

ORIGINAL RESEARCH

Open access

Operationalizing "Material Similarity" for Domain Adaptation: A Definitional Boundary for Composition-Transfer Learning

Sanjay Kulkarni¹, Meenal Joshi^{1*}, Rohan Patil², Aniket Deshmukh²

Abstract

Material similarity functions as the implicit premise underlying domain adaptation in materials informatics, yet it remains largely undefined, rendering transfer claims difficult to evaluate and compare. This work establishes a task-dependent, operational definition of material similarity for composition-transfer learning, grounded in four orthogonal dimensions: composition, structure, property behavior, and local atomic environment. By examining current usage patterns, it reveals systematic ambiguities arising from single-dimension interpretations and demonstrates how these ambiguities contribute to inconsistent transfer outcomes. The proposed framework introduces threshold-based criteria and a weighted similarity score that jointly determine when transfer is justified, alongside boundary conditions that constrain successful adaptation. Gray zones and edge cases are analyzed to show how trade-offs across dimensions must be explicitly quantified rather than implicitly assumed. The resulting formulation repositions similarity as a measurable, task-specific precondition rather than a descriptive label, enabling pre-transfer validation and standardized reporting. This shift provides a conceptual foundation for reproducible domain adaptation, aligning materials machine learning with more rigorous principles of transferability and model generalization.

Keywords Materials informatics, Transfer learning, Domain adaptation, Material similarity, Compositional similarity, Chemical similarity

*Correspondence:

Meenal Joshi

meenal.joshi@gmail.com

¹ Department of Computational Materials Engineering, Faculty of Technology, Savitribai Phule Pune University, Pune, India

² Department of Materials Data Systems, Faculty of Engineering, IIT Bombay, Mumbai, India

Introduction

Domain adaptation and transfer learning have emerged as essential strategies in computational materials engineering, allowing models trained on well-characterized source domains to be repurposed for data-scarce target domains [1-7]. A representative claim in the literature reads: "Because material A and material B are similar, we transfer a model trained on A to B." Such statements appear in studies ranging from alloy design to crystal property prediction [8-10]. Yet the pivotal term—"similar"—is rarely defined. Does similarity refer to shared elemental constituents, identical crystal prototypes, overlapping property ranges, or analogous local atomic environments? The absence of a precise boundary renders the transfer hypothesis untestable and makes cross-paper comparisons impossible.

This article provides a definitional boundary analysis of "material similarity" tailored to domain adaptation in composition-transfer learning. It does not present new experiments, datasets, or performance benchmarks; instead, it offers a conceptual clarification grounded in existing scholarship. The analysis proceeds by first establishing why similarity

matters for the transfer hypothesis, then dissecting current usage patterns and their inherent confusions, mapping the four primary dimensions of material similarity, and finally articulating an operational definition together with minimum reporting requirements.

The motivation is practical as well as theoretical. As databases such as the Materials Project and high-throughput computational screening generate ever-larger composition spaces, researchers increasingly rely on transfer learning to extrapolate across chemically related but not identical systems [11, 12]. Bartók *et al.* demonstrated that unified machine-learning frameworks can model both molecules and materials within a single representational space [13], yet they stopped short of specifying similarity criteria for domain shifts. Similarly, Chen *et al.* introduced graph networks as a universal framework for molecules and crystals [14], highlighting the need for robust transfer mechanisms but leaving the notion of “similar materials” implicit. Batra *et al.* surveyed emerging materials intelligence ecosystems and noted the growing dependence on transfer learning, again without formalizing similarity [12]. More targeted work on domain adaptation explicitly examines when transfer succeeds, but treats similarity as a black-box precondition rather than an operational variable [15, 16].

The present boundary analysis therefore fills a critical gap. By restricting attention to composition-transfer learning—where the source and target differ primarily in elemental makeup or stoichiometry while sharing broader materials classes—this paper establishes definitional boundaries that are both task-dependent and measurable. The result is a framework that allows researchers to state with precision when a similarity claim justifies domain adaptation and when it does not. In doing so, it shifts the field from anecdotal assertions of similarity toward a reproducible, dimensionally explicit standard.

Why "Material Similarity" Matters for Domain Adaptation

The transfer hypothesis in materials machine learning rests on a single intuitive premise: a model trained on source domain *S* will generalize to target domain *T* provided *S* and *T* are “similar enough.” Without an operational definition of “similar enough,” however, the hypothesis cannot be falsified or validated systematically. Positive transfer occurs when the domain gap is small enough that fine-tuning or reweighting yields improved performance on *T*; negative transfer occurs when the gap is large and the transferred model degrades target-domain accuracy relative to training from scratch; neutral transfer occupies the ambiguous middle ground where outcomes are unpredictable [16, 17].

