ACS MATERIALS LETTERS

www.acsmaterialsletters.org

Impact of Domain Knowledge on the Property Prediction of Specialized Machine Learning Models

Published as part of ACS Materials Letters special issue "Machine Learning for Materials Chemistry".

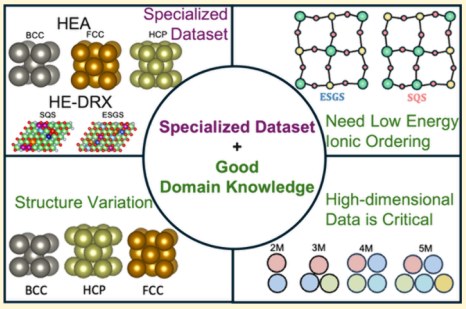
Lin Wang, Tanjin He, and Bin Ouyang*

Cite This: *ACS Materials Lett.* 2025, 7, 2708–2715

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: **Developing transferable machine learning models is trending in data-driven materials research. However, how to apply such models to a specific research domain remains unclear. In this work, we choose high-entropy materials as a platform with a specialized data set containing 145,323 DFT-relaxed materials. This data set is used to explore the role of domain-specific knowledge in training effective models. Our tests with three representative graph neural network architectures indicate the model complexity has much smaller influence on performance than the data itself. Specifically, the consideration of low-energy atomic ordering, structures with diverse elemental coverage, and high-order interactions significantly influences the model performance. We also find that domain knowledge-driven sampling can greatly enhance unsupervised learning techniques. This research highlights that developing specialized data sets is more beneficial than further complicating deep learning architectures. Additionally, physics-inspired sampling algorithms are crucially needed for better machine learning models for a specific materials research domain.**



Pretrained foundation models using large data sets are becoming central to data-driven materials research. Recent work shows that graph neural networks (GNNs) trained on generic materials data sets can reach or exceed chemical accuracy (e.g., 1 kcal/mol or 43.36 meV/atom).[1−6] Despite these significant advancements, the next key challenge is adapting these models to specialized domains. In this study, we addressed this challenge with systematic benchmark on a comprehensive specialized[7−9] data set of 18,810 HE-DRX[10] and 126,513 HEA compositions,[11,12] comparable in size to leading generic data sets[2,13−15] but focused exclusively on high-entropy materials. Benchmarking three popular GNNs revealed that model complexity has a limited impact on prediction accuracy. The uniformly low mean absolute errors (MAEs) suggest that current GNNs are already sufficiently good at capturing both structural and compositional information.[16] Instead, domain knowledge (e.g., ionic ordering, phase behavior, and high-dimensional interactions) plays a much larger role in improving the performance. Additionally, our results show that training data quality matters more than quantity: informed sampling based on physical insight outperforms unsupervised approaches like PCA[17] or UMAP.[18] These findings highlight that effective domain adaptation depends more on embedding material knowledge and smart data selection than on increasing model complexity.

IMPACT OF DATA SIZE AND MODEL VARIATIONS

Three representative GNN models, CGCNN,[6] M3GNet,[4] and CHGNet,[5] are evaluated on both HE-DRX and HEA data sets. CGCNN is the first crystal graph-based convolutional model;[6] M3GNet introduces the first universal interatomic potential trained on Materials Project data;[4] and CHGNet represents a leading universal GNN-based force field.[5] These models effectively represent many other existing efforts. To assess performance across data size, we performed 70%−15%−15% train−validation−test splits using 1%, 2%, 5%, 10%, 20%, 50%, and 100% of each data set, with five independent runs per size to ensure robustness. All data used DFT-relaxed structures, with E_{hull} values as target property.[19−24] Results are shown in Figure 1 and Tables S1−S8, with green dots denoting average

Received: April 28, 2025
Revised: June 24, 2025
Accepted: June 26, 2025

https://doi.org/10.1021/acsmaterialslett.5c00726
ACS Materials Lett. 2025, 7, 2708−2715

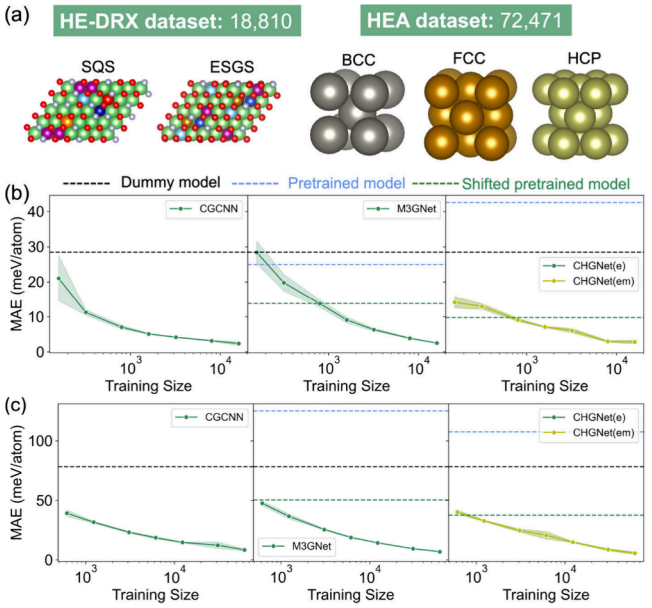


Figure 1. Data set and model performance with partial data set variations. (a) Schematic of the HE-DRX data set and HEA data set and types of structures included. SQS stands for special quasi-random structures, and ESGS stands for electrostatic ground states. (b) Performance of all three models on HE-DRX data set with 0.70−0.15−0.15 splitting in terms of predicting E_{hull}. (c) Performance of all three models on HEA data set with 0.70−0.15−0.15 splitting in terms of predicting E_{hull}. The black dashed lines are a reference of a dummy model that uses the average of a data set to predict all data. The blue and green dashed lines are predicting performance using pretrained models from M3GNet[4] and CHGNet[5] that were trained based on the Materials Project data set. The blue dashed line indicates the performance by directly applying the corresponding model, while the green dashed line is the performance after enforcing the centering of all data on y = x. For CHGNet, the training has been done both without (denoted as e) and with (denoted as em) magnetic moments. The shaded area indicates the error from five independent trainings with different random seeds of selecting samples from our complete data set.

