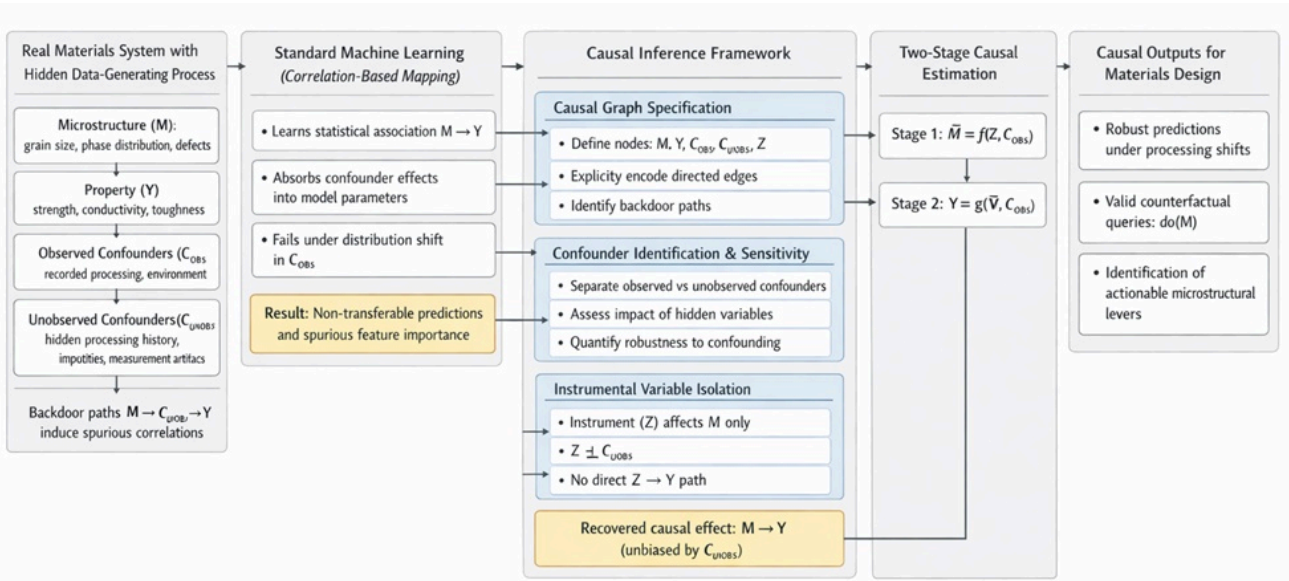


Figure 1. Causal Identification Architecture for Microstructure–Property Mapping Under Unobserved Confounding

Unobserved confounders (C_{unobs}) send arrows to both microstructure (M) and property (Y), creating a backdoor path. Observed confounders (C_{obs}) follow the same pattern but can be statistically adjusted. The instrumental variable (Z) points only to M, providing an exogenous source of variation. Microstructure in turn points to property. The graph visually identifies which paths must be blocked or isolated to recover the causal effect of M on Y.

Together, these five components form a coherent conceptual framework that transforms microstructure–property mapping from a correlational exercise into a causally grounded enterprise suitable for robust design under realistic data limitations.

Table 2 formalizes the roles and identification conditions of each variable type within the causal framework, clarifying how causal effects are recovered under unobserved confounding.

Table 2. Identification Roles of Variables in the Microstructure–Property Causal Framework

Variable Type	Symbol	Functional Role in Framework	Identifiability Status	Key Assumptions	Failure Risk

Microstructure	M	Treatment variable; mediates causal effect on Y	Target of estimation	Manipulability (conceptual intervention)	Measurement error, representation bias
Property	Y	Outcome variable	Observable	Correct measurement and alignment	Noise, testing artifacts
Observed Confounders	C_{obs}	Block backdoor paths via adjustment	Measurable	Sufficient measurement coverage	Omitted variable bias if incomplete
Unobserved Confounders	C_{unobs}	Source of bias in M – Y relationship	Not directly observable	Bounded influence (via sensitivity analysis)	Severe bias if strong and ignored
Instrumental Variable	Z	Provides exogenous variation in M	Identifiable if valid	Relevance, exclusion, independence	Weak instrument, violation of exclusion
Predicted Microstructure	$\hat{M}$	Purged variation used for causal estimation	Derived	Correct first-stage specification	Model misspecification
Counterfactual State	$do(M)$	Hypothetical intervention on M	Inferred	Structural correctness of causal graph	Invalid causal structure

Relation to Existing Methods

Standard machine learning approaches in computational materials engineering focus on learning statistical correlations between microstructural descriptors and properties. These models optimize for accurate prediction within the observed data distribution but remain silent on whether altering a microstructural feature will causally change the property. The causal inference framework proposed here shifts the objective from correlation to intervention. It equips researchers to answer design questions of the form “what happens to the property if the microstructure is changed while holding confounders fixed?” This distinction is essential for materials design, where decisions require reliable counterfactual predictions rather than pattern matching [2].