Work on chemical similarity in machine learning emphasizes that similarity judgments underpin every transfer decision, yet the community lacks consensus metrics [18–20]. Domain-adaptation studies further illustrate that transfer success is highly sensitive to the source–target relationship, yet do not supply criteria for deciding when that relationship is sufficiently close [15, 21, 22]. In the absence of such criteria, authors frequently assert similarity on the basis of superficial features—shared elements, for example—without evidence that those features are relevant to the prediction task.

This definitional vacuum has three practical consequences. First, it impedes cumulative science: one study’s “similar” materials may be another’s “dissimilar” ones, rendering meta-analyses unreliable. Second, it encourages over-optimistic reporting; negative-transfer cases are under-reported because the underlying similarity claim is never quantified and therefore never falsified. Third, it hinders benchmark design: without agreed similarity boundaries, it is impossible to construct controlled transfer tasks that isolate compositional effects from structural or property effects.

The gap is particularly acute for composition-transfer learning, where source and target share the same material class (alloys or inorganic crystals) but differ in elemental identity or stoichiometry. Here, similarity cannot be reduced to database provenance or broad chemical family membership. Reviews of materials intelligence ecosystems note the proliferation of transfer-learning applications yet call for better foundational concepts [12]. Likewise, unified modeling frameworks demonstrate that representations can span molecules and materials, but do not address how to decide when two points in that representation space are close enough for reliable transfer [13].

Material similarity therefore functions as the hidden linchpin of the entire domain-adaptation enterprise. An operational definition converts it from a rhetorical device into a testable precondition. By specifying which dimensions must align and at what thresholds, researchers can move from post-hoc rationalization (“the transfer worked, therefore the materials must have been similar”) to pre-hoc justification (“the materials satisfy criteria C1–C4, therefore transfer is justified”). This shift is essential if composition-transfer learning is to mature from an empirical practice into a principled methodology.

Current Usage and Confusions

Examination of recent literature reveals multiple, often implicit, operationalizations of “material similarity” that introduce systematic ambiguity. One recurrent interpretation reduces compositional similarity to elemental overlap, whereby alloys are deemed comparable solely because they share constituent species, particularly transition metals. Although studies on chemical similarity document this tendency, they emphasize that elemental presence alone neglects stoichiometry and electronic structure [18, 19, 23, 24], a limitation underscored by contrasts such as CoCrFeNi and CoO, which share cobalt yet diverge fundamentally in bonding and properties, undermining transfer learning for formation energies. A closely related assumption equates similarity with structural identity, inferring transferability from shared prototypes such as face-centered-cubic (FCC) arrangements. While graph-based frameworks demonstrate the utility of structural representations within consistent prototypes [14], they also require explicit accommodation of compositional variation, as evidenced by the divergent elastic and thermal responses of FCC aluminum and FCC nickel, where naive model transfer frequently induces negative transfer. Beyond these structural considerations, similarity is often inferred from overlapping property regimes, as in claims that materials are analogous because they exhibit wide band gaps; however, domain-adaptation studies indicate that convergence in observable properties does not ensure alignment of underlying trends [15, 21]. Attention to local environments introduces further nuance, with coordination numbers or polyhedral geometries frequently invoked as sufficient descriptors, yet unified ML frameworks demonstrate that such features remain strongly modulated by element-specific contexts [13]. Another, more implicit, criterion derives from shared database provenance, particularly in multi-property prediction settings, though this collapses under scrutiny given the chemical and physical heterogeneity within large repositories [11, 12, 25]. These heterogeneous usages are rarely articulated, and the absence of explicit prioritization renders similarity claims effectively incomparable, producing a fragmented literature in which successful transfer in one context coexists with negative transfer in another. Although prior work acknowledges this conceptual dispersion, it does not establish a unified boundary definition [15, 16, 18], motivating the present analysis to reinterpret these practices through orthogonal dimensions and to demonstrate the limitations of treating any single dimension in isolation for composition-transfer learning.

Dimensions of Material Similarity

Material similarity for domain adaptation comprises four interrelated yet conceptually distinct dimensions.

Table 1 consolidates the manuscript’s definitional contribution by distinguishing the four dimensions of material similarity, specifying what each dimension measures, and clarifying why none is sufficient in isolation for transfer justification.