MAE and shaded areas representing variation. A dummy model predicting the data set mean is shown as a black dashed line (Figure S13). We also tested pretrained formation energy (E_{form}) models, M3GNet-MP-2018.6.1-Eform[4] and CHGNet 0.3.0,[5] on our data sets. Their performance is shown as blue and green dashed lines (Figures S14−S15). Although both E_{hull} and E_{form} represent relative energies,[4,5] they are still fundamentally distinct. Green dashed lines include a force-centering shift (y = x) of model prediction as suggested by Deng et al.,[25] while the blue lines reflect direct use of model. To mitigate systematic offsets due to data set bias, we applied the "shifted model" approach, which is considered the upper-bound performance for pretrained models.

It can be seen from Figure 1 that for all GNN models, performance differences diminish with sufficient data (e.g., >1000 DFT-relaxed structures). On the HE-DRX data set, CGCNN, M3GNet, and CHGNet (e and em models) achieve MAEs of 1.95, 2.45, 2.02, and 2.07 meV/atom, respectively; for HEA, the MAEs are 7.62, 6.75, 4.98, and 7.40 meV/atom. Such values are much lower than the direct prediction with pretrained formation energy model (Figures S14−S15). Despite their differing architectural complexities, CGCNN is much simpler without explicit three-body interactions. Their comparable performance suggests that for high-entropy materials, model architecture matters less once enough data is available. Current deep learning models are already sophisticated enough for specialized materials discovery tasks. Even though building a very complex generic model will certainly have other application scenarios, if one only needs to understand the specific design space of materials, further complicating the ML model on generic data sets will not be as beneficial as building a specialized data set and more efficient sampling.

A preliminary fine-tuning from pretrained M3GNet-MP-2018.6.1-Eform[4] using up to 10% of specialized data sets was conducted to predict E_{form} and derive E_{hull}, as shown in Figure S16. The fine-tuned model showed improved E_{hull} predictions, confirming the benefits of fine-tuning. However, to further push the MAE below 10 meV/atom for both HE-DRX and HEA, fine-tuning still requires at least a few thousand data points, e.g., 2000 samples for HE-DRX and 7000 for HEA, while the fine-tuned model shows a similar MAE to the retrained model when trained with this level of data size. These observations suggest that while fine-tuning a pretrained generic model can effectively capture general trends within a specific compositional space, it offers limited benefit in minimizing prediction error. In such cases, increasing the amount of high-quality, domain-specific training data remains the most effective strategy for improving the accuracy. Motivated by these findings, we designed three controlled tasks to systematically investigate how domain knowledge, such as atomic ordering, structural variations, and higher-order interactions, can be leveraged to further enhance GNN-based materials predictions.

IMPACT OF ATOMIC ORDERING

Understanding and distinguishing different ordering states are crucial in high-entropy materials research. The HE-DRX system, with its rich variety of ordering states,[20,26−29] is ideal for testing machine learning models' ability to capture local ordering. We enumerated the compositional space for two classic HE-DRX ordering types, illustrated in Figure 2(a). The ESGS notation, shown by the equation at the top of each schematic, stands for an electrostatic ground state, which is obtained by minimizing the Ewald summation of the structure. Conversely, the SQS structures are generated by minimizing the cluster occupancy difference to approach the random limit.[30] Both structural types are observed in HE-DRX and other ionic systems,[31] correlating to short-range ordered states and fully random states, respectively.

With the selected data set, the machine learning training was designed by splitting the data set in different ways. First, the data set was divided into ESGS-only and SQS-only groups. As shown in Figure 2(b)−(c), we performed 82−18 train-validation splits on ESGS structures and tested these models on SQS structures, denoted as ESGS→SQS. Conversely, SQS structures were used for training and validation, with testing on ESGS structures (SQS→ESGS; Figure 2(b) and Figure 2(d)). Across all models, MAEs were consistently lower for ESGS→SQS, with CGCNN showing the lowest MAE value (Figure 2(c)−(d)). The contrast between Figure 2(c) and Figure 2(d) underscores the importance of domain knowledge in model training. A dummy model predicting ESGS energies by the average achieved a lower MAE (24.2 meV/atom, Supporting Information Figure S17) than all GNN models for the SQS→

https://doi.org/10.1021/acsmaterialslett.5c00726
ACS Materials Lett. 2025, 7, 2708−2715