The framework also relates closely to domain adaptation techniques. Domain adaptation methods attempt to adjust models when the distribution of microstructural inputs or processing conditions shifts between source and target domains. They reweight samples or learn invariant representations to improve transfer performance. However, domain adaptation treats the shift as a black-box statistical mismatch without identifying its root cause. In contrast, the causal framework explicitly attributes distribution shifts to changes in unobserved confounders and uses instrumental variable strategies to isolate invariant causal pathways. As a result, causal models often transfer more reliably across processing histories or material classes because they capture the underlying mechanisms rather than merely compensating for statistical differences [9, 10].

Similarly, the framework advances beyond traditional extrapolation in microstructure–property modeling. Many existing machine learning studies emphasize improving out-of-distribution generalization through regularization or physics-informed constraints. Yet without a causal lens, extrapolation remains fragile because models may latch onto spurious correlations that do not persist under new conditions. Causal models, by removing confounding bias and focusing on direct effects, provide a principled basis for extrapolation. They identify relationships that remain stable even when unobserved processing variables change, offering greater confidence when predicting properties for microstructures generated under novel synthesis routes [12, 15].

Finally, the framework enhances interpretability in a deeper sense than post-hoc explainability tools. Feature importance scores or attention maps in standard models highlight which microstructural descriptors correlate with properties, but they do not distinguish causal drivers from confounded associations. The causal graph specification and confounder identification steps produce interpretable directed relationships that explain why certain microstructural changes affect properties. This level of mechanistic insight supports hypothesis generation and guides experimental validation far more effectively than correlational explanations alone [2, 11, 21].

Overall, the causal framework complements and extends existing methods rather than replacing them. It builds on domain adaptation and extrapolation techniques by supplying the missing causal foundation, thereby enabling more robust, transferable, and design-oriented microstructure–property mapping in data-driven materials engineering.

Challenges and Open Questions

Despite its promise, implementing the causal framework for microstructure–property mapping faces several practical and theoretical challenges. The first major challenge is discovering causal structure from data. Causal discovery algorithms can infer directed relationships among variables, yet they typically require large sample sizes and strong assumptions about the data-generating process. Microstructural datasets in materials science are often limited in scale, and the high-dimensional nature of image-based or graph-based descriptors further complicates automated discovery. Integrating domain knowledge with these algorithms remains an open question that must be addressed for the framework to scale [22–24].

A second challenge lies in finding valid instrumental variables. Materials processing rarely includes true randomization, making it difficult to identify variables that affect microstructure without directly influencing properties or correlating with unobserved confounders. Researchers must rely on domain expertise to propose candidate instruments, such as natural batch-to-batch variations or deliberately randomized characterization settings. Justifying the exclusion and independence assumptions for these instruments in complex material systems continues to demand careful theoretical and empirical scrutiny [18].

Third, standard instrumental variable methods assume linear relationships, yet microstructural data are inherently nonlinear and high-dimensional. Grain morphologies, defect networks, and phase distributions cannot be adequately captured by simple scalar features. Developing nonlinear and high-dimensional extensions of instrumental variable estimation suitable for image or graph data represents a critical open avenue for future research [4].

Fourth, unobserved confounding often manifests in time-series form. Processing history is sequential: each thermal or mechanical step builds on the previous one, accumulating hidden effects that influence both evolving microstructure and final properties. Adapting the framework to handle time-series causal inference while respecting the temporal ordering of processing events is essential for additive manufacturing and other dynamic processes [23, 25, 26].

Fifth, validating causal claims without ground-truth interventions remains inherently difficult. Because full randomization of microstructural features is often impractical, the framework must incorporate rigorous sensitivity analysis and robustness checks. Comparing causal predictions against correlational baselines on held-out data with documented confounder shifts provides one practical validation route, but systematic benchmarks for causal performance in materials contexts are still lacking [13, 14, 17].

