# Machine Learning-Assisted Prediction of Gas Adsorption and Thermo-Physical Properties in Advanced Carbon-Based Materials

Wesley Jimenez
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
wesley540@unr.edu

Harish Joshi
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
hjoshi@oregonstate.edu

Zhi Wei
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
zhiw@uc.edu

Aapo Warner
Department of Computer Science, University of Houston, Houston, TX, USA.
aapowork@uh.edu

**Abstract**

The prediction of gas adsorption and thermo-physical properties in advanced carbon-based materials, such as carbon foams, graphene derivatives, and hexagonal boron nitride hybrids, has become a central challenge in the design of next-generation sorbents and thermal management systems. Traditional experimental and simulation-based approaches are constrained by high computational costs, limited throughput, and the combinatorial complexity of material space. Machine learning offers a transformative paradigm by enabling rapid, accurate, and scalable property prediction from structural and compositional descriptors. This paper presents a system-level analysis of machine learning-assisted prediction frameworks, focusing on the architectural, infrastructural, and governance dimensions that underpin their deployment. We examine the trade-offs between representation learning techniques, uncertainty quantification, and data quality in the context of gas adsorption on doped and defective carbon surfaces. We further explore the integration of thermophysical property prediction into broader socio-technical infrastructures, including material informatics databases, open science policies, and reproducibility standards. Drawing on recent advances in first-principles calculations and experimental characterization, we discuss how machine learning models can bridge the gap between atomistic simulations and engineering applications. The paper also critically assesses issues of algorithmic fairness, robustness against out-of-distribution data, and the policy implications of using predictive models in safety-critical contexts such as toxic gas capture and thermal protection systems. By synthesizing insights from computational materials science, data engineering, and governance studies, we propose a framework for sustainable, equitable, and reliable deployment of machine learning in carbon-based material design.

**Keywords**

machine learning, gas adsorption, thermo-physical properties, carbon-based materials, material informatics, data infrastructure, governance, robustness, fairness, hexagonal boron nitride, carbon foam, predictive modeling.

## 1. Introduction

The development of advanced carbon-based materials for gas separation, storage, and thermal management has accelerated in recent decades, driven by the need for efficient carbon capture, toxic gas protection, and lightweight thermal insulation. Materials such as activated carbon, carbon foams, graphene oxide frameworks, and doped hexagonal boron nitrides exhibit remarkable tunability in their adsorption selectivity and thermal conductivity, yet the experimental exploration of their vast composition-structure-property space remains prohibitively time-consuming. Computational screening via density functional theory and molecular dynamics offers physical accuracy but suffers from high computational costs that limit the number of configurations that can be evaluated. Machine learning has emerged as a complementary approach that can learn complex structure-property relationships from existing data, enabling high-throughput prediction and rational design [1,2]. However, the successful integration of machine learning into materials research requires careful consideration of the entire system: from data generation and curation to model selection, validation, deployment, and governance.

This paper adopts a systems perspective to examine the role of machine learning in predicting gas adsorption and thermo-physical properties of carbon-based materials. We focus not only on algorithmic advances but also on the infrastructural, ethical, and policy dimensions that determine practical impact. The structure of the paper is as follows. Section 2 provides a theoretical background on gas adsorption mechanisms and thermo-physical properties relevant to carbon-based materials. Section 3 discusses machine learning architectures commonly employed, including graph neural networks, random forests, and Gaussian processes, and highlights the trade-offs between interpretability and accuracy. Section 4 addresses data infrastructure, including database curation, metadata standards, and the challenge of small datasets in materials science. Section 5 examines deployment considerations in both academic and industrial settings, including scalability, real-time prediction, and integration with experimental workflows. Section 6 tackles robustness, fairness, and policy implications, emphasizing the need for uncertainty quantification and bias mitigation when models are used to guide safety-critical decisions. Section 7 presents case studies and cross-domain comparisons, including the prediction of gas adsorption on doped boron nitrides and the thermo-physical characterization of carbon foams. Section 8 outlines future directions, followed by a conclusion that synthesizes the main findings.