Table 1. Four Dimensions of Material Similarity: Definitions, Observable Indicators, and Transfer Relevance

Dimension	Core definitional question	What is being compared?	Observable indicators / operational proxies	Why this dimension matters for transfer	Why it is insufficient alone	Typical failure mode if used alone
Compositional similarity	Do source and target contain comparable	Element sets, atomic fractions,	Element overlap ratio; stoichiometric	Governs whether the model	Shared elements do not guarantee	Superficial element overlap is

	elemental identities and stoichiometric proportions?	stoichiometric ranges, substitution patterns	distance; fractional composition vectors; compositional embedding distance	encounters chemically related input space and comparable elemental interactions	comparable bonding, phase behavior, or response surfaces	mistaken for transferability even when structure or bonding diverges
Structural similarity	Do source and target share comparable crystallographic organization?	Crystal prototypes, space groups, lattice relations, coordination-number distributions, phase class	Prototype match; symmetry match; lattice-parameter ratios; coordination-distribution overlap	Important when target properties depend strongly on long-range order and crystallographic constraints	Identical prototype does not ensure similar energetics, electronic structure, or composition-response mapping	Prototype identity is over-interpreted despite major compositional or property mismatch
Property similarity	Do source and target occupy similar task-relevant property regimes?	Property distributions, ranges, gradients, trends with composition or structure	Range overlap; property-distribution distance; slope similarity; rank-order consistency	Directly links similarity assessment to the actual prediction task rather than generic material resemblance	Similar scalar outputs may arise from very different underlying mechanisms or representations	Models transfer between domains with overlapping outputs but incompatible generative structure
Local-environment similarity	Do source and target exhibit comparable atom-centered neighborhoods?	Bond lengths, neighbor shells, local coordination geometries, short-range order	Bond-length distribution overlap; local-descriptor distance; coordination-geometry frequency; environment embeddings	Captures short-range physics often missed by global composition or prototype labels	Similar local motifs may still sit inside very different global structural or property landscapes	Over-reliance on local descriptors hides larger-scale mismatch across phases or property manifolds

Each dimension captures a different facet of the source–target relationship and carries task-dependent weight.

Compositional similarity is defined through element identity and stoichiometric proportion, interrogating not only which elements are shared but also whether their atomic fractions occupy comparable regimes. Similarity is therefore elevated when source and target compositions preserve both elemental membership and relative proportions, whereas it degrades under the introduction of novel species or substantial concentration disparities. Structural similarity extends this logic to crystallographic organization, encompassing space-group identity, prototype correspondence, lattice-parameter ratios, and coordination-number distributions; however, even within a shared prototype, deviations such as lattice distortion necessitate quantitative treatment rather than categorical equivalence. Property similarity reframes the notion of correspondence in terms of task-relevant observables, evaluating the extent to which source and target distributions and their underlying gradients align, whether in formation-energy landscapes or electronic characteristics such as band gaps and density of states [26]. This perspective foregrounds the relationship between observable outputs and their

compositional dependencies, rather than treating property overlap as intrinsically sufficient. Local-environment similarity further refines this hierarchy by emphasizing atom-centered descriptors, including bond-length distributions, coordination geometries, and element-specific neighbor shells, thereby capturing short-range order inaccessible to global compositional or structural descriptors. Under these conditions, the relative salience of each dimension becomes inherently task-dependent: formation-energy prediction is governed primarily by compositional and local-environment alignment due to the dominance of nearest-neighbor interactions, whereas elastic-constant prediction is more strongly conditioned by structural correspondence, reflecting the influence of long-range order on mechanical response. This task-specific weighting is implicitly acknowledged in the domain-adaptation study by M., while the graph-network framework of Chen *et al.* offers a unified representational basis through which all four dimensions may be encoded within a shared embedding space.

This multi-dimensional representation clarifies why single-dimension claims are misleading and why an operational definition must integrate all four axes with explicit, task-specific weighting.

Figure 1 presents the manuscript's central contribution as a hierarchical decision framework showing how task specification, four-dimensional similarity assessment, threshold testing, and boundary-condition screening jointly determine whether composition-transfer learning is justified.

