Q-CatNet: Leveraging Quantum and Graph Features for Catalyst Simulation and Discovery

Ericsson T. Chenebuah *
University of Ottawa
National Research Council Canada

Michel Nganbe
University of Ottawa

David Liu
University of Ottawa

Alain Tchagang
University of Ottawa
National Research Council Canada

Abstract

The rational design of heterogeneous catalysts demands predictive tools that balance quantum-level accuracy with efficiency. We present Q-CatNet, a Quantum-informed Crystal Network derived from the Crystal Graph Convolutional Neural Network (CGCNN) model that integrates enriched edge descriptors with global state attributes to capture both local and system-wide physics. Ablation studies show that edge features alone yield only marginal gains over a plain CGCNN, while the joint edge–global representation in Q-CatNet achieves a Mean Absolute Error (MAE) score of 0.294 eV, improving accuracy by ~30%. Benchmarking against baselines confirms its robustness: Q-CatNet outperforms the image-based Fourier Transformed Crystal Property (FTCP) representation by 46% and surpasses three state-of-the-art graph-based architectures for material informatics, namely GINConv, NNConv, and SchNet, by margins ranging from 8% to 38%. These results highlight Q-CatNet as a generalizable and physically consistent framework for accelerating adsorption energy prediction, offering a practical route toward efficient catalyst screening and discovery.

1 Introduction

Historically, designing effective catalysts has been a difficult task, often complicated by the intricate interplay of several factors that influence catalytic activity, selectivity, and stability. Traditional methods, including high-throughput density functional theory (DFT) calculations [1] and Edisonian-based empirical screening [2], while effective, are often constrained by computational expense and the immense chemical and structural space. To overcome such a challenge and accelerate the discovery and design of novel catalysts, the field of materials science is increasingly turning to advanced approaches at the expense of traditional techniques. In recent years, moreover, deep learning techniques have become transformative data-driven tools in materials design and informatics, enabling unprecedented efficiency in predictive modeling and optimization. Classic Graph Neural Network (GNN) frameworks like Crystal Graph Convolutional Neural Network (CGCNN) [3], and MatErials Graph Network (MEGNet) [4], have been validated on both inorganic and organic materials databases such as the Materials Project (MP) [5], JARVIS-DFT [6], and Open Quantum Materials Database (OQMD) [7]. In the context of catalysis, GNNs have been successfully applied for rapid prediction of catalytic activity and selectivity, enabling efficient screening of novel materials [8-11]. These applications span diverse catalytic systems or descriptors, including homogeneous, heterogeneous,

*All correspondence should be directed to echen013@uottawa.ca

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

electro-, and photocatalysis, further highlighting their versatility and broad applicability in advancing catalytic science. Moreover, among the myriad of descriptors that serve as predictive proxies in modeling catalytic performance with GNNs, the adsorption energy (ΔE_{ads}) stands out as arguably the most critical and universally applicable parameter for explaining and predicting catalytic activity and microkinetics [12]. By definition, ΔE_{ads} is the strength of interaction between a reactant or intermediate species and the catalyst surface, and is a measure of the prominent energetic landscape for surface reactions to occur. As such, accurately predicting ΔE_{ads} through machine learning techniques, particularly in the context of the *Sabatier* principle, has become a significant focus of research and has been pursued by numerous investigators. In this regard, state-of-the-art models have actually been employed for baseline ΔE_{ads} prediction in prior works, laying a strong foundation for subsequent advancements. For instance, the original Open Catalyst 2020 (OC20) [13] dataset study utilized three established GNN architectures, namely CGCNN [3], SchNet [14], and DimeNet++ [15] to predict relaxed ΔE_{ads} values from initial crystal structures. From the study, they achieved best in-domain MAE scores on test set at 0.6149 eV, 0.6387 eV, and 0.5620 eV, respectively [13]. Despite these significant advancements, a substantial gap remains between current ML performance and the level of experimental chemical accuracy typically desired for thermochemical properties, which is at approximately 0.0434 eV/atom, or 1 kcal/mol [16]. This discrepancy likely stems from the intrinsic complexities of adsorbate–bulk surface interactions, including quantum mechanical effects such as charge transfer, binding site specificity, and orbital hybridization, which remain challenging to capture with current ML descriptors and architectures. These limitations highlight the ongoing need for innovative approaches in descriptor design and model development, aimed at bridging this accuracy gap and advancing the predictive capabilities of ML in catalysis.

To make a contribution in this regard, the present study proposes a Quantum-informed Catalysis Network (Q-CatNet) model for improving prediction capabilities of ΔE_{ads} as a primary descriptor of catalytic activity, alongside other target properties relevant to simulating adsorbate–bulk surface interactions. The Q-CatNet model integrates improved Coulombic interaction modeling and periodic crystalline features to better represent adsorbate–bulk surface dynamics. This study demonstrates the model's applicability using four distinct adsorbate species (C, H, N, and O) sourced from the OC20 dataset [13].

2 Methodology

Given a catalyst system represented in its conventional unit cell, the Q-CatNet model encodes it as an input graph G defined by node, edge, and global-state attributes. Using supervised learning, the model learns a mapping function $f(G) \mapsto \Delta E_{ads}$ which predicts the target adsorbate energy