## 2. Theoretical Framework: Gas Adsorption and Thermo-Physical Properties

The adsorption of gases such as carbon dioxide, ammonia, hydrogen cyanide, and chlorine on carbon-based surfaces is governed by a combination of physisorption and chemisorption mechanisms, which depend on the electronic structure, pore topology, and surface functionalization of the material [3]. In hexagonal boron nitride (h-BN) systems, the introduction of metal dopants or cyclic carbon-metal hybrid sites can significantly alter the binding energies and charge transfer characteristics, thereby enhancing selectivity for specific gas molecules [4]. Similarly, carbon foams—porous materials derived from coal tar pitch, petroleum pitch, or biomass—exhibit thermo-physical properties such as thermal conductivity,

specific heat capacity, and thermal diffusivity that are highly sensitive to the raw material and processing conditions [5]. Understanding these properties is essential for applications in thermal protection systems, catalysis, and gas storage.

Traditional computational methods rely on first-principles density functional theory to compute adsorption energies and electronic density of states, but such calculations are limited to a few hundred structures per study. For example, a systematic first-principles investigation of gas adsorption on doped h-BN sheets requires evaluating multiple dopants, gas molecules, and adsorption sites, quickly leading to a combinatorial explosion [6]. Moreover, thermo-physical property prediction often requires molecular dynamics simulations that capture phonon transport and thermal boundary resistance, which are computationally intensive. Machine learning models can learn the mapping from structural descriptors (e.g., coordination numbers, bond lengths, pore size distributions) to target properties, enabling orders-of-magnitude speedup while maintaining acceptable accuracy [7]. However, the reliability of such models depends critically on the quality and diversity of the training data, as well as the choice of representation.

## 3. Machine Learning Architectures for Property Prediction

A variety of machine learning architectures have been applied to gas adsorption and thermo-physical property prediction in carbon-based materials, each with distinct advantages and limitations. Random forests and gradient boosting methods are widely used for their interpretability and robustness against overfitting when working with tabular data derived from experimental measurements or coarse-grained descriptors [8]. These models can handle mixed numerical and categorical features, such as dopant type, temperature, and pressure, but their performance saturates when the descriptor space becomes high-dimensional or when the underlying physics is highly non-linear.

Deep neural networks, particularly graph neural networks, have become increasingly popular because they can directly learn from atomic coordinates and bonding topologies without manual feature engineering [9]. Graph neural networks operate on the molecular graph representation, where nodes represent atoms and edges represent bonds, and they iteratively update node embeddings through message passing. This architecture captures local chemical environments and long-range interactions, which are critical for adsorption energy prediction on doped surfaces. Nevertheless, graph neural networks require large training sets—often thousands of structures—to generalize well, and they are prone to overfitting on small datasets common in materials research. Gaussian process regression offers a Bayesian alternative that provides built-in uncertainty estimates, which is invaluable for active learning and safe decision-making [10]. The trade-off is reduced scalability to large datasets, as Gaussian processes scale cubically with the number of training points.

The choice of architecture must also consider the nature of the property being predicted. Gas adsorption energies are usually single-value regression targets, while thermo-physical properties such as thermal conductivity may vary with temperature and require models that can capture temperature dependence. Multi-task learning, where a single model predicts multiple correlated properties simultaneously, can improve data efficiency and generalization [11]. Hybrid approaches that combine physics-based knowledge (e.g., empirically derived potentials) with machine learning corrections are also emerging as a promising avenue, as they embed physical constraints while leveraging data-driven flexibility [12].

## 4. Data Infrastructure and Governance

The success of machine learning in material science hinges on the availability of large, well-curated, and interoperable datasets. Data infrastructure involves the collection, storage, annotation, and dissemination of experimental and computational property data. Public repositories such as the Materials Project, the Open Quantum Materials Database, and the NIST Thermo-Physical Properties Database provide foundational resources, but they often lack detailed metadata on synthesis conditions, sample purity, and measurement uncertainties [13]. For carbon-based materials, additional challenges arise from the variety of allotropes (e.g., graphene, carbon nanotubes, porous carbons) and the sensitivity of properties to defect density and surface chemistry.

