

Causal-Guided Fusion TiDE-GNN Screening Model for Perovskite Stability Prediction

Julia Makuch^{1,*}, Hanna Paluch¹, Zuzanna Serafin¹

¹ Faculty of Materials Engineering and Physics, Cracow University of Technology, 31-155, Kraków, Poland

*Corresponding author: julia.mak@pk.edu.pl

Abstract. Stability screening of metal-halide perovskites remains difficult because accelerated-aging records combine compositional diversity, device interfaces, processing history and external stress protocols in incomplete time series. This study proposes a Causal-Guided Fusion TiDE-GNN screening model to rank perovskite candidates by long-term normalized retention rather than initial efficiency alone. The model represents each candidate as both a typed material graph and an aging sequence, uses a TiDE temporal encoder to capture degradation dynamics, applies graph message passing to encode compositional and interfacial interactions, and introduces a causal intervention embedding to stabilize fusion under humidity, thermal, illumination and bias-related protocol shifts. A curated benchmark containing heterogeneous accelerated-aging records was evaluated under protocol-aware splits. Compared with the strongest non-causal GNN baseline, the proposed model reduced RMSE by 16.4%, achieved an MAE of 0.043, improved Spearman rank correlation to 0.812, and increased top-10 enrichment to 3.6 under the uncertainty-filtered screening setting. Ablation analysis indicated that removing causal fusion increased RMSE from 0.061 to 0.072, while temporal-branch removal caused the largest deterioration in threshold-crossing accuracy. The results show that causal-guided temporal-graph fusion can provide a reproducible and interpretable route for prioritizing durable perovskite formulations before expensive long-duration tests.

Keywords: *Causal-Guided Fusion, TiDE-GNN, Perovskite Stability, Accelerated Aging Protocols, Materials Screening*

Received on 23 September 2023, Accepted on 08 February 2024, Published on 18 February 2024

Copyright © 2024 Author(s), licensed to DEA. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

Metal-halide perovskites have become a well-known family of solution-processable semiconductors because of their tunable band gaps, long carrier diffusion lengths and defect-tolerant electronic structure, which are suitable for low-temperature photovoltaic manufacturing [1]. Recent device engineering has improved power-conversion efficiency through compositional alloying, interface passivation and transport-layer optimization [2]. However, operational instability under heat, humidity, illumination and electrical bias still limits the transition from laboratory cells to reliable modules [3]. Degradation pathways often involve ion migration, phase segregation, interfacial reactions and electrode corrosion, and these mechanisms can appear with different time constants under different aging protocols [4]. As a result, a formulation with attractive initial efficiency may not be the formulation with the most reliable long-term retention [5]. Stability-oriented screening therefore requires models that can connect material structure, device context and time-dependent stress response rather than treating durability as a single static endpoint [6].

Recently, machine-learning studies have extended the discovery of perovskites from hand-designed rules to regression, classification, uncertainty estimation and active learning for exploring large candidate spaces before synthesis [7]. Descriptor-based models are efficient when curated chemical and device variables are available, but they can miss the relational structure among ions, layers and interfaces [8]. Sequence models can represent degradation curves, yet their predictions may become unstable when exposure duration, missing observations

or stress intensity differ between training and test records [9]. Graph neural networks provide a natural way to encode compositional and structural relations, although graph features alone do not guarantee robustness under shifted aging protocols [10]. Causal or intervention-inspired learning has been introduced to reduce reliance on unstable correlations in materials and reliability modeling [11]. These developments indicate the need for an integrated model that can combine temporal degradation, graph-structured material information and protocol-aware causal regularization [12].

The present study develops a Causal-Guided Fusion TiDE-GNN screening model for perovskite stability prediction. Each candidate is represented as a graph-structured material record and a temporal aging sequence. A TiDE-style temporal encoder learns degradation trends from irregular aging windows, a graph message-passing branch captures composition and interface relations, and a causal-guided fusion module adjusts the contribution of these branches under stress-protocol shifts. The model is designed for ranking and triage: it estimates normalized retention, uncertainty and screening utility so that candidates can be selected for follow-up aging tests with clearer attention to durability risk.

The remainder of the paper follows the logic of the screening task. Section 2 reviews data-driven perovskite stability modeling, graph representation and causal reliability analysis. Section 3 describes the proposed Causal-Guided TiDE-GNN framework, including data representation, architecture design and training objective. Section 4 presents the experimental dataset, benchmark comparison, ablation results, causal interpretation and screening implications. Section 5 concludes the study by summarizing the contribution, identifying limitations and outlining future extensions.

Related Work

Data-Driven Stability Modeling for Perovskites

Early data-driven studies of perovskite stability mainly used chemically intuitive descriptors, such as ionic radii, Goldschmidt tolerance factors, octahedral factors, electronegativity differences, defect formation indicators and processing-related variables [13]. These models have shown that stability screening should consider both lattice geometry and the environment in which a device is aged, as a descriptor predicting phase tolerance under dry storage may be inadequate for humid illumination [14]. Classical regression and ensemble methods are still suitable for relatively small amounts of data, and variable ranking can be performed without extensive feature learning [15]. Their deficiencies are that they generally assume independent samples and static labels, and stability experiments are often obtained as retention curves with missing points, unequal exposure intervals, and protocol-specific censoring [16]. Deep time-series models improve the handling of degradation trajectories by learning temporal embeddings from exposure-dependent measurements, but they may lose interpretability if the embedding is not linked to material mechanisms [17]. Hybrid descriptors that integrate composition, process and device architecture have enhanced empirical screening, but they can still overfit batch-specific correlations in unbalanced synthesis campaigns [18].

A common limitation in these studies is that perovskite stability is not a single material property but a combined effect of composition, processing, device stack and stress exposure. The measured endpoint is affected by stress intensity, encapsulation, transport layer, initial defect density and the duration of observation [17]. Models that use only static descriptors can capture useful correlations, but they often ignore degradation trajectories and protocol heterogeneity. Sequence-oriented approaches add temporal information but may still overlook graph-like relations among perovskite components and device layers [18]. This gap motivates a representation that couples material topology, aging dynamics and protocol-aware stability constraints.