These challenges highlight fertile ground for interdisciplinary collaboration between causal inference experts and materials scientists. Addressing them will determine whether the proposed framework transitions from conceptual guidance to routine practice in computational materials engineering.

Implications for Materials Design

The causal framework carries direct implications for how experimental campaigns, modeling workflows, and discovery pipelines are structured in materials science. For experimental design, it encourages deliberate recording of potential confounders whenever feasible—such as exact thermal profiles, cooling rates, batch identifiers, and environmental conditions. Where complete measurement is impossible, the framework advocates designing studies that incorporate natural or controlled variation suitable for instrumental variable analysis. Randomizing selected processing parameters within feasible limits can generate the exogenous variation needed for causal identification, transforming routine experimentation into a source of interventional insight [18–20].

In machine learning model development, the framework calls for a shift from purely predictive objectives to causal estimation. Developers should embed causal graph specification early in the pipeline, explicitly test for unobserved confounding through sensitivity analysis, and report not only predictive accuracy but also robustness to hypothetical confounder shifts. This practice discourages over-reliance on correlation-based benchmarks and promotes models that generalize under distribution changes. By prioritizing instrumental variable estimation and counterfactual reasoning, modelers can produce outputs that directly support decision-making rather than merely describing past observations [2, 9, 11].

For materials discovery more broadly, the framework enables a new paradigm of counterfactual design. Instead of screening large libraries of microstructures for correlated property improvements, researchers can query the causal model to predict the property change resulting from targeted microstructural interventions while holding processing confounders fixed. This capability distinguishes actionable levers—features that can be engineered to improve performance—from mere correlates that offer no reliable design pathway. In high-entropy alloys or additively manufactured components, for instance, the framework can clarify whether refining grain size or adjusting texture will causally enhance toughness under realistic production variability [12, 15, 27].

Collectively, these implications move materials design from trial-and-error guided by correlation toward rational, mechanism-based optimization. The framework equips the community to make confident statements about how microstructural changes will translate into property gains even when processing conditions evolve, ultimately accelerating the development of next-generation materials with precisely tailored performance.

Toward Causal Materials Science

The vision articulated here is a causal materials science in which models capture invariant mechanisms rather than transient statistical associations. Such models would remain reliable under distribution shifts caused by new processing routes, different laboratories, or scaled production, and would directly support counterfactual design questions essential to innovation. By addressing unobserved confounding through explicit graphs, instrumental variables, and sensitivity analysis, the field can move beyond brittle correlational predictions toward robust, transferable understanding [2, 3, 7].

A practical roadmap supports this transition. In the short term, researchers can apply instrumental variable strategies to problems that already contain natural experiments—such as batch-to-batch variation in industrial datasets or randomized characterization protocols. These focused applications will demonstrate tangible gains in causal reliability for specific microstructure–property pairs. In the medium term, the community should invest in nonlinear causal inference methods capable of handling high-dimensional microstructural images and graphs. Advances in causal representation learning can be adapted to extract causally meaningful features from electron micrographs or tomography data [25, 28, 29]. In the long term, the goal is seamless integration of data-driven causal discovery with established physical knowledge, creating hybrid models that respect thermodynamic and kinetic constraints while learning from observational data [4, 22, 23].

Success will be measured by concrete criteria. A mature causal model should predict the effect of a microstructural intervention on a target property with substantially lower error than correlation-based baselines when tested under known

confounder shifts. Where standard models exhibit large degradation in performance, causal counterparts should maintain stability, demonstrating that the framework has isolated true mechanisms. Achieving this level of robustness across multiple material classes would mark a genuine paradigm shift in computational and data-driven materials engineering.

Ultimately, adopting the causal framework does not require discarding existing machine learning tools; it reframes their use around interventional questions. The result will be a materials science better equipped to design rather than merely describe, to extrapolate rather than interpolate, and to innovate with confidence under the realistic constraints of unobserved confounding.

Conclusion

Microstructure–property mapping requires causal inference, not just correlation. Unobserved confounders such as processing history, trace impurities, and measurement artifacts systematically bias standard machine learning models, causing predictions to fail under even modest distribution shifts. The conceptual framework presented here addresses this limitation through five integrated components: causal graph specification, confounder identification, instrumental variable estimation, counterfactual reasoning, and validation strategy. By making these elements explicit, the framework enables robust predictions that remain valid when processing conditions change and supports the interventional questions at the heart of materials design.