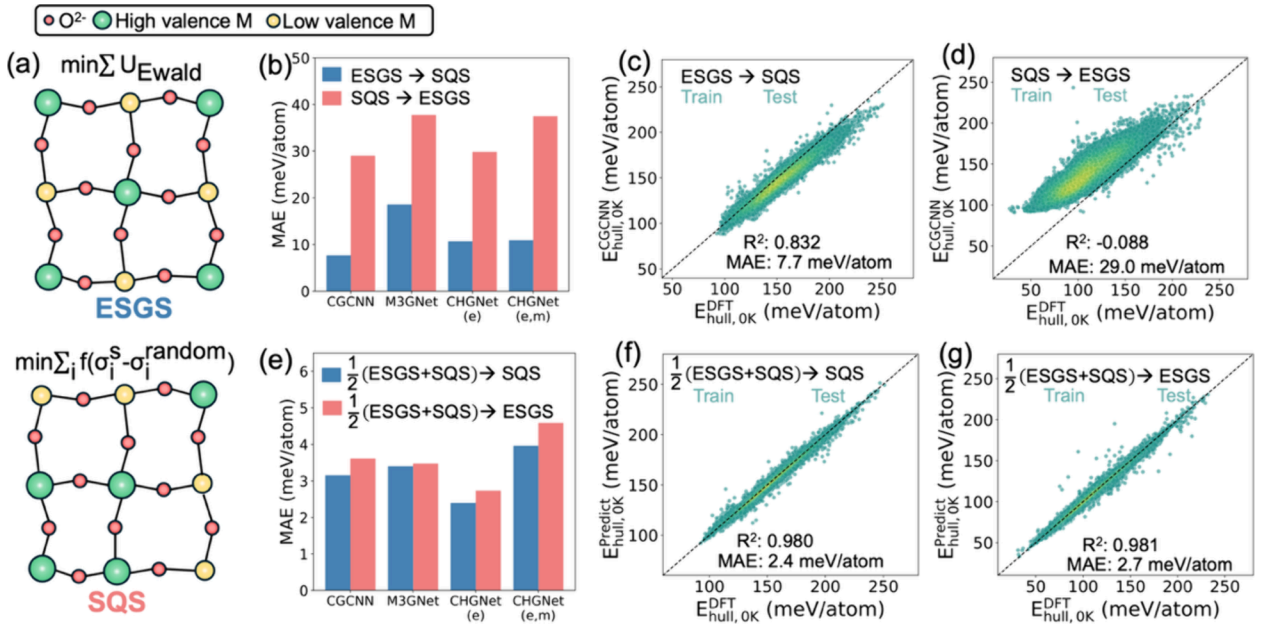


Figure 2. Predicted performance across different atomic orderings. (a) Schematic of how electrostatic ground state (ESGS) structure and special quasi random structure (SQS) are generated. (b) The performance of different GNN models in terms of using one order to predict the other order. The E_{hull} values are used as labels for training. (c, d) The parity plots of all test data sets for the best model (CGCNN) for this task. (e) The performance of different GNN models in terms of using mixed ESGS and SQS as the training data to predict SQS or ESGS structures. (f, g) The parity plots of all test data sets for the best model (CHGNet (e)) for this task.

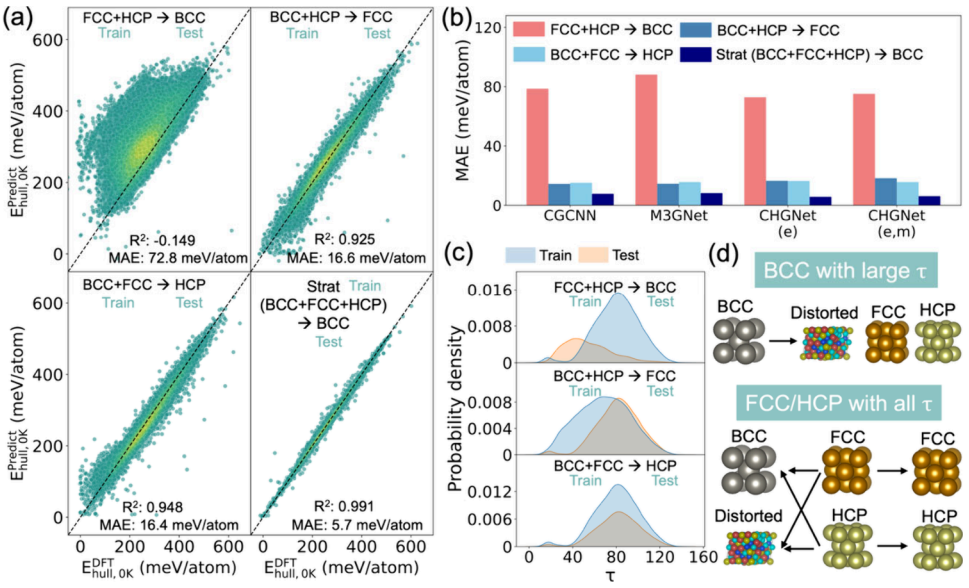


Figure 3. Predicted performance across different crystal structures. (a) Parity plots of all test data sets from the best model (CHGNet, e) for four types of tasks. The specific task is labeled in each panel. The notation "Strat" indicates that stratified sampling has been performed to generate the same data size as the case of FCC+HCP→BCC. (b) The performance of all four models for all four tasks. (c) The distribution of different training and testing data sets as a function of τ. (d) Illustration of the phase transition appears after relaxation for structures with different τ.

ESGS task, highlighting the risk of relying solely on SQS structures. Although SQS structures are often used in ML studies of high-entropy materials,[32−37] our results suggest that incorporating ionic ordering information (e.g., ESGS) is critical.[20,38,39] To further validate this, we created another data set combining half ESGS and half SQS structures, with the same total size as the earlier splits. Testing was done on the remaining unseen structures. As shown in Figure 2(e)−(g), including both ESGS and SQS significantly reduced MAEs across all models. Additional parity plots are provided in Supporting Information Figures S18−S20.