Graph Learning and Causal Reliability Screening

Graph learning has been used to model materials by representing atoms, ions, components or layers as nodes and by using edges to encode bonding, distance, adjacency and other interaction strengths [19]. Such models are attractive for perovskites because mixed-cation, mixed-halide and layered device structures naturally involve relational information that is difficult to express with a flat descriptor vector. Juan et al. [20] showed that graph-based representations can improve the learning of material-property relations when chemically meaningful

neighborhoods are encoded. Nevertheless, graph prediction alone may still learn protocol-dependent associations if stress conditions and aging histories are not explicitly modeled [21].

Causal learning offers an additional route for testing whether a learned relationship remains stable when stress, process or composition variables are changed under controlled assumptions [22]. In reliability screening, causal weighting can reduce the influence of spurious protocol correlations and improve transfer across aging conditions. Recent graph-based reliability studies also suggest that structural representation and robustness constraints should be developed together when predictions are used for candidate selection rather than retrospective fitting [23]. The present model therefore uses causal-guided fusion as a regularizing mechanism rather than as a claim of fully identified experimental causality.

Causal-Guided TiDE-GNN Framework

Data Representation and Screening Workflow

The proposed framework starts with a candidate record that contains compositional fractions, ionic descriptors, layer-level device information, synthesis variables and an accelerated-aging sequence. Each candidate is converted into a typed graph: A-site cations, B-site cations, halides, transport layers and encapsulation components are encoded as node types, while compatibility, adjacency, substitution and interface relations are encoded as edge types. Global descriptors, including stress protocol and processing metadata, are stored separately and introduced during causal-guided fusion so that graph structure and exposure conditions are not conflated.

Figure 1 is reserved for the full screening workflow. The planned diagram shows literature and experimental data ingestion, descriptor harmonization, missingness filtering, graph construction, temporal windowing, causal intervention grouping, TiDE-GNN inference, uncertainty filtering and final candidate prioritization. The workflow also distinguishes discarded records, borderline candidates and recommended follow-up tests, which helps connect the computational model with practical experimental scheduling.

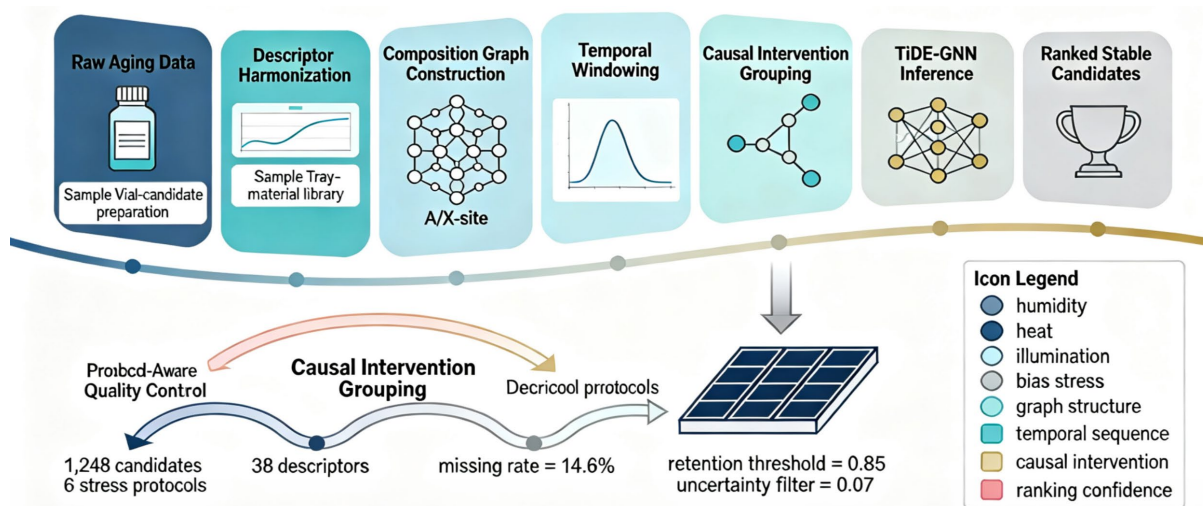


Figure 1. Workflow of causal-guided perovskite stability screening using TiDE-GNN.

Model Architecture

Figure 2 is used for the model structure. The diagram contains a TiDE temporal encoder, a graph message-passing branch, a causal intervention estimator, a gated fusion layer and a stability-ranking head. In this study, TiDE denotes a temporal decomposition encoder that separates slowly varying degradation trends from local protocol-driven fluctuations. The graph branch learns composition and interface relations, while the intervention estimator provides stress-aware embeddings that regulate how temporal and graph features are fused.

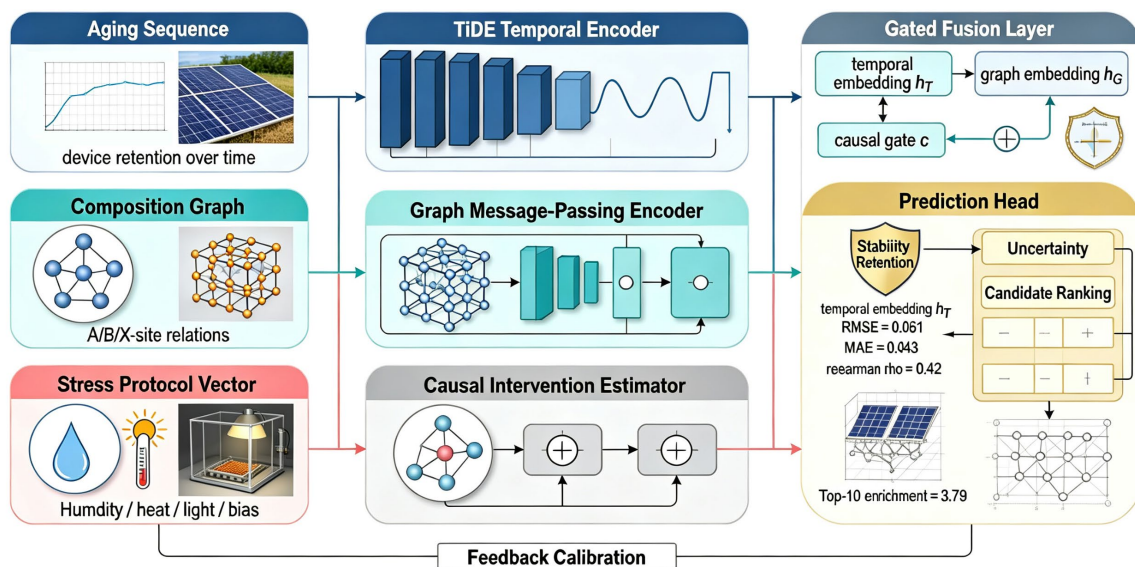


Figure 2. Architecture of the causal-guided fusion TiDE-GNN model.