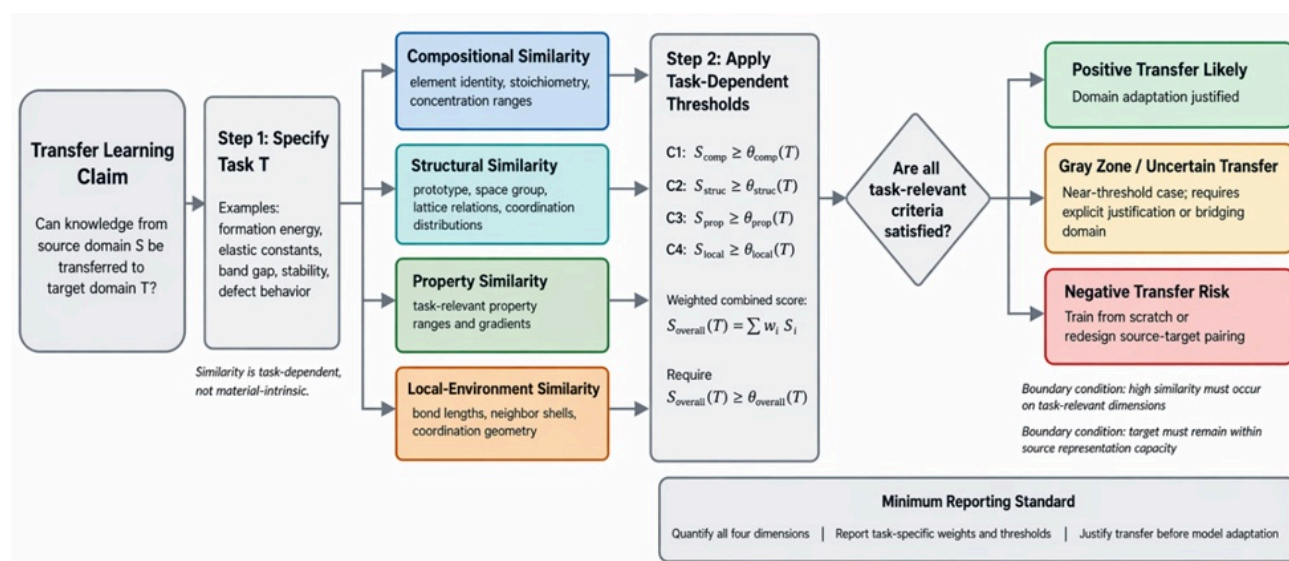


Figure 1. A Hierarchical Decision Framework for Operationalizing Material Similarity in Composition-Transfer Learning

Proposed Operational Definition

Two materials or material families, denoted A and B, are considered similar for transfer learning on a task T only when four conditions are jointly satisfied: compositional similarity surpasses a task-dependent threshold $\theta_{comp}(T)$; structural similarity exceeds $\theta_{struct}(T)$; property similarity across all task-relevant observables meets $\theta_{prop}(T)$; and a combined similarity score, constructed as a weighted aggregation of the four dimensions with task-specific weights, exceeds $\theta_{overall}(T)$. This formulation is explicitly task-contingent rather than material-intrinsic, such that correspondence in one predictive context does not imply transferability in another, reflecting the shifting dominance of underlying mechanisms. The thresholds θ are therefore not universal constants but must be determined empirically for each task family through controlled variation of source–target distances [16, 17].

This definition imposes immediate operational constraints. Any assertion of similarity must specify the precise dimensions evaluated and the numerical thresholds employed, and must be anchored to a clearly defined task T; decontextualized

claims are thereby excluded. When multiple predictive objectives are considered, each requires an independent assessment, as cross-task generalization cannot be assumed. Accordingly, a minimum reporting standard emerges in which studies invoking material similarity for domain adaptation provide quantitative scores across all four dimensions, justify the selected thresholds, and identify the dimensions most heavily weighted for the task at hand. These requirements extend existing calls for transparency in domain adaptation and refine prior recommendations on chemical similarity and transfer-learning diagnostics [15, 16, 18].

Enforcing this framework transforms similarity from an implicit assumption into a testable condition. Evaluation shifts toward explicit interrogation of which dimensions were quantified, how thresholds were defined, and whether they were satisfied, enabling both critical assessment and prospective decision-making. In practice, this reframing allows practitioners to assess transfer viability a priori, reducing reliance on post hoc identification of negative transfer. The boundary articulated here thus recasts material similarity as a rigorous prerequisite for composition-transfer learning rather than a descriptive convenience.

Boundary Conditions for Transfer Success

The operational definition of material similarity delineates necessary yet insufficient conditions for effective domain adaptation, as successful composition-transfer learning depends on a set of boundary constraints that regulate the transition from positive to neutral or negative transfer. Central to this framework is the requirement that similarity aligns with task-relevant dimensions: when predictive performance is governed by compositional factors, as in formation-energy or stability estimation, compositional correspondence must exceed $\theta_{comp(T)}$, a condition whose violation has been shown to precipitate rapid degradation in transfer efficacy despite superficial agreement elsewhere [15, 21]. This dependence shifts in structure-dominated tasks, such as elastic-constant prediction, where structural alignment assumes a gating role. At the same time, divergence along task-irrelevant dimensions does not inherently disrupt transfer, provided such variation remains subordinate within the learned feature hierarchy; graph-network architectures, for instance, accommodate this tolerance by attenuating non-critical features during fine-tuning [14].