Figure 2 demonstrates that model performance heavily depends on the sampling of local orderings, independent of the architecture. Training on randomly selected or biased structures, even with large data size, can be risky. We hypothesize that ESGS structures offer broader coordination diversity due to charge state effects, whereas SQS structures, reflecting random stoichiometry, offer less local variation as

https://doi.org/10.1021/acsmaterialslett.5c00726
ACS Materials Lett. 2025, 7, 2708−2715

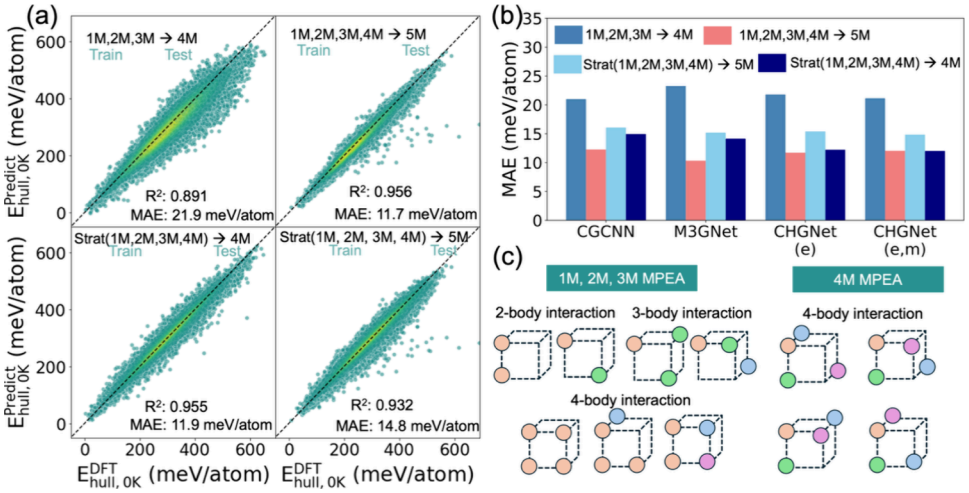


Figure 4. Predicted performance for higher-dimensional interactions. (a) Parity plots of all test data sets from the best model (CHGNet (em)) for four types of tasks. The specific task is labeled in each panel. The notation "Strat" indicates that stratified sampling has been performed to generate the same data size as the case of 1M,2M,3M→4M. (b) The performance of all four models for all four tasks. (c) The illustration of selected cluster interactions in 1M,2M,3M (left panel) and representative unique cluster interactions in 4M MPEAs. Circles with different colors indicate different species.

they do not depend on charge states of ions. Additionally, ESGS structures are often closer to ground states:[31] 95.18% of ESGS structures have lower energy than corresponding SQS structures (Supporting Information Figure S21). Including such structures thus better captures chemical bonding variations[40] and reduces systematic errors.[25]

IMPACT OF CRYSTAL STRUCTURAL VARIATION

Beyond ionic ordering, another major challenge is capturing variation across different crystal structures. To test model transferability across structures, we designed four training schemes (Figure 3(a)). In the first three tasks, models were trained on two structure types (e.g., BCC + FCC) to predict the third (e.g., HCP). Data sizes were balanced by selecting equal numbers of BCC, FCC, and HCP samples with an 82−18 random train-validation split. Multiple random seeds ensured robust evaluation, and detailed results are shown in Supporting Information Table S9. Testing performance, where the model predicts unseen structure types, is summarized in Figure 3(a), with example parity plots for the best model (CHGNet, energy-only) and additional models shown in Supporting Information Figures S22−S24. Notably, models trained on FCC and HCP data perform poorly when predicting BCC structures, yielding a high MAE of 78.5 meV/atom. In contrast, models trained on BCC+HCP to predict FCC and BCC+FCC to predict HCP achieved much lower MAEs of 14.5 and 15.1 meV/atom, respectively. This sharp contrast suggests fundamental differences in the complexity or the learnability of the three structure types.

The divergent behaviors observed in Figure 3(a)−(b) highlight that even large data sets (>10,000 data points) with full elemental coverage do not guarantee useful models. To investigate the physical origin of this divergence, we plotted training and testing data sets against our previously established descriptor τ,[11] defined as $\tau = VEC \times (VEC_{std} + VEC) + \sqrt[3]{\log(X_{avg})}$, where VEC, VEC_{std}, and X_{avg} represent average valence electron concentration, its standard deviation, and average electronegativity, respectively. As shown in Figure 3(c), BCC structures exhibit a distinct τ distribution compared to those of FCC and HCP, reflecting fundamental differences in electronic properties. This divergence stems from mechanical instabilities: many high-τ HEAs initialized as BCC relax into FCC, HCP, or amorphous structures, as summarized in Figure 3(d) and our prior work.[11] Consequently, BCC structures lack high-τ compositions, while FCC and HCP structures span a broader τ range except at very low τ, where they can relax into BCC. This data set artifact explains the poor performance when using only FCC and HCP data to predict BCC structures. Including BCC structures ensures τ-space coverage and significantly improves model performance. This is confirmed by the fourth task in Figure 3(a), where random sampling across all three structure types yields a low MAE of 5.7 meV/atom when predicting unseen BCC data. Additionally, to determine the amount of BCC data needed for accurate BCC predictions, we incrementally added varying fractions of BCC compounds to the FCC and HCP training set and evaluated the model on the remaining BCC. As shown in Figure S25, we found that including approximately 4,500 BCC compounds will reduce the MAE to around 10 meV/atom. This type of analysis—searching for the amount of data needed for efficient modeling—has also been explored in other systems.[41]

IMPACT OF HIGHER-ORDER INTERACTION

Beyond structural phases, another critical challenge in high-entropy materials is the prediction of higher-dimensional chemical spaces. To assess this, we trained models on HEA data containing 1, 2, or 3 metals (82−18 train−validation split) and tested them on 4-metal data. This was compared to training on 1−4 metal data and testing on 5-metal data (top right panel of Figure 4(a)). The best model, CHGNet (em), is shown in the top left panel of Figure 4(a). Training on 1−4 metal data yields significantly better performance (MAE: 11.7 meV/atom) than training on 1−3 metal data only (MAE: 21.9 meV/atom). However, the four-metal data set is much larger (61,425 samples) compared to the combined 1M+2M+3 M data set (11,046 samples). To control for training size, we performed stratified sampling to match data set sizes and then tested on the remaining 4 or 5 M data. The results are shown

https://doi.org/10.1021/acsmaterialslett.5c00726
ACS Materials Lett. 2025, 7, 2708−2715