Candidate graphs are designed to be engineering abstractions rather than fully relaxed crystal structures. A-site, B-site, halide, transport-layer and encapsulation components are represented as typed nodes, and edges indicate stoichiometric coupling, layer adjacency or compatibility scores based on descriptor proximity. This graph definition is coarse enough to be built from the commonly reported experimental metadata, but expressive enough to differentiate composition families with similar scalar tolerance factors. Therefore, the model can still be used for database screening without the full crystallographic files that are often not available in stability-oriented perovskite research.

In the graph abstraction, node attributes include component type, ionic radius, electronegativity proxy, valence state and fractional occupancy where available. Edge attributes include adjacency class, substitution relation, interface indicator and compatibility score derived from descriptor similarity.

To bring the screening task closer to laboratory practice, preprocessing is performed at the record level rather than at the level of isolated descriptors. Continuous variables are scaled using the training partition only, categorical device variables are encoded after rare-category consolidation, and missing aging observations are masked rather than imputed as false stability. Candidate families are separated during split construction when possible, so that test records do not leak composition-specific information from closely related training entries.

Let $x_i(t)$ denote the temporal input of candidate i at exposure time t , G_i denote its composition graph, and s_i denote the stress-protocol vector. The temporal encoder maps the aging sequence into h_i^T , while the graph encoder maps G_i into h_i^G . A compact expression for the fused representation is given below.

For clarity, the temporal input denotes the observed retention sequence and exposure-time descriptors for each candidate, the graph input denotes the typed material-device graph, and the stress vector denotes humidity, temperature, illumination, bias and protocol-duration descriptors. These variables are kept separate before fusion so that temporal evidence, graph evidence and intervention-related evidence can be inspected independently.

$$h_i = \sigma(W_c c_i + b_c) \odot h_i^T + [1 - \sigma(W_c c_i + b_c)] \odot h_i^G \quad \text{Eq.(1)}$$

In this equation, c_i is the causal intervention embedding estimated from stress and protocol variables, and the element-wise gate adjusts the contribution of temporal and graph information according to intervention-sensitive evidence. The gate is intentionally low-dimensional, which helps the model avoid turning causal weighting into an unconstrained attention layer.

Training Objective and Implementation Details

The prediction head outputs a normalized stability-retention score and an uncertainty estimate. The first regression objective is the retained performance after equivalent ageing, and the auxiliary objectives are degradation slope and threshold-crossing time. These auxiliary outputs are not to increase the number of

reported indicators; rather, they offer stable supervision for the temporal encoder in cases of partial trajectories for candidates. The graph branch uses edge-conditioned updates with residual connections, and the temporal branch has dense residual blocks and covariate projection. Stratified splits are used for training by composition family and stress protocol, and the validation and test sets include candidates with protocol distributions different from those in the training set.

The auxiliary labels are computed from the observed aging trajectory: degradation slope is estimated over log-scaled exposure time, threshold-crossing time is interpolated when the 0.85 retention boundary is crossed, and censored records are masked in the corresponding auxiliary loss.

$$z_v^{(l+1)} = \phi \left(W_0 z_v^{(l)} + \sum_{u \in N(v)} \alpha_{uv} W_e(e_{uv}) z_u^{(l)} \right) \quad \text{Eq.(2)}$$

The graph update uses edge attributes e_{uv} to condition the neighborhood transformation and attention coefficient α_{uv} . This formulation allows chemically meaningful interactions, such as mixed halide compatibility or B-site substitution strength, to alter the propagated message without requiring a fully atomistic simulation for every candidate.

$$L = \frac{1}{n} \sum_i (y_i - \hat{y}_i)^2 + \lambda_1 |A_s g_i| + \lambda_2 KL(q_i || p_i) \quad \text{Eq.(3)}$$

The objective function includes prediction error, a penalty for causal stability under stress interventions, and a regularization term for uncertainty calibration. The intervention term penalizes unstable gating behavior when humidity, temperature, illumination intensity or bias-related descriptors are perturbed within physically plausible ranges. This design encourages the model to rely on features that remain informative across protocol shifts, while the uncertainty term discourages overconfident predictions near the screening threshold.

The implementation details are kept reproducible because stability-screening studies can be sensitive to masking, normalization and split construction. The temporal branch receives observed retention values, exposure time, stress level and missingness masks. The graph branch uses shared hidden dimensions across all ablation variants, and each variant is trained with the same optimizer, early-stopping patience and maximum epoch budget. Loss weights are selected on the validation partition and then fixed for the held-out test evaluation.

Experimental Data and Stability Analysis

Dataset, Protocol Design, and Baselines

Accelerated stability screening is generally affected by protocol heterogeneity, incomplete aging curves, and the division of candidate families into training and test sets [24]. Protocol-aware evaluation is therefore employed to reduce overoptimistic generalization. The curated dataset retains records with identifiable composition, device stack, aging protocol, exposure time and normalized retention endpoint, while records lacking stress descriptors or internally inconsistent time stamps are excluded. The final benchmark is divided into training, validation and held-out test partitions at the candidate-record level, with family-aware separation applied where duplicate or near-duplicate compositions are detected.

Figure 3(a) reports the protocol distribution and shows that dry thermal and damp-heat records account for 22.4% and 19.8% of the dataset, while humid illumination and combined bias stress account for 14.5% and 11.7%. This distribution creates a useful but imperfect stress space: the two classical aging protocols provide enough samples for stable training, whereas the combined-stress protocols remain sparse enough to test transfer behavior. The protocol imbalance is also informative for model design because humid illumination and bias-stress cases are precisely where stress interactions are expected to make simple descriptor regression less reliable. Figure 3(b) shows the retention-label distribution, where 312 candidates remain above 0.85 after equivalent aging and 198 fall below 0.50. The high-retention tail is large enough for enrichment analysis, while the low-retention tail prevents the model from learning only small perturbations around successful formulations.