Governance of data includes establishing standards for data provenance, versioning, and citation. Without rigorous data governance, machine learning models trained on aggregated data may inherit systematic biases or errors from disparate sources. For example, adsorption energies computed with different exchange-correlation functionals in density functional theory can vary by several kilojoules per mole, and a model trained on mixed functional data may produce inconsistent predictions [14]. Therefore, the materials informatics community has advocated for the use of ontologies and minimal information models, such as the Chemical Information Ontology and the Materials Science Common Data Dictionary, to ensure machine readability and comparability [15]. Furthermore, data sharing policies that incentivize open access while protecting intellectual property are needed to accelerate progress. Federated learning frameworks, where models are trained across multiple institutions without centralizing raw data, offer a compromise for proprietary datasets [16].

## 5. Deployment and Scalability

Deploying machine learning models for gas adsorption and thermo-physical property prediction in real-world settings requires attention to computational efficiency, latency, and integration with existing experimental or process control systems. In industrial contexts, models may be used for real-time monitoring of adsorption columns or for guiding the synthesis of new carbon foams with tailored thermal properties. Scalability concerns arise when models must process thousands of candidate materials or operate on resource-constrained edge devices. Model compression techniques, such as pruning, quantization, and knowledge distillation, can reduce inference time and memory footprint while preserving accuracy [17].

Another deployment challenge is the distribution shift between training data and operational conditions. For instance, a model trained on adsorption data at room temperature and low pressure may perform poorly at elevated temperatures or high pressures encountered in industrial capture processes. Continual learning and domain adaptation strategies are necessary to maintain model performance over time as new data becomes available [18]. Additionally, the integration of machine learning with experimental high-throughput automation—often termed self-driving laboratories—requires careful orchestration of data flows, model retraining cycles, and uncertainty-driven experimental design. The governance of such autonomous loops raises questions about accountability and human oversight, particularly when models propose novel but untested material compositions.

## 6. Robustness, Fairness, and Policy Implications

Robustness in machine learning-assisted property prediction refers to the model's ability to maintain accurate and reliable outputs when faced with noisy inputs, sensor drift, or out-of-distribution samples. For safety-critical applications such as the adsorption of toxic gases (e.g.,

hydrogen cyanide, chlorine), a model that overestimates adsorption capacity could lead to insufficient protection in emergency response scenarios [19]. Uncertainty quantification becomes essential: Bayesian methods, ensemble approaches, and conformal prediction provide confidence intervals that flag predictions with low certainty for human review. In materials design, robustness also encompasses adversarial robustness, where small perturbations in input descriptors should not drastically change predicted properties. Current research on adversarial training in graph neural networks has shown promise, but its application in materials science remains limited [20].

Fairness concerns in this domain relate to the equitable distribution of research benefits and the avoidance of biases that disadvantage certain material classes or processing routes. For example, if training data overrepresents noble metal dopants or high-temperature synthesis conditions, the resulting model may systematically underpredict performance for earth-abundant alternatives or low-energy processing methods [21]. Such biases can inadvertently reinforce unsustainable practices or limit the exploration of environmentally benign materials. Policymakers and funding agencies must consider guidelines that require diverse data sampling, transparent reporting of model limitations, and inclusion of underrepresented material families in benchmark datasets.

The policy implications extend to the regulatory landscape for materials used in air purification, carbon capture, and thermal management. If machine learning predictions are used as evidence for material certification or safety approvals, standards for model validation and uncertainty reporting must be established. International organizations such as the International Organization for Standardization and the National Institute of Standards and Technology are beginning to develop frameworks for the verification and validation of computational models in materials science [22]. These efforts need to be adapted to the unique characteristics of machine learning, including the risk of overfitting, the black-box nature of deep networks, and the difficulty of reproducing results across different software stacks.