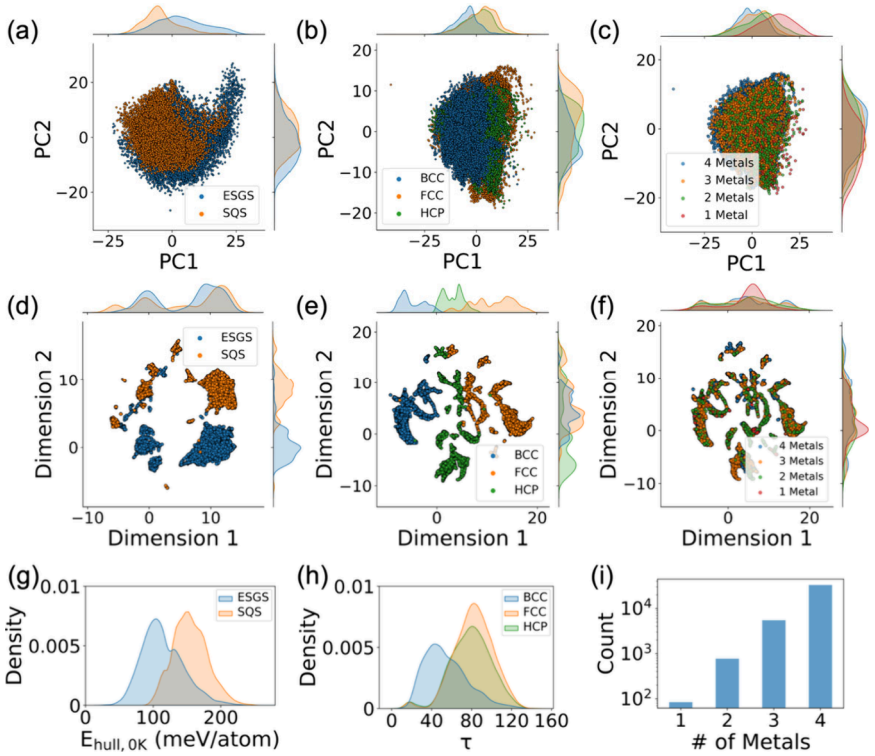


Figure 5. Unsupervised and domain knowledge sampling strategy. (a−c) Principal component analysis (PCA) among compositional and structural features from Matminer[48] to distinguish (a) ESGS structures with SQS structures; (b) BCC, FCC, and HCP structures; and (c) structures with different metals. (d−f) Uniform Manifold Approximation and Projection (UMAP) among compositional and structural features from Matminer[48] to distinguish (d) ESGS structures with SQS structures; (e) BCC, FCC, and HCP structures; and (f) structures with different metals. (g−i) Physics-inspired clustering of data to distinguish (g) ESGS structures with SQS structures; (h) BCC, FCC, and HCP structures; and (i) structures with different numbers of metals.

in the bottom two panels of Figure 4(a), with performance metrics for all four ML models summarized in Figure 4(b) and Supporting Information Figures S26−S28 and Tables S10−S11. Across all tasks, model performances are very similar, reinforcing the key role of data dimensionality and sampling.

Comparison of Strat(1M−4M) → 4M and Strat(1M−4M) → 5M results (bottom panels, Figure 4a) shows that including 4M data significantly improves model accuracy for 4M predictions, reducing the MAE from 21.9 to 11.9 meV/atom. For 5M predictions, the MAE reaches 14.8 meV/atom, comparable to models trained on the full 1M−4M data set. These findings highlight the value of incorporating high-dimensional data to improve predictive performance. As illustrated in Figure 4c, 4M data sets introduce four-species interactions absent in lower-dimensional data, underscoring the benefit of domain knowledge from higher-order interactions in enhancing model capability.

■ SAMPLING STRATEGY

The constantly observed similar performance among Figures 1−4 further confirms the fact that the model complexity[42] makes little difference compared to the training data. Instead, effective sampling strategies are more critical for training a specialized ML model for a particular chemical space. Recent studies[17,18,43] have shown promising advances in using unsupervised learning to enhance data selection, but efficient sampling remains an overlooked challenge, as further illustrated by Figure 5. As shown in Figure 5, we compared physically inspired data clustering with two widely used unsupervised learning-based methods: principal component analysis (PCA)[17] and Uniform Manifold Approximation and Projection (UMAP).[18] PCA and UMAP have been used to cluster materials space to enhance ML sampling, which are benchmarked against domain knowledge-driven clustering, using three key criteria: (i) ordering (ESGS vs SQS), (ii) structural type (BCC, FCC, HCP), and (iii) compositional dimensionality. For PCA and UMAP, standard compositional and structural features were used: structural descriptors involve packing fraction, Ewald energy,[44] maximum packing efficiency,[45] short-range order,[46] XRD diffraction, radial distribution function, structural complexity[47]) while compositional features are adopted from Matminer[48] (details in computational method and Figures S29−S30). As a result, the featurization process generates 351 structural features for HE-DRX data set and 362 structural features for HEA data set, while 187 compositional features are generated for both data sets.