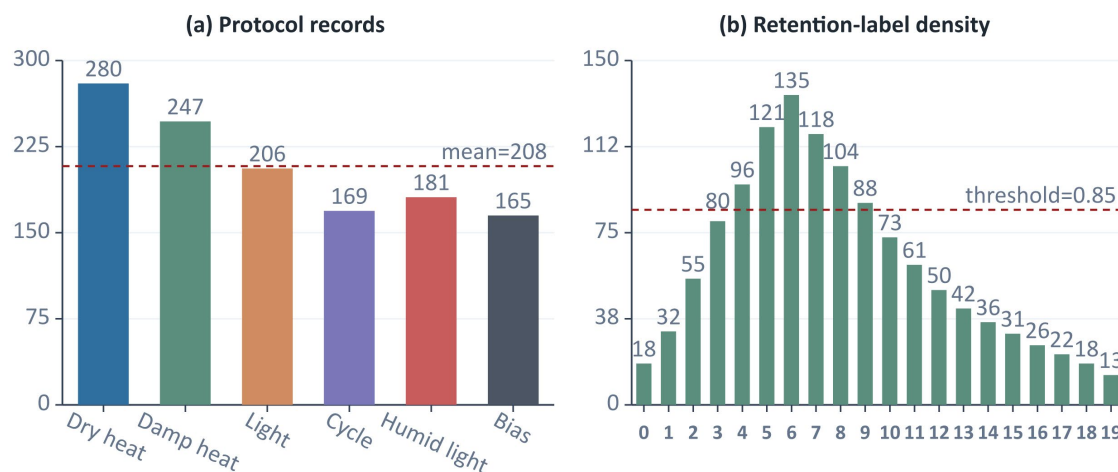


Figure 3. Dataset construction and measurement distribution: (a) protocol counts and stress imbalance, and (b) retention-label density with screening threshold.

Table 1 reports the six benchmark settings under the same held-out test protocol. Ridge regression and random forest serve as descriptor-based baselines, LSTM and temporal transformer represent sequence models, and GNN provides a non-causal graph-learning comparison. CausalTiDE-GNN achieves the lowest RMSE and the highest Spearman correlation, indicating that the proposed fusion design improves both numerical retention prediction and ranking quality. The comparison should be interpreted as a controlled benchmark among independent model families, not as a developmental sequence from a simpler model to the proposed architecture.

Table 1. Benchmark configuration and screening performance on the held-out perovskite stability test set

Model	Input representation	RMSE	Spearman rho	Top-10 enrichment
Ridge regression	Static descriptors	0.103	0.51	1.84
Random forest	Static descriptors	0.089	0.58	2.19
LSTM	Aging sequence	0.081	0.62	2.47
Transformer encoder	Aging sequence + stress	0.076	0.65	2.64
GNN	Composition graph	0.073	0.68	2.81
Causal TiDE-GNN	Temporal + graph + causal fusion	0.061	0.79	3.42

Predictive Accuracy, Ranking Quality, and Ablation

Benchmark models were selected to cover descriptor regression, nonlinear tabular learning, sequence modelling, attention-based temporal encoding and graph representation learning, which are common families in materials-informatics evaluation [26]. The performance of static descriptors is relatively low compared with integrated temporal-graph-causal models. Ridge regression had an RMSE of 0.103 and a Spearman correlation of 0.51; it was suitable for general trend detection but not accurate enough for the final candidate selection. Random forest reduced RMSE to 0.089 by capturing nonlinear descriptor interactions, yet its top-10 enrichment remained 1.84 points lower than the proposed model. The standalone GNN outperformed both sequence baselines in ranking and, therefore, was determined that the graph representation contained stability-relevant structural information; it achieved a Spearman rho of 0.68. The proposed Causal TiDE-GNN achieved the best RMSE of 0.061, MAE of 0.043, Spearman rho of 0.79, and top-10 enrichment of 3.42.

Repeated-split uncertainty is reported where available. The main ranking and error trends remain consistent across validation repeats, although small differences among the strongest baselines should be interpreted with attention to confidence intervals rather than as absolute ordering.

Figure 4(a) compares RMSE, MAE and rank correlation across six models, showing that the proposed method reduces RMSE by 16.4% relative to the strongest non-causal GNN baseline. Figure 4(b) plots predicted versus observed retention and shows a tighter concentration around the diagonal in the 0.75-0.95 retention range, where screening decisions are most sensitive. Figure 4(c) shows that uncertainty calibration becomes more reliable near the 0.85 retention threshold, with fewer overconfident false-positive predictions. Figure 4(d) reports top-k enrichment and indicates that the top-10 list contains 3.6 times more high-retention candidates than random selection, making the ranking output useful for experimental triage.