## 7. Case Studies and Cross-Domain Comparisons

To illustrate the concepts discussed, we examine two representative case studies. The first involves the prediction of gas adsorption properties of metal-doped hexagonal boron nitrides. Recent first-principles studies have systematically computed the adsorption energies of carbon monoxide, ammonia, hydrogen cyanide, cyanogen chloride, and chlorine on various doped h-BN surfaces [11]. These calculations reveal that the introduction of cyclic carbon-metal defects significantly alters the electronic band structure and charge redistribution, leading to enhanced chemisorption for certain gases. Machine learning models trained on these first-principles data can predict adsorption energies for unseen dopants and gas combinations with mean absolute errors below 0.2 eV, enabling rapid screening of thousands of candidate materials [23]. The success of these models depends on the use of graph-based representations that capture the local environment around the doping site and the adsorbate.

The second case study focuses on the thermo-physical properties of carbon foams produced from different raw materials, such as coal tar pitch and biomass-derived precursors. Experimental characterization of thermal conductivity and specific heat capacity across a range of densities and processing temperatures reveals a strong correlation between precursor chemistry and final properties [5]. Machine learning models using features such as elemental composition, pyrolysis temperature, and porosity can predict thermal conductivity with R-squared values exceeding 0.85, allowing manufacturers to tune processing parameters for

target applications. However, the limited number of experimental data points (often fewer than 100) necessitates the use of regularization techniques and transfer learning from related material classes, such as carbon aerogels and graphitic foams [24].

Cross-domain comparisons with other porous materials, such as metal-organic frameworks and zeolites, highlight the generalizability of machine learning approaches. While the structural complexity of carbon-based materials is lower than that of metal-organic frameworks, the strong dependence on defect chemistry and surface functionalization requires careful descriptor engineering. Transfer learning from large databases of metal-organic frameworks to carbon systems has shown mixed results, suggesting that domain-specific pretraining is often necessary [25]. Conversely, models trained on carbon data have been successfully applied to predict adsorption in boron nitride nanotubes, indicating that electronic similarity can enable cross-material transfer.

## 8. Future Directions

The integration of machine learning with first-principles and experimental methods is advancing rapidly, but several directions remain underdeveloped. First, the incorporation of dynamic properties such as diffusion coefficients and heat transfer rates under non-equilibrium conditions into machine learning frameworks will be essential for designing materials for real-world processes. Second, the development of foundation models pre-trained on large corpora of crystal structures and molecular configurations—analogous to large language models—could provide a powerful starting point for property prediction across diverse carbon-based systems. Third, the combination of active learning with automated synthesis and characterization platforms (self-driving labs) offers a path toward autonomous discovery of optimal materials for gas capture and thermal management, but requires robust algorithms for experimental design and uncertainty management. Finally, the ethical and policy dimensions discussed in this paper call for interdisciplinary collaboration between material scientists, computer scientists, ethicists, and regulators to ensure that machine learning tools serve societal goals without exacerbating inequalities or compromising safety.

## 9. Conclusion

Machine learning has become an indispensable tool for predicting gas adsorption and thermo-physical properties in advanced carbon-based materials. Its ability to accelerate discovery and optimization is contingent upon a holistic systems approach that addresses data infrastructure, model architecture, deployment scalability, robustness, fairness, and policy. By examining the trade-offs between model complexity and interpretability, the challenges of small and heterogeneous datasets, and the governance of autonomous workflows, this paper has argued that the successful integration of machine learning into materials research requires more than algorithmic innovation. It demands a socio-technical infrastructure that promotes reproducibility, equity, and safety. The case studies on doped hexagonal boron nitrides and carbon foams illustrate both the potential and the pitfalls of current approaches. As the field moves toward larger, more autonomous prediction systems, continued attention to the structural, ethical, and policy dimensions will be critical to realizing the full benefits of machine learning for carbon-based material design.

## References

1. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. Nature, 559(7715), 547-555.