Beyond dimensional alignment, the residual domain gap must remain within the adaptive capacity of the model, since even nominal compliance with similarity criteria can be undermined by pronounced shifts in property gradients that exceed architectural flexibility, rendering transferred parameters ineffective or detrimental [6, 16]. This constraint interacts directly with data availability, as limited source datasets fail to encode stable compositional mappings, leading to unreliable transfer even between closely aligned materials [8, 9]. A further limitation arises when the target domain extends beyond the representational scope of the source, particularly through the introduction of previously unseen local environments or structural prototypes, where extrapolation becomes inherently unstable despite advances in unified modeling frameworks [13]. Under these conditions, transfer learning succeeds primarily when alignment is concentrated in the dominant task-relevant dimensions and the target remains embedded within the learned manifold, thereby minimizing domain discrepancy and facilitating efficient fine-tuning. Deviations from these constraints necessitate alternative strategies, including training from first principles or introducing intermediate domains, thereby recasting the operational definition as a predictive criterion rather than a retrospective justification.

Table 2 translates the proposed definition into an actionable decision matrix by linking similarity configurations and boundary conditions to expected transfer outcomes and appropriate modeling choices.

Table 2. Decision Matrix for Composition-Transfer Learning: Boundary Conditions, Expected Outcomes, and Recommended Research Actions

Scenario	Similarity profile across four dimensions	Boundary-condition status	Expected transfer outcome	Interpretation	Recommended research action

A. Strongly justified transfer	High similarity on all task-relevant dimensions; weighted score clearly exceeds overall threshold	All key conditions satisfied	Positive transfer likely	Source-target gap is narrow enough for transfer to function as meaningful prior knowledge	Proceed with domain adaptation; report all four scores, task weights, and thresholds explicitly
B. Selectively justified transfer	High similarity on the two or three dimensions most important for task T; modest mismatch on clearly secondary dimensions	Conditions mostly satisfied; residual gap manageable	Positive or neutral transfer plausible	Transfer may work when non-dominant dimensions do not control the prediction function	Proceed cautiously; justify weighting logic and evaluate against from-scratch baseline
C. Near-threshold gray zone	Mixed profile; one major dimension near threshold or multiple dimensions only marginally acceptable	Conditions ambiguous	Uncertain / unstable transfer	Transfer claim is not yet falsified, but not adequately justified either	Use bridging domain, stronger diagnostics, ablation tests, and explicit sensitivity analysis before claiming success
D. Hidden mismatch despite surface similarity	High score on one salient dimension only (for example, same elements or same prototype) but low scores elsewhere	Critical conditions violated	Negative transfer risk high	Apparent similarity is rhetorical rather than operational because alignment is one-dimensional	Do not justify transfer on single-dimension grounds; either redesign pairing or train from scratch
E. Representation-capacity violation	Moderate similarity scores but target introduces novel elements, local environments, or prototypes absent from source manifold	Condition on source representation capacity violated	Negative transfer likely	Even acceptable scalar similarity cannot rescue extrapolation beyond learned representation support	Expand source coverage, add intermediate source domains, or redesign representation before transfer
F. Data-volume failure	Similarity is high, but source data are sparse or poorly distributed	Source sufficiency condition violated	Neutral or negative transfer possible	Similarity cannot compensate for weak source-domain learning	Increase source-domain volume/coverage before transfer claims are made
G. Task-misaligned transfer	Materials appear globally similar, but the highest-scoring dimensions are not the ones most relevant to task T	Task dependence violated	Neutral or negative transfer likely	Similarity has been assessed, but on the wrong basis for the stated prediction objective	Recompute similarity using task-specific weighting and redefine transfer rationale