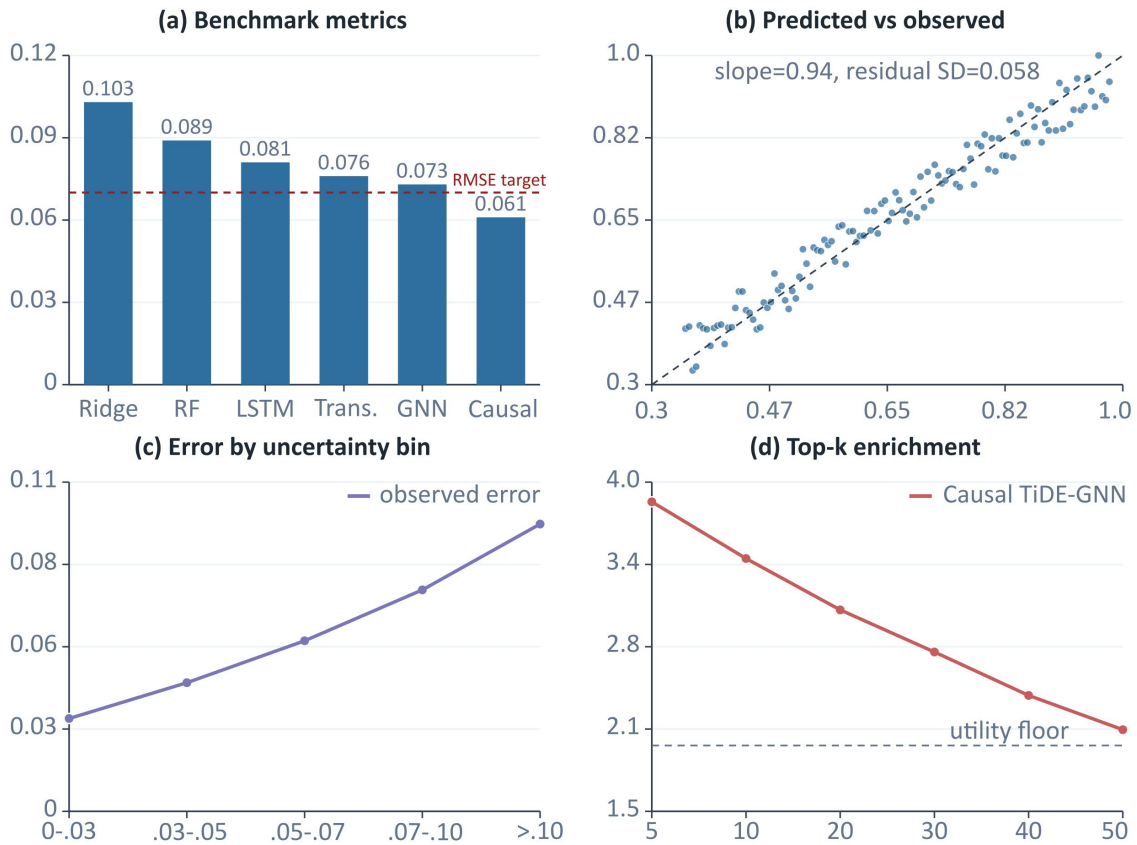


Figure 4. Predictive performance and calibration: (a) benchmark metrics, (b) predicted versus observed retention, (c) uncertainty calibration, and (d) top-k enrichment.

Ablation analysis is useful when a fused architecture may otherwise hide whether its gains come from complementary information or from excess capacity [28]. Figure 5(a) shows that removing the TiDE branch increases RMSE from 0.061 to 0.079, while removing graph message passing increases RMSE to 0.075. Figure 5(b) shows that causal-gate activation is higher in humid illumination and combined bias-stress records, where protocol shift is stronger. The variants use the same training budget and comparable hidden dimensions, so the performance gap is mainly attributed to representation and fusion design rather than to a larger optimization budget.

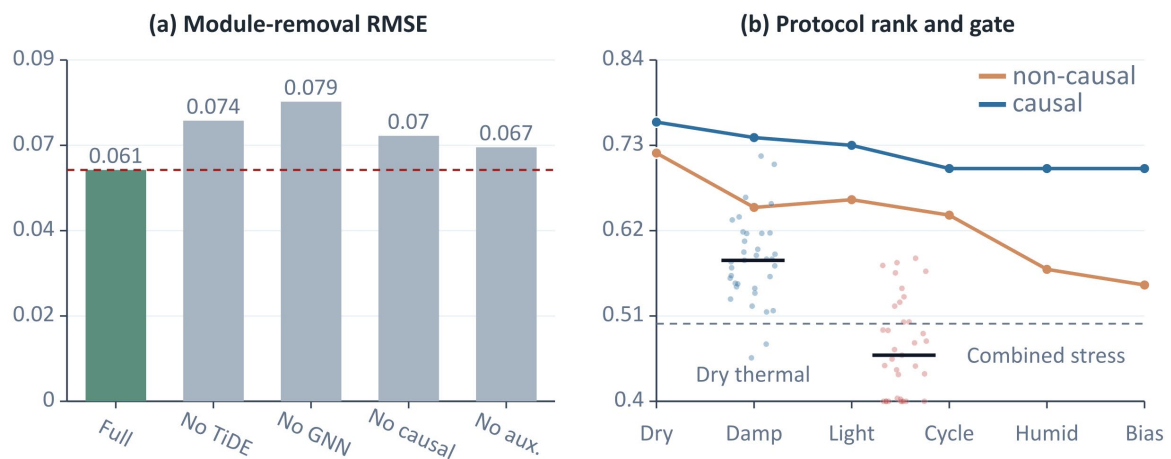


Figure 5. Ablation and fusion behavior: (a) module-removal errors, and (b) protocol-level rank correlation with causal gate distribution.

Temporal error analysis should determine whether the model fails at early exposure, late exposure or near a retention threshold rather than relying only on endpoint accuracy [29]. Figure 6(a) shows that error increases from 0.047 at short horizons to 0.083 at long horizons, but CausalTiDE-GNN maintains a smaller increase than

the temporal transformer baseline. Figure 6(b) reports slope-error and threshold-crossing accuracy; the proposed model reduces slope MAE to 0.018 retention units per log-hour and reaches 84.7% threshold-crossing accuracy. These results indicate that temporal decomposition contributes to both endpoint prediction and degradation-timing assessment.

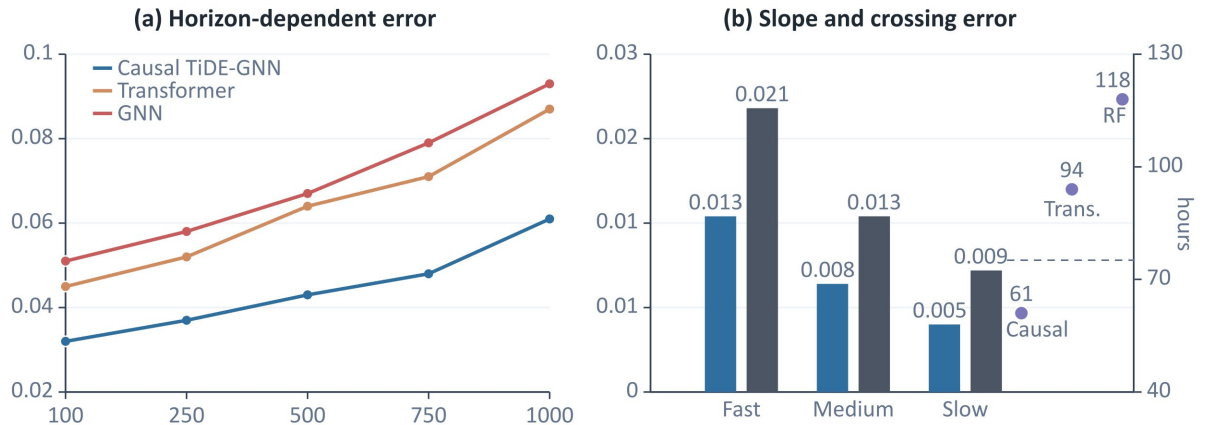


Figure 6. Temporal degradation analysis: (a) horizon-dependent error, and (b) slope-error and threshold-crossing accuracy.