2. Raccuglia, P., Elbert, K. C., Adler, P. D. F., Falk, C., Wenny, M. B., Mollo, A., ... & Zeller, M. (2016). Machine-learning-assisted materials discovery using failed experiments. Nature, 533(7601), 73-76.

3. Sing, K. S. W., Everett, D. H., Haul, R. A. W., Moscou, L., Pierotti, R. A., Rouquerol, J., & Siemieniewska, T. (1985). Reporting physisorption data for gas/solid systems with special reference to the determination of surface area and porosity. Pure and Applied Chemistry, 57(4), 603-619.

4. Zhang, M., Wang, C., Yu, W., Zhu, Y., & Zhu, P. (2024). Tuning gas adsorption in hexagonal boron nitride by metal and cyclic carbon-metal doping: A first-principles perspective. Physical Chemistry Chemical Physics, 26(12), 9123-9137.

5. Al-Majali, M. R., Zhang, M., Al-Majali, Y. T., & Trembly, J. P. (2025). Impact of raw material on thermo-physical properties of carbon foam. The Canadian Journal of Chemical Engineering, 103(3), 1309-1318.

6. Sun, W., & Ceder, G. (2017). Bayesian optimization for materials design. Journal of the American Chemical Society, 139(38), 13341-13348.

7. Ward, L., Agrawal, A., Choudhary, A., & Wolverton, C. (2016). A general-purpose machine learning framework for predicting properties of inorganic materials. npj Computational Materials, 2, 16028.

8. Fernández, M., & Barnard, A. S. (2019). Identification of structural features that drive colloidal stability in gold nanoparticles using machine learning. Journal of Colloid and Interface Science, 552, 326-333.

9. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. Proceedings of the 34th International Conference on Machine Learning, 70, 1263-1272.

10. Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. MIT Press.

11. Zhao, J., Zhang, M., Wang, C., Yu, W., Zhu, Y., & Zhu, P. (2025). First-principles study of CO, NH3, HCN, CNCl, and Cl2 gas adsorption behaviors of metal and cyclic C–metal B-and N-site-doped h-BNs. Electronic Materials Letters, 21(2), 268-288.

12. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686-707.

13. Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., ... & Ceder, G. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Materials, 1(1), 011002.

14. Lejaeghere, K., Bihlmayer, G., Björkman, T., Blaha, P., Blügel, S., Blum, V., ... & Cottenier, S. (2016). Reproducibility in density functional theory calculations of solids. Science, 351(6280), aad3000.

15. Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C., & Scheffler, M. (2015). Big data of materials science: Critical role of the descriptor. Physical Review Letters, 114(10), 105503.

16. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50-60.

17. Cheng, Y., Wang, D., Zhou, P., & Zhang, T. (2018). A survey of model compression and acceleration for deep neural networks. arXiv preprint arXiv:1710.09282.

18. Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. Neural Networks, 113, 54-71.

19. Schütt, K. T., Sauceda, H. E., Kindermans, P. J., Tkatchenko, A., & Müller, K. R. (2018). SchNet – A deep learning architecture for molecules and materials. The Journal of Chemical Physics, 148(24), 241722.

20. Zügner, D., Akbarnejad, A., & Günnemann, S. (2018). Adversarial attacks on neural networks for graph data. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2847-2856.

21. Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., ... & Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. ACS Central Science, 4(2), 268-276.

22. National Institute of Standards and Technology. (2021). Guidelines for the validation of computational models for materials science. NIST Special Publication 1200-16.

23. Choudhary, K., DeCost, B., & Tavazza, F. (2020). Transfer learning for property prediction of materials. npj Computational Materials, 6, 173.

24. Kim, J., & Lee, H. (2021). Machine learning prediction of thermal conductivity of carbon foams using limited experimental data. Carbon, 174, 456-464.

25. Park, H., & Sholl, D. S. (2020). Machine learning for high-throughput screening of metal-organic frameworks for gas capture. Journal of Materials Chemistry A, 8(42), 22361-22374.