Boundary Cases and Gray Zones

Even with an operational definition and explicit boundary conditions, regions of ambiguity persist where similarity scores approach threshold values or vary across dimensions, necessitating careful interpretation of borderline cases. When identical elements appear in differing stoichiometries, as in TiO_2 (rutile prototype) and TiO (rock-salt structure), compositional overlap remains partial while structural alignment collapses and local environments diverge, producing a discontinuous property landscape in which formation-energy transfer fails as the domain gap exceeds $\theta_{\text{struc}(T)}$ and $\theta_{\text{local}(T)}$. A contrasting scenario arises when distinct chemistries share a common prototype: MgO and NaCl exhibit negligible compositional overlap yet converge structurally, enabling transfer in elasticity-dominated tasks where coordination geometry governs response and structural similarity outweighs elemental disparity. This asymmetry is further exposed in systems such as body-centered-cubic (BCC) iron and face-centered-cubic (FCC) iron, where identical composition coexists with structural divergence; here, transfer may remain viable for magnetic-moment prediction because the governing mechanism is predominantly element-driven, rendering structural variation secondary.

Such cases foreground the non-universality of similarity thresholds, which must be empirically calibrated within each task domain, as even substantial compositional overlap may prove insufficient under certain conditions, a sensitivity implicitly demonstrated in alloy-design workflows where minor stoichiometric perturbations shift systems across similarity boundaries [9]. The interplay between dimensions becomes particularly evident when compensatory effects emerge, allowing strong alignment in structural or local-environment descriptors to offset compositional mismatch, or vice versa, though the extent of such trade-offs remains inherently task-specific. Frameworks such as the AtomSets hierarchy developed by Chen *et al.* [27] provide a means to formalize these interactions, yet do not prescribe universal weighting schemes. These gray zones therefore do not undermine the framework but instead delimit the space within which judgment operates, requiring explicit documentation of dimensional trade-offs and justification of aggregate similarity relative to defined thresholds. In this way, similarity is recast from a binary classification into a structured, reproducible landscape grounded in quantifiable criteria.

Relation to Other Concepts

Material similarity for domain adaptation is related to but distinct from several established concepts in machine learning and materials informatics.

It differs from domain-discrepancy measures such as Maximum Mean Discrepancy or adversarial domain alignment. Those metrics quantify distributional shift after the fact using feature embeddings; material similarity, by contrast, is a prior assessment performed before any target data are seen. Hu *et al.* [1] explores when transfer works precisely by probing this distinction, showing that low discrepancy alone does not guarantee success unless the underlying material similarity criteria are also satisfied [28].

Material similarity also extends—but is not identical to—chemical similarity as operationalized for molecules. Fingerprint-based metrics (Tanimoto coefficients, molecular graphs) are well-developed for organic compounds, yet crystals introduce periodic boundary conditions and stoichiometry that fingerprints do not capture. Axelrod *et al.* [11] review chemical similarity in materials machine learning and notes the gap; the four-dimensional framework proposed here bridges that gap by incorporating structural and local-environment axes absent from purely molecular definitions.

Finally, material similarity is a conceptual predecessor to data-driven transfer-learning success predictors such as the H-score or task-similarity metrics. Those predictors evaluate transfer potential after source and target data are embedded; material similarity supplies an interpretable, human-readable prior that explains why certain embeddings succeed or fail. Kong *et al.* [4] and Choudhary *et al.* [29] demonstrate the value of such priors in multi-property prediction, yet they treat similarity as an implicit input rather than an explicitly defined boundary.

By positioning material similarity at the intersection of these concepts, the present definition provides a missing foundational layer. It supplies the “why” behind observed discrepancies and success predictors, enabling researchers to diagnose transfer failures at the materials level rather than the model level alone.

Implications for Transfer Learning Practice

Adopting the operational definition carries immediate consequences for three stakeholder groups.

For researchers reporting transfer learning, the requirements are clear: cease claiming “the materials are similar” without quantification. Every manuscript must report similarity scores along all four dimensions, specify task-relevant weights, and justify the thresholds applied. Papers that omit this information should be revised before submission. The minimum reporting standard aligns with the transparency already advocated in large-dataset transfer studies by Hoffmann *et al.* [6] and alloy-specific workflows by Sun *et al.* [8].

For reviewers, the new standard supplies concrete evaluation criteria. A reviewer can now ask: “Which dimensions were quantified? What were the exact thresholds? Were they exceeded for the stated task?” Claims that rely on database provenance or single-dimension assertions (element overlap alone, for example) must be flagged as insufficient. This shifts peer review from subjective judgment to verifiable boundary compliance.

For benchmark designers, the framework suggests the construction of similarity-controlled transfer tasks. Standard datasets should include pre-computed similarity matrices across the four dimensions for representative source–target pairs. Such benchmarks would enable systematic studies of positive, negative, and neutral transfer as a function of controlled similarity distances. Batra *et al.* [12] called for more structured materials intelligence ecosystems; similarity-controlled benchmarks constitute a direct implementation of that vision.