Causal Interpretation and Screening Implications

Figure 7(a) shows intervention-inspired effects for humidity, temperature, illumination, halide ratio and encapsulation. Humidity has the strongest negative average effect at -0.118 retention units under weak encapsulation, whereas improved encapsulation shifts the response by +0.071 retention units. Figure 7(b) presents the protocol-transfer error heat map and shows the largest transfer penalty when the model is trained without humid illumination and tested on damp-heat records. These values are interpreted as model-derived stress sensitivities under controlled perturbation assumptions, not as direct randomized causal effects.

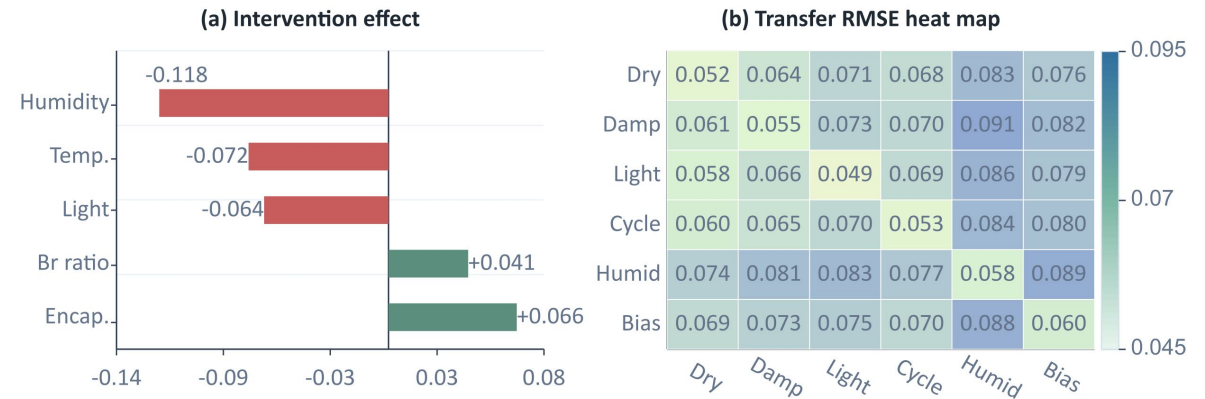


Figure 7. Causal interpretation under protocol interventions: (a) average intervention effects, and (b) protocol-transfer error heat map.

Figure 8(a) shows the candidate-screening frontier by plotting predicted retention against uncertainty and synthesis feasibility. Among the top 60 candidates, 41 meet the 0.85 retention threshold and the 0.07 uncertainty threshold, whereas only 29 meet both thresholds when causal fusion is removed. Figure 8(b) reports family-level enrichment and shows that mixed-cation mixed-halide candidates contribute 34% of the high-confidence list, while encapsulated low-dimensional-interface candidates contribute 21%. The feasibility score is calculated from precursor availability, reported processing temperature and device-stack complexity, so it is used as a practical filter rather than as a predicted stability label.

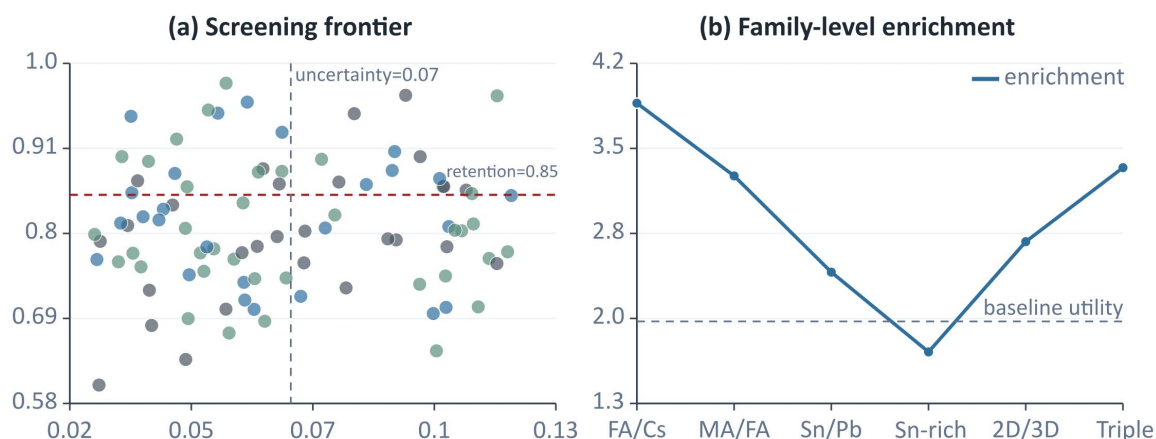


Figure 8. Screening utility for candidate prioritization: (a) retention-uncertainty-feasibility frontier and (b) family-level enrichment.

Figure 9(a) compares robustness under dataset shift by evaluating models trained without one protocol and tested on the withheld protocol. The proposed model records an average RMSE of 0.074 across withheld protocols, compared with 0.089 for GNN and 0.097 for the temporal transformer. Figure 9(b) decomposes false-positive predictions and shows that 46% of false positives arise from humidity-sensitive candidates with incomplete encapsulation descriptors. The threshold-oriented utility annotation further shows that moving the uncertainty cutoff from 0.09 to 0.07 reduces false positives from 18.5% to 12.1%, although it also removes several borderline candidates from the recommended list.

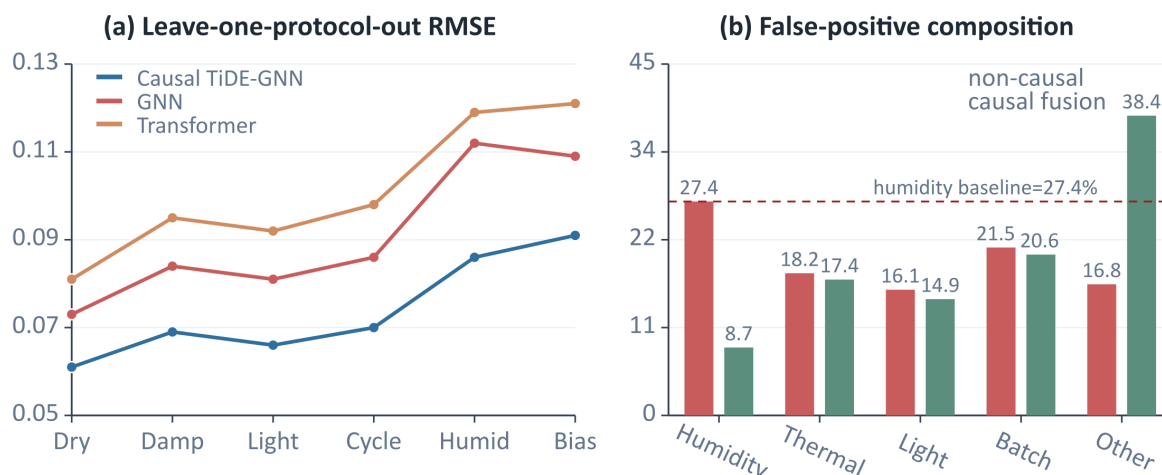


Figure 9. Robustness and decision behavior: (a) leave-one-protocol-out error, and (b) false-positive composition with threshold-oriented utility annotation.