Figure 5(a)−(c) shows that PCA fails to form meaningful clusters, even when important physical features like Ewald summation, valence electron concentration, and element count are included. This failure stems from PCA's linear nature, which struggles with intrinsically nonlinear relationships, as also noted by Li et al.[18] In contrast, UMAP, a nonlinear method, shows improved clustering, particularly separating ESGS and SQS structures effectively (Figure 5(d), Figure S31). However, UMAP still has its limitations. As shown in Figure 5(e), while it distinguishes BCC, FCC, and HCP domains, it does not capture subtler relationships explaining why FCC+HCP→BCC predictions outperform others. UMAP also fails to cluster structures based on the number of elements

https://doi.org/10.1021/acsmaterialslett.5c00726
ACS Materials Lett. 2025, 7, 2708−2715

(Figure 5(f)). In comparison, domain-knowledge-based clustering (Figure 5(g)−(i)) effectively organizes structures by ordering, phase type, and compositional dimensionality, enabling more efficient clustering of the important physical factors pointed out in Figures 2−4.

Overall, we present a comprehensive investigation showing that for high-entropy materials, specialized data set design and domain knowledge in sampling are far more critical than model complexity. This work lays a foundation for more effective machine learning strategies in both high-entropy and broader materials science research.

COMPUTATIONAL METHODS

All electronic structure calculations were performed using Vienna Ab-initio Simulation Package (VASP)[49−51] with PBE functional[52] and PAW pseudopotential.[53,54] A reciprocal space discretization of 25 k-points per $Å^{-1}$ was used to sample the Brillouin zone.[55] The kinetic energy cutoff is 520 eV. The Methfessel-Paxton scheme is used to smooth the partial occupancies for each orbital.[56] The smearing width is 0.2 eV. The convergence criteria for the self-consistent field (SCF) step are 10^{-5} eV. Geometric optimizations were executed until the force on each atom is less than 0.05 eV/Å. The details of our DFT data set are in the Supporting Information. For all three ML architectures used (e.g., CGCNN, M3GNET, and CHGNet), a two-body cutoff radius of 8 Å and a three-body cutoff of 4 Å were applied during graph construction. Details are in Supporting Information (Detailed Computational Methods and Figures S3−S12).

ASSOCIATED CONTENT

Data Availability Statement

The data sets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request. The authors are also in the process of opening source of all raw data given it takes extra effort for such an enormous data set.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsmaterialslett.5c00726.

Detailed computational method; benchmark results for the architectures in CGCNN, M3GNet, and CHGNet, and parity plots showing the prediction performance of CGCNN, M3GNet, and CHGNet across different types of domain knowledge (including various ionic orderings, structural variations, and higher-dimensional interactions) (PDF)

AUTHOR INFORMATION

Corresponding Author

Bin Ouyang − *Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32304, United States;* orcid.org/0000-0002-8181-6815; Email: bouyang@fsu.edu

Authors

Lin Wang − *Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32304, United States*

Tanjin He − *Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, United States;* orcid.org/0000-0001-8834-8703

Complete contact information is available at: https://pubs.acs.org/10.1021/acsmaterialslett.5c00726

Author Contributions

CRediT: **Lin Wang** investigation, methodology, resources, software; **Tanjin He** formal analysis, methodology, validation; **Bin Ouyang** conceptualization, formal analysis, funding acquisition, resources, supervision, validation, writing - original draft, writing - review & editing.

Notes

The authors declare no competing financial interest.

A preprint version of this manuscript has been previously posted:[57] Wang, L.; He, T.; Ouyang, B. The Impact of Domain Knowledge on Universal Machine Learning Models. 2024, Version 1. *ChemRxiv*. 10.26434/chemrxiv-2024-fmq8p (accessed August 05, 2024).

ACKNOWLEDGMENTS

L.W. and B.O. would like to acknowledge funding support from startup funding from Florida State University. The Computational resources are provided by the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS), the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science and the U.S. Department of Energy under contract no. DE-AC02-05CH11231 and Research Computing Center (RCC) at Florida State University. The computation and data processing are also supported by the supercomputing resources from the Department of Energy's Office of Energy Efficiency and Renewable Energy at the National Renewable Energy Laboratory. Argonne National Laboratory's contribution is based upon work supported by Laboratory Directed Research and Development (LDRD) funding from Argonne National Laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy under contract no. DE-AC02-06CH11357.

REFERENCES

(1) Batatia, I.; Benner, P.; Chiang, Y.; Elena, A. M.; Kovács, D. P.; Riebesell, J.; Advincula, X. R.; Asta, M.; Avaylon, M.; Baldwin, W. J. et al. A foundation model for atomistic materials chemistry. 2023, *arXiv*, 2401.00096. https://ui.adsabs.harvard.edu/abs/2024arXiv240100096B (accessed December 01, 2023).

(2) Yang, H.; Hu, C.; Zhou, Y.; Liu, X.; Shi, Y.; Li, J.; Li, G.; Chen, Z.; Chen, S.; Zeni, C. et al. MatterSim: A Deep Learning Atomistic Model Across Elements, Temperatures and Pressures. *arXiv*, 2024, 2405.04967 http://arxiv.org/abs/2405.04967 (accessed May 08 2024).

(3) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling deep learning for materials discovery. *Nature* **2023**, *624*, 80−85.

(4) Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2022**, *2*, 718−728.

(5) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* **2023**, *5*, 1031−1041.

(6) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120*, 145301.

(7) Liu, X.; Ouyang, B.; Zeng, Y. Balancing autonomy and expertise in autonomous synthesis laboratories. *Nat. Comput. Sci.* **2025**, *5*, 92−94.