Collectively, these implications convert material similarity from a rhetorical convenience into a community-wide reporting norm. The result will be fewer irreproducible claims, clearer meta-analyses, and accelerated progress toward reliable composition-transfer learning across alloys and crystals.

Conclusion

This study reframes material similarity from an informal descriptor into a formally bounded, task-dependent construct that governs the validity of composition-transfer learning. By decomposing similarity into four orthogonal dimensions and embedding them within a thresholded, weighted framework, it resolves longstanding ambiguities that have limited comparability across studies and obscured the causes of transfer success or failure. The analysis demonstrates that similarity cannot be inferred from isolated features or assumed from shared provenance, but must instead be established through explicit, quantitative alignment with task-relevant mechanisms. Boundary conditions and gray-zone cases further clarify that transfer outcomes emerge from the interaction between dimensional alignment, model capacity, and data sufficiency, rather than from similarity alone.

Beyond conceptual clarification, the framework introduces a shift in practice: transfer is justified not retrospectively but through verifiable preconditions that can be reported, tested, and reproduced. This repositioning has implications for experimental design, benchmarking, and peer evaluation, enabling the field to move from anecdotal reasoning toward principled decision-making. In doing so, it establishes material similarity as a foundational layer linking chemical representation, model behavior, and domain adaptation performance, thereby supporting the maturation of materials machine learning into a more predictive and systematically grounded discipline.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 21 Jun 2022

Revised: 18 Oct 2022

Accepted: 27 Dec 2022

Published online: 18 January 2023

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

- Hu J, Liu D, Fu N, Dong R. Improving realistic material property prediction using domain adaptation based machine learning [Preprint]. arXiv; 2023. arXiv:2308.02937.
- Goetz A, Durmaz AR, Müller M, Thomas A, Britz D, Kerfriden P, et al. Addressing materials' microstructure diversity using transfer learning. *NPJ Comput Mater*. 2022;8(1):27.
<https://doi.org/10.1038/s41524-022-00703-z>.
- Jha D, Choudhary K, Tavazza F, Liao WK, Choudhary A, Campbell C, et al. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. *Nat Commun*. 2019;10(1):5316.
<https://doi.org/10.1038/s41467-019-13297-w>.
- Kong S, Guevarra D, Gomes CP, Gregoire JM. Materials representation and transfer learning for multi-property prediction. *Appl Phys Rev*. 2021;8(2):021409.
<https://doi.org/10.1063/5.0047066>.
- Feng S, Zhou H, Dong H. Application of deep transfer learning to predicting crystal structures of inorganic substances. *Comput Mater Sci*. 2021;195:110476.
<https://doi.org/10.1016/j.commatsci.2021.110476>.

Hoffmann N, Schmidt J, Botti S, Marques MAL. Transfer learning on large datasets for the accurate prediction of material properties. *Digit Discov.* 2023;2(5):1368-79.
<https://doi.org/10.1039/D3DD00030C>.

Ford E, Maneparambil K, Kumar A, Sant G, Neithalath N. Transfer (machine) learning approaches coupled with target data augmentation to predict the mechanical properties of concrete. *Mach Learn Appl.* 2022;8:100271.
<https://doi.org/10.1016/j.mlwa.2022.100271>.

Sun H, Zhang H, Ren G, Zhang C. A knowledge transfer framework for general alloy materials properties prediction. *Materials (Basel).* 2022;15(21):7442.
<https://doi.org/10.3390/ma15217442>.

Jiang L, Zhang Z, Hu H, He X, Fu H, Xie J. A rapid and effective method for alloy materials design via sample data transfer machine learning. *NPJ Comput Mater.* 2023;9(1):26.
<https://doi.org/10.1038/s41524-023-00979-9>.

Lee J, Asahi R. Transfer learning for materials informatics using crystal graph convolutional neural network. *Comput Mater Sci.* 2021;190:110314.
<https://doi.org/10.1016/j.commatsci.2021.110314>.

Axelrod S, Schwalbe-Koda D, Mohapatra S, Damewood J, Greenman KP, Gómez-Bombarelli R. Learning matter: Materials design with machine learning and atomistic simulations. *Acc Mater Res.* 2022;3(3):343-57.
<https://doi.org/10.1021/accountsmr.1c00238>.

Batra R, Song L, Ramprasad R. Emerging materials intelligence ecosystems propelled by machine learning. *Nat Rev Mater.* 2021;6(8):655-78.
<https://doi.org/10.1038/s41578-020-00255-y>.