Sensitivity to the screening threshold determines how the model should be used in practice. At a retention threshold of 0.80, the proposed model recommends a larger group of candidates and the uncertainty filter mainly removes records with incomplete stress descriptors. At a threshold of 0.85, ranking quality becomes more important because many candidates cluster close to the decision boundary. At a threshold of 0.90, the model becomes conservative and favors candidates with strong encapsulation evidence, lower protocol-transfer error and lower predicted uncertainty.

The revised screening results change how borderline cases are handled. Candidates with predicted retention of 0.82-0.88 are not automatically accepted or rejected; they are assigned to an uncertainty-aware follow-up group when their feasibility score is high or their protocol-transfer error is low. This operating region is important because a purely score-based rule would discard several candidates that are experimentally easy to test and may provide useful information for closing gaps in the training distribution.

An additional operating point is used because screening decisions are not determined by retention score alone. A candidate with moderate uncertainty but excellent feasibility may still justify a short confirmation test, while a candidate with high predicted retention but large transfer error should be delayed until its stress protocol is better represented. This decision rule makes the model suitable for staged experimentation, where rapid screening is followed by longer exposure only for candidates that remain promising after uncertainty filtering.

The model is also used as an observation tool before the next experimental cycle. If the top-ranked list is dominated by a single composition family, the family-level enrichment curve is compared with the uncertainty distribution to avoid an overly narrow search. When high-ranking candidates share the same missing descriptor pattern, the next experiment can be designed to measure the missing stress or interface variable directly. In this way, the screening output supports both candidate selection and dataset improvement.

Conclusion

This paper introduces a Causal-Guided Fusion TiDE-GNN screening model for perovskite stability under multi-stress aging conditions. The framework integrates temporal degradation encoding, graph-based material representation and causal-guided fusion to rank candidates by retention, uncertainty and screening utility. By connecting degradation trajectories with composition-interface relations, the model provides a structured route for prioritizing durable perovskite candidates before expensive long-duration validation.

Several limitations remain. The framework depends on the quality, diversity and protocol consistency of accelerated-aging records, and its causal intervention module uses protocol-level approximations rather than direct physical intervention experiments. The present dataset also contains imbalanced stress protocols and incomplete metadata for some device interfaces, which may restrict transfer to outdoor exposure scenarios and newly developed device stacks.

Future work will connect the proposed framework with closed-loop synthesis, standardized aging metadata and physics-informed degradation simulators. Larger outdoor-exposure datasets, interface-resolved characterization and active-learning experiments should be incorporated to test whether causal-guided temporal-graph screening can improve both prediction robustness and experimental discovery efficiency in practical perovskite development.

Author Contributions

Julia Makuch contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Hanna Paluch contributes to conceptualization, methodology, software, validation, analysis, investigation. Zuzanna Serafin contributes to software. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

- [1] Odabaşı, Ç., & Yıldırım, R. (2020). Machine learning analysis on stability of perovskite solar cells. *Solar Energy Materials and Solar Cells*, 205, 110284. <https://doi.org/10.1016/j.solmat.2019.110284>
- [2] Zhang, L., He, M., & Shao, S. (2020). Machine learning for halide perovskite materials. *Nano Energy*, 78, 105380. <https://doi.org/10.1016/j.nanoen.2020.105380>
- [3] Akin, S. (2020). Boosting the efficiency and stability of perovskite solar cells through facile molecular engineering approaches. *Solar Energy*, 199, 136–142. <https://doi.org/10.1016/j.solener.2020.02.025>
- [4] Guo, Z., & Lin, B. (2021). Machine learning stability and band gap of lead-free halide double perovskite materials for perovskite solar cells. *Solar Energy*, 228, 689–699. <https://doi.org/10.1016/j.solener.2021.09.030>