The framework relates to and extends existing methods in domain adaptation, extrapolation, and interpretability while highlighting distinct advantages for causal reasoning. It also surfaces key challenges—causal discovery at scale, identification of valid instruments, handling of nonlinearity and time-series data, and rigorous validation—that define productive directions for future research. Despite these obstacles, the payoff is substantial: materials models that capture invariant mechanisms, generalize reliably, and directly inform rational microstructural engineering.

The call to the community is clear. Computational materials scientists should begin incorporating causal graph thinking and sensitivity analysis into their workflows, seek opportunities for instrumental variable analysis in both laboratory and industrial datasets, and prioritize counterfactual predictions alongside traditional performance metrics. As the volume of microstructural data continues to grow, the transition from correlation to causation will determine whether data-driven approaches fulfill their promise of accelerating discovery or remain limited to pattern recognition within narrow domains.

Adopting this causal perspective offers a pathway toward a more reliable, transferable, and design-oriented materials science—one capable of delivering the next generation of high-performance materials under the realistic conditions of unobserved confounding.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 24 Nov 2024

Revised: 29 Jan 2025

Accepted: 09 May 2025

Published online: 18 July 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

Li Z, Yoon J, Zhang R, Rajabipour F, Srubar WV 3rd, Dabo I, et al. Machine learning in concrete science: Applications, challenges, and best practices. *NPJ Comput Mater.* 2022;8(1):127.
<https://doi.org/10.1038/s41524-022-00810-x>.

Pilania G. Machine learning in materials science: From explainable predictions to autonomous design. *Comput Mater Sci.* 2021;193:110360.
<https://doi.org/10.1016/j.commatsci.2021.110360>.

Ziatdinov M, Nelson CT, Zhang X, Vasudevan RK, Eliseev E, Morozovska AN, et al. Causal analysis of competing atomistic mechanisms in ferroelectric materials from high-resolution scanning transmission electron microscopy data. *NPJ Comput Mater.* 2020;6(1):127.
<https://doi.org/10.1038/s41524-020-00396-2>.

Sun X, Bahmani B, Vlassis NN, Sun W, Xu Y. Data-driven discovery of interpretable causal relations for deep learning material laws with uncertainty propagation. *Granul Matter.* 2022;24(1):1.
<https://doi.org/10.1007/s10035-021-01137-y>.

Tian X, Bai J, Gong X, Lei R, Yan Y. Leveraging causal discovery and causal inference to improve interpretability of Charpy impact toughness. *Comput Mater Sci.* 2025;259:114164.
<https://doi.org/10.1016/j.commatsci.2025.114164>.

Ting JY, Barnard AS. Data-driven causal inference of process-structure relationships in nanocatalysis. *Curr Opin Chem Eng.* 2022;36:100818.
<https://doi.org/10.1016/j.coche.2022.100818>.

Liu R, Hao J, Li J, Wang S, Liu H, Zhou Z, et al. Causal inference machine learning leads original experimental discovery in CdSe/CdS core/shell nanoparticles. *J Phys Chem Lett.* 2020;11(17):7232-8.
<https://doi.org/10.1021/acs.jpclett.0c02115>.

Seike K, Manaka H, Miura Y. Causal inference in statistics insights into stress-induced ferroelectric states in SrTiO₃: Disentangling piezoelectric and flexoelectric effects from birefringence images. *Sci Technol Adv Mater Methods.* 2025;5(1):2503698.
<https://doi.org/10.1080/27660400.2025.2503698>.

Hu J, Liu D, Fu N, Dong R. Realistic material property prediction using domain adaptation based machine learning. *Digit Discov.* 2024;3(2):300-12.
<https://doi.org/10.1039/D3DD00162H>.

Soares E, Shirasuna VY, Brazil EV, Aquino Gutierrez KF, Cerqueira R, Zubarev D, et al. Causality-driven feature selection and domain adaptation for enhancing chemical foundation models in downstream tasks. *Mach Learn Sci Technol.* 2025;6(1):015017.
<https://doi.org/10.1088/2632-2153/adabb1>.

Kamarthi A, Kaliyamoorthy B. Causal inference-informed explainable selective weighted ensemble for predicting ultra-high-performance concrete compressive strength. *Knowl Based Syst.* 2025;330:114770.
<https://doi.org/10.1016/j.knosys.2025.114770>.

Zhao YM, Zhang JY, Liaw PK, Yang T. Machine learning-based computational design methods for high-entropy alloys. *High Entropy Alloys Mater.* 2025;3(1):41-100.
<https://doi.org/10.1007/s44210-025-00055-5>.