Bartók AP, De S, Poelking C, Bernstein N, Kermode JR, Csányi G, et al. Machine learning unifies the modeling of materials and molecules. *Sci Adv.* 2017;3(12).
<https://doi.org/10.1126/sciadv.1701816>.

Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. *Chem Mater.* 2019;31(9):3564-72.
<https://doi.org/10.1021/acs.chemmater.9b01294>.

Zhao X, Shao F, Zhang Y. A novel joint adversarial domain adaptation method for rotary machine fault diagnosis under different working conditions. *Sensors (Basel).* 2022;22(22):9007.
<https://doi.org/10.3390/s22229007>.

Parvez MR, Chang KW. Evaluating the values of sources in transfer learning. In: *Proceedings of the 2021 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies; 2021 Jun 6-11; Online. Stroudsburg (PA): Association for Computational Linguistics; 2021. p. 5084-116.*
<https://doi.org/10.18653/v1/2021.naacl-main.402>.

Wong LJ, McPherson S, Michaels AJ. Assessing the value of transfer learning metrics for RF domain adaptation [Preprint]. *arXiv*; 2022. arXiv:2206.08329.
<https://doi.org/10.48550/arXiv.2206.08329>.

Chatterjee M, Roy K. Chemical similarity and machine learning-based approaches for the prediction of aquatic toxicity of binary and multicomponent pharmaceutical and pesticide mixtures against *Aliivibrio fischeri*. *Chemosphere.* 2022;308(Pt 3):136463.
<https://doi.org/10.1016/j.chemosphere.2022.136463>.

Skinninger MA, Dejong CA, Franczak BC, McNicholas PD, Magarvey NA. Comparative analysis of chemical similarity methods for modular natural products with a hypothetical structure enumeration algorithm. *J Cheminform.* 2017;9(1):46.

<https://doi.org/10.1186/s13321-017-0234-y>.

Wu F, Courty N, Jin S, Li SZ. Improving molecular representation learning with metric learning-enhanced optimal transport. *Patterns* (N Y). 2023;4(4):100714.

<https://doi.org/10.1016/j.patter.2023.100714>.

Wang Z, Goetz J, Brenning A. Transfer learning for landslide susceptibility modeling using domain adaptation and case-based reasoning. *Geosci Model Dev*. 2022;15(23):8765-84.

<https://doi.org/10.5194/gmd-15-8765-2022>.

Liu F, Liu Q, Bannur S, Pérez-García F, Usuyama N, Zhang S, et al. Compositional zero-shot domain transfer with text-to-text models. *Trans Assoc Comput Linguist*. 2023;11:1097-113.

https://doi.org/10.1162/tacl_a_00585.

Park K, Ko YJ, Durai P, Pan CH. Machine learning-based chemical binding similarity using evolutionary relationships of target genes. *Nucleic Acids Res*. 2019;47(20).

<https://doi.org/10.1093/nar/gkz743>.

Wang HC, Botti S, Marques MAL. Predicting stable crystalline compounds using chemical similarity. *NPJ Comput Mater*. 2021;7(1):12.

<https://doi.org/10.1038/s41524-020-00481-6>.

Hwang J, Iwasaki Y. Improving efficiency of autonomous material search via transfer learning from nontarget properties. *Sci Technol Adv Mater Methods*. 2023;3(1):2254202.

<https://doi.org/10.1080/27660400.2023.2254202>.

Zhu J, Zhang X, Guo M, Li J, Hu J, Cai S, et al. Restructured single parabolic band model for quick analysis in thermoelectricity. *NPJ Comput Mater*. 2021;7(1):116.

<https://doi.org/10.1038/s41524-021-00587-5>.

Chen C, Ong SP. AtomSets as a hierarchical transfer learning framework for small and large materials datasets. *NPJ Comput Mater*. 2021;7(1):173.

<https://doi.org/10.1038/s41524-021-00639-w>.

Xu Q, Shen S, de Souza RS, Chen M, Ye R, She Y, et al. From images to features: Unbiased morphology classification via variational auto-encoders and domain adaptation. *Mon Not R Astron Soc*. 2023;526(4):6391-400.

<https://doi.org/10.1093/mnras/stad3181>.

Gupta V, Choudhary K, Mao Y, Wang K, Tavazza F, Campbell C, et al. MPpredictor: An artificial intelligence-driven web tool for composition-based material property prediction. *J Chem Inf Model*. 2023;63(7):1865-71.

<https://doi.org/10.1021/acs.jcim.3c00307>.