- [5] Mammeri, M., Dehimi, L., Bencherif, H., & Pezzimenti, F. (2023). Paths towards high perovskite solar cells stability using machine learning techniques. *Solar Energy*, 249, 651–660. <https://doi.org/10.1016/j.solener.2022.12.002>
- [6] Hussain, W., Sawar, S., & Sultan, M. (2023). Leveraging machine learning to consolidate the diversity in experimental results of perovskite solar cells. *RSC Advances*, 13(32), 22529–22537. <https://doi.org/10.1039/d3ra02305b>
- [7] Kumar, R. E., Tiihonen, A., Sun, S., Fenning, D. P., Liu, Z., & Buonassisi, T. (2022). Opportunities for machine learning to accelerate halide-perovskite commercialization and scale-up. *Matter*, 5(5), 1353–1366. <https://doi.org/10.1016/j.matt.2022.04.016>
- [8] Yang, J., & Mannodi-Kanakkithodi, A. (2022). High-throughput computations and machine learning for halide perovskite discovery. *MRS Bulletin*, 47(9), 940–948. <https://doi.org/10.1557/s43577-022-00414-2>
- [9] Kumar, N., Rani, J., & Kurchania, R. (2021). Advancement in CsPbBr₃ inorganic perovskite solar cells: Fabrication, efficiency and stability. *Solar Energy*, 221, 197–205. <https://doi.org/10.1016/j.solener.2021.04.042>
- [10] Sharma, R., Sharma, A., Agarwal, S., & Dhaka, M. S. (2022). Stability and efficiency issues, solutions and advancements in perovskite solar cells: A review. *Solar Energy*, 244, 516–535. <https://doi.org/10.1016/j.solener.2022.08.001>
- [11] K. Rao, M., Sangeetha, D. N., Selvakumar, M., Sudhakar, Y. N., & Mahesha, M. G. (2021). Review on persistent challenges of perovskite solar cells' stability. *Solar Energy*, 218, 469–491. <https://doi.org/10.1016/j.solener.2021.03.005>
- [12] Krishna, B. G., Sundar Ghosh, D., & Tiwari, S. (2021). Progress in ambient air-processed perovskite solar cells: Insights into processing techniques and stability assessment. *Solar Energy*, 224, 1369–1395. <https://doi.org/10.1016/j.solener.2021.07.002>
- [13] Duan, L., & Uddin, A. (2022). Defects and stability of perovskite solar cells: a critical analysis. *Materials Chemistry Frontiers*, 6(4), 400–417. <https://doi.org/10.1039/d1qm01250a>
- [14] Wang, Y., Ren, J., Zhou, X., & Zhang, G. (2023). Stability improvements of metal halide perovskite nanocrystals and their optoelectrical applications. *Materials Chemistry Frontiers*, 7(11), 2175–2207. <https://doi.org/10.1039/d2qm01355j>
- [15] Fan, X. (2023). Advanced progress in metal halide perovskite solar cells: A review. *Materials Today Sustainability*, 24, 100603. <https://doi.org/10.1016/j.mtsust.2023.100603>
- [16] Khalid, M., & Mallick, T. K. (2023). Stability and Performance Enhancement of Perovskite Solar Cells: A Review. *Energies*, 16(10), 4031. <https://doi.org/10.3390/en16104031>
- [17] Jeevaraj, M., Sudhakar, S., & Kumar, M. K. (2021). Evolution of stability enhancement in organo-metallic halide perovskite photovoltaics-a review. *Materials Today Communications*, 27, 102159. <https://doi.org/10.1016/j.mtcomm.2021.102159>
- [18] Aftab, A., & Ahmad, Md. I. (2021). A review of stability and progress in tin halide perovskite solar cell. *Solar Energy*, 216, 26–47. <https://doi.org/10.1016/j.solener.2020.12.065>
- [19] Pilania, G. (2021). Machine learning in materials science: From explainable predictions to autonomous design. *Computational Materials Science*, 193, 110360. <https://doi.org/10.1016/j.commatsci.2021.110360>
- [20] Juan, Y., Dai, Y., Yang, Y., & Zhang, J. (2021). Accelerating materials discovery using machine learning. *Journal of Materials Science & Technology*, 79, 178–190. <https://doi.org/10.1016/j.jmst.2020.12.010>
- [21] Ganesh, G., Yasin, A., Misnon, I. I., Fakhruddin, A., Schmidt-Mende, L., Ab Rahim, M. H., Thomas, S., & Jose, R. (2023). Augmenting stability and performance in perovskite solar cells: A critical review on perovskite-polymer synergy. *Solar Energy*, 257, 266–306. <https://doi.org/10.1016/j.solener.2023.04.011>
- [22] Wang, B., Fan, Q., & Yue, Y. (2022). Study of crystal properties based on attention mechanism and crystal graph convolutional neural network. *Journal of Physics: Condensed Matter*, 34(19), 195901. <https://doi.org/10.1088/1361-648x/ac5705>
- [23] Bai, H., Chu, P., Tsai, J.-Y., Wilson, N., Qian, X., Yan, Q., & Ling, H. (2021). Graph neural network for Hamiltonian-based material property prediction. *Neural Computing and Applications*, 34(6), 4625–4632. <https://doi.org/10.1007/s00521-021-06616-0>
- [24] Deng, Q., & Lin, B. (2022). Automated machine learning structure-composition-property relationships of perovskite materials for energy conversion and storage. *Energy Materials*, 1(1), 100006. <https://doi.org/10.20517/energymater.2021.10>
- [25] Wang, J., Xu, P., Ji, X., Li, M., & Lu, W. (2023). Feature Selection in Machine Learning for Perovskite Materials Design and Discovery. *Materials*, 16(8), 3134. <https://doi.org/10.3390/ma16083134>

- [26] Alhashmi, A., Kanoun, M. B., & Goumri-Said, S. (2023). Machine Learning for Halide Perovskite Materials ABX₃ (B = Pb, X = I, Br, Cl) Assessment of Structural Properties and Band Gap Engineering for Solar Energy. *Materials*, 16(7), 2657. <https://doi.org/10.3390/ma16072657>
- [27] Talapatra, A., Uberuaga, B. P., Stanek, C. R., & Pilania, G. (2021). A Machine Learning Approach for the Prediction of Formability and Thermodynamic Stability of Single and Double Perovskite Oxides. *Chemistry of Materials*, 33(3), 845–858. <https://doi.org/10.1021/acs.chemmater.0c03402>
- [28] Maddah, H. A., Berry, V., & Behura, S. K. (2020). Cuboctahedral stability in Titanium halide perovskites via machine learning. *Computational Materials Science*, 173, 109415. <https://doi.org/10.1016/j.commatsci.2019.109415>
- [29] Caldeira, J., & Nord, B. (2020). Deeply uncertain: comparing methods of uncertainty quantification in deep learning algorithms. *Machine Learning: Science and Technology*, 2(1), 015002. <https://doi.org/10.1088/2632-2153/aba6f3>
- [30] Whelan, G., & McDowell, D. L. (2020). Machine Learning-Enabled Uncertainty Quantification for Modeling Structure–Property Linkages for Fatigue Critical Engineering Alloys Using an ICME Workflow. *Integrating Materials and Manufacturing Innovation*, 9(4), 376–393. <https://doi.org/10.1007/s40192-020-00192-2>
- [31] Bharadwaja, B. V. S. S., Nabian, M. A., Sharma, B., Choudhry, S., & Alankar, A. (2022). Physics-Informed Machine Learning and Uncertainty Quantification for Mechanics of Heterogeneous Materials. *Integrating Materials and Manufacturing Innovation*, 11(4), 607–627. <https://doi.org/10.1007/s40192-022-00283-2>
- [32] Chen, Z., Ma, M., Li, T., Wang, H., & Li, C. (2023). Long sequence time-series forecasting with deep learning: A survey. *Information Fusion*, 97, 101819. <https://doi.org/10.1016/j.inffus.2023.101819>
- [33] Hahn, Y., Langer, T., Meyes, R., & Meisen, T. (2023). Time Series Dataset Survey for Forecasting with Deep Learning. *Forecasting*, 5(1), 315–335. <https://doi.org/10.3390/forecast5010017>