(8) Zeng, Y.; Ouyang, B.; Liu, J.; Byeon, Y. W.; Cai, Z.; Miara, L. J.; Wang, Y.; Ceder, G. High-entropy mechanism to boost ionic conductivity. *Science* **2022**, *378*, 1320−1324.
(9) Ouyang, B.; Zeng, Y. The rise of high-entropy battery materials. *Nat. Commun.* **2024**, *15*, 973.
(10) Wang, L.; Sunariwal, N.; He, Y.; Kim, D. h.; Yeon, D. h.; Zeng, Y.; Cabana, J.; Ouyang, B. Elemental Stability Rules for High Entropy Disordered Rocksalt Type Li-Ion Battery Positive Electrodes. *Adv. Energy Mater.* **2025**, *15*, 2404982.
(11) Wang, L.; Ouyang, B. Phase Selection Rules of Multi-Principal Element Alloys. *Adv. Mater.* **2024**, *36*, No. e2307860.
(12) Wang, L.; Zhang, Y.; James, K.; Ouyang, B. General Model for d-Center Prediction in Multi-Principal-Element Alloys. *J. Am. Chem. Soc.* **2025**, DOI: 10.1021/jacs.5c05649.
(13) Barroso-Luque, L.; Shuaibi, M.; Fu, X.; Wood, B. M.; Dzamba, M.; Gao, M.; Rizvi, A.; Zitnick, C. L.; Ulissi, Z. W. Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv*, 2410.12771, 2024.
(14) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials* **2013**, *1*, 011002.
(15) Schmidt, J.; Wang, H. C.; Cerqueira, T. F. T.; Botti, S.; Marques, M. A. L. A dataset of 175k stable and metastable materials calculated with the PBEsol and SCAN functionals. *Sci. Data* **2022**, *9*, 64.
(16) Bartel, C. J.; Trewartha, A.; Wang, Q.; Dunn, A.; Jain, A.; Ceder, G. A critical examination of compound stability predictions from machine-learned formation energies. *npj Computational Materials* **2020**, *6*, 97.
(17) Qi, J.; Ko, T. W.; Wood, B. C.; Pham, T. A.; Ong, S. P. Robust training of machine learning interatomic potentials with dimensionality reduction and stratified sampling. *npj Computational Materials* **2024**, *10*, 43.
(18) Li, K.; DeCost, B.; Choudhary, K.; Greenwood, M.; Hattrick-Simpers, J. A critical examination of robustness and generalizability of machine learning prediction of materials properties. *npj Computational Materials* **2023**, *9*, 55.
(19) Ouyang, B.; Wang, J.; He, T.; Bartel, C. J.; Huo, H.; Wang, Y.; Lacivita, V.; Kim, H.; Ceder, G. Synthetic accessibility and stability rules of NASICONs. *Nat. Commun.* **2021**, *12*, 5752.
(20) Lun, Z.; Ouyang, B.; Kwon, D. H.; Ha, Y.; Foley, E. E.; Huang, T. Y.; Cai, Z.; Kim, H.; Balasubramanian, M.; Sun, Y.; et al. Cation-disordered rocksalt-type high-entropy cathodes for Li-ion batteries. *Nat. Mater.* **2021**, *20*, 214−221.
(21) Ong, S. P.; Wang, L.; Kang, B.; Ceder, G. Li-Fe-P-O2 Phase Diagram from First Principles Calculations. *Chem. Mater.* **2008**, *20*, 1798−1807.
(22) Wang, L.; Wang, J.; Ouyang, B. Computational Investigation of MAX as Intercalation Host for Rechargeable Aluminum-Ion Battery. *Adv. Energy Mater.* **2023**, *13*, 2302584.
(23) Sun, W.; Dacek, S. T.; Ong, S. P.; Hautier, G.; Jain, A.; Richards, W. D.; Gamst, A. C.; Persson, K. A.; Ceder, G. The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2016**, *2*, No. e1600225.
(24) He, Y.; He, Z.; Ouyang, B. Design principle of disordered rocksalt type overlithiated anode for high energy density batteries. *Mater. Horiz* **2024**, *11*, 6049−6056.
(25) Deng, B.; Choi, Y.; Zhong, P.; Riebesell, J.; Anand, S.; Li, Z.; Jun, K.; Persson, K. A.; Ceder, G. Systematic softening in universal machine learning interatomic potentials. *npj Computational Materials* **2025**, *11*, 9.
(26) Wang, L.; He, Z.; Ouyang, B. Data driven design of compositionally complex energy materials. *Comput. Mater. Sci.* **2023**, *230*, 112513.
(27) Cai, Z.; Zhang, Y. Q.; Lun, Z.; Ouyang, B.; Gallington, L. C.; Sun, Y.; Hau, H. M.; Chen, Y.; Scott, M. C.; Ceder, G. Thermodynamically Driven Synthetic Optimization for Cation-Disordered Rock Salt Cathodes. *Adv. Energy Mater.* **2022**, *12*, 2103923.
(28) Ouyang, B.; Artrith, N.; Lun, Z.; Jadidi, Z.; Kitchaev, D. A.; Ji, H.; Urban, A.; Ceder, G. Effect of Fluorination on Lithium Transport and Short-Range Order in Disordered-Rocksalt-Type Lithium-Ion Battery Cathodes. *Adv. Energy Mater.* **2020**, *10*, 1903240.
(29) Lun, Z.; Ouyang, B.; Cai, Z.; Clément, R. J.; Kwon, D.-H.; Huang, J.; Papp, J. K.; Balasubramanian, M.; Tian, Y.; McCloskey, B. D.; et al. Design Principles for High-Capacity Mn-Based Cation-Disordered Rocksalt Cathodes. *Chem.* **2020**, *6*, 153−168.
(30) Zunger, A.; Wei, S.; Ferreira, L. G.; Bernard, J. E. Special quasirandom structures. *Phys. Rev. Lett.* **1990**, *65*, 353−356.
(31) Li, L.; Ouyang, B.; Lun, Z.; Huo, H.; Chen, D.; Yue, Y.; Ophus, C.; Tong, W.; Chen, G.; Ceder, G.; et al. Atomic-scale probing of short-range order and its impact on electrochemical properties in cation-disordered oxide cathodes. *Nat. Commun.* **2023**, *14*, 7448.
(32) Meng, H.; Yu, R.; Tang, Z.; Wen, Z.; Chu, Y. Formation ability descriptors for high-entropy carbides established through high-throughput methods and machine learning. *Cell Reports Physical Science* **2023**, *4*, 101512.
(33) Zhang, J.; Xu, B.; Xiong, Y.; Ma, S.; Wang, Z.; Wu, Z.; Zhao, S. Design high-entropy carbide ceramics from machine learning. *npj Computational Materials* **2022**, *8*, 5.
(34) Levämäki, H.; Tasnádi, F.; Sangiovanni, D. G.; Johnson, L. J. S.; Armiento, R.; Abrikosov, I. A. Predicting elastic properties of hard-coating alloys using ab-initio and machine learning methods. *npj Computational Materials* **2022**, *8*, 17.
(35) Guo, W.; Wang, L.; Zhu, L. a.; Ye, Y.; Zhang, Z.; Yang, B.; Zhang, S.; Bai, S. Accelerated discovery and formation mechanism of high-entropy carbide ceramics using machine learning based on low-cost descriptors. *J. Alloys Compd.* **2024**, *1004*, 175929.
(36) Rittiruam, M.; Setasuban, S.; Noppakhun, J.; Saelee, T.; Ektarawong, A.; Aumnongpho, N.; Boonchuay, S.; Khajondetchairit, P.; Praserthdam, S.; Alling, B.; et al. First-Principles Density Functional Theory and Machine Learning Technique for the Prediction of Water Adsorption Site on PtPd-Based High-Entropy-Alloy Catalysts. *Advanced Theory and Simulations* **2023**, *6*, 2200926.
(37) Zhu, S.; Sarıtürk, D.; Arróyave, R. Accelerating CALPHAD-based phase diagram predictions in complex alloys using universal machine learning potentials: Opportunities and challenges. *Acta Mater.* **2025**, *286*, 120747.
(38) Zhang, R.; Zhao, S.; Ding, J.; Chong, Y.; Jia, T.; Ophus, C.; Asta, M.; Ritchie, R. O.; Minor, A. M. Short-range order and its impact on the CrCoNi medium-entropy alloy. *Nature* **2020**, *581*, 283−287.
(39) Zhang, R.; Zhao, S.; Ophus, C.; Deng, Y.; Vachhani, S. J.; Ozdol, B.; Traylor, R.; Bustillo, K. C.; Morris, J. W., Jr; Chrzan, D. C.; et al. Direct imaging of short-range order and its impact on deformation in Ti-6Al. *Sci. Adv.* **2019**, *5*, No. eaax2799.
(40) Huang, W.; Urban, A.; Rong, Z.; Ding, Z.; Luo, C.; Ceder, G. Construction of ground-state preserving sparse lattice models for predictive materials simulations. *npj Computational Materials* **2017**, *3*, 30.
(41) Huang, L.; Ling, C. Practicing deep learning in materials science: An evaluation for predicting the formation energies. *J. Appl. Phys.* **2020**, *128*, 124901.
(42) Riebesell, J.; Goodall, R. E. A.; Benner, P.; Chiang, Y.; Deng, B.; Lee, A. A.; Jain, A.; Persson, K. A. Matbench Discovery - A framework to evaluate machine learning crystal stability predictions. 2023, *arXiv*, 2308.14920. https://ui.adsabs.harvard.edu/abs/2023arXiv230814920R (accessed August 01, 2023).
(43) Rondinelli, J. M.; Chen, W. ET-AL: Entropy-targeted active learning for bias mitigation in materials data. *Applied Physics Reviews* **2023**, *10*, 021403.
(44) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **1921**, *369*, 253−287.
(45) Ward, L.; Liu, R.; Krishna, A.; Hegde, V. I.; Agrawal, A.; Choudhary, A.; Wolverton, C. Including crystal structure attributes in

machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B* **2017**, *96*, 024104.
(46) Zimmermann, N. E. R.; Horton, M. K.; Jain, A.; Haranczyk, M. Assessing Local Structure Motifs Using Order Parameters for Motif Recognition, Interstitial Identification, and Diffusion Path Characterization. *Frontiers in Materials* **2017**, DOI: 10.3389/fmats.2017.00034.
(47) Krivovichev, S. V. Structural complexity of minerals: information storage and processing in the mineral world. *Mineralogical Magazine* **2013**, *77*, 275−326.
(48) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; et al. Matminer: An open source toolkit for materials data mining. *Comput. Mater. Sci.* **2018**, *152*, 60−69.
(49) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15−50.
(50) Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B Condens Matter* **1994**, *49*, 14251−14269.
(51) Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B Condens Matter* **1993**, *47*, 558−561.
(52) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(53) Blochl, P. E. Projector augmented-wave method. *Phys. Rev. B Condens Matter* **1994**, *50*, 17953−17979.
(54) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758−1775.
(55) Monkhorst, H. J.; Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **1976**, *13*, 5188−5192.
(56) Methfessel, M.; Paxton, A. T. High-precision sampling for Brillouin-zone integration in metals. *Phys. Rev. B Condens Matter* **1989**, *40*, 3616−3621.
(57) Wang, L.; He, T.; Ouyang, B. The Impact of Domain Knowledge on Universal Machine Learning Models. *ChemRxiv*, 2024, DOI: 10.26434/chemrxiv-2024-fmq8p (accessed August 05, 2024).

https://doi.org/10.1021/acsmaterialslett.5c00726
ACS Materials Lett. 2025, 7, 2708−2715