Loh WW. Causal inference with unobserved confounding: Leveraging negative control outcomes using lavaan. *Multivar Behav Res.* 2025;60(5):1029-41.
<https://doi.org/10.1080/00273171.2025.2507742>.

Kallus N, Zhou A. Minimax-optimal policy learning under unobserved confounding. *Manag Sci.* 2021;67(5):2870-90.
<https://doi.org/10.1287/mnsc.2020.3699>.

Zang Y, Koutsourelakis PS. PSP-GEN: Stochastic inversion of the process-structure-property chain in materials design through deep, generative probabilistic modeling. *Acta Mater.* 2025;284:120600.
<https://doi.org/10.1016/j.actamat.2024.120600>.

Kong D, Yang S, Wang L. Identifiability of causal effects with multiple causes and a binary outcome. *Biometrika.* 2022;109(1):265-72.
<https://doi.org/10.1093/biomet/asab016>.

Miller LC, Shaikh SJ, Jeong DC, Wang L, Gillig TK, Godoy CG, et al. Causal inference in generalizable environments: Systematic representative design. *Psychol Inq.* 2019;30(4):173-202.
<https://doi.org/10.1080/1047840X.2019.1693866>.

Hahn J, Ridder G. Instrumental variable estimation of nonlinear models with nonclassical measurement error using control variables. *J Econom.* 2017;200(2):238-50.
<https://doi.org/10.1016/j.jeconom.2017.06.008>.

Plagborg-Møller M, Wolf CK. Instrumental variable identification of dynamic variance decompositions. *J Polit Econ.* 2022;130(8):2164-202.
<https://doi.org/10.1086/720141>.

Xie Z, Chen Z, Jiang Y, Yao J, Zhang P, Lei H, et al. Causal relationships between epilepsy and the microstructure of the white matter: A Mendelian randomization study. *Medicine (Baltimore).* 2024;103(44):e40090.
<https://doi.org/10.1097/MD.00000000000040090>.

Chen D, Zhao H, He J, Pan Q, Zhao W. An causal XAI diagnostic model for breast cancer based on mammography reports. In: *Proceedings - 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2021 Dec 9-12; Virtual. Piscataway (NJ): IEEE; 2021. p. 3341-9.*
<https://doi.org/10.1109/BIBM52615.2021.9669648>.

Barakati K, Molak A, Nelson C, Zhang X, Takeuchi I, Kalinin SV. Causal discovery from data assisted by large language models. *Appl Phys Lett*. 2025;127(12):121904.
<https://doi.org/10.1063/5.0272287>.

Mogensen SW, Rathsmann K, Nilsson P. Causal discovery in a complex industrial system: A time series benchmark. In: Locatello F, Didelez V, editors. *Proceedings of the Third Conference on Causal Learning and Reasoning*. *Proc Mach Learn Res*. 2024;236:1218-36.

Shen X, Raghavan S, Przybelski SA, Lesnick TG, Ma S, Reid RI, et al. Causal structure discovery identifies risk factors and early brain markers related to evolution of white matter hyperintensities. *Neuroimage Clin*. 2022;35:103077.
<https://doi.org/10.1016/j.nicl.2022.103077>.

Cai S, Peng H, Liu R, Chen P. Causal-oriented representation learning for time-series forecasting based on the spatiotemporal information transformation. *Commun Phys*. 2025;8:242.
<https://doi.org/10.1038/s42005-025-02170-6>.

Castri L, Mghames S, Hanheide M, Bellotto N. CAnDOIT: Causal discovery with observational and interventional data from time series. *Adv Intell Syst*. 2024;6(12):2400181.
<https://doi.org/10.1002/aisy.202400181>.

Ni B, Glaser B, Taheri-Mousavi SM. End-to-end prediction and design of additively manufacturable alloys using a generative AlloyGPT model. *NPJ Comput Mater*. 2025;11(1):294.
<https://doi.org/10.1038/s41524-025-01768-2>.

Ahuja K, Mahajan D, Wang Y, Bengio Y. Interventional causal representation learning. In: Krause A, Brunskill E, Cho K, Engelhardt B, Sabato S, Scarlett J, editors. *Proceedings of the 40th International Conference on Machine Learning*. *Proc Mach Learn Res*. 2023;202:372-407.

Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: An overview. *IEEE Signal Process Mag*. 2018;35(1):53-65.
<https://doi.org/10.1109/MSP.2017.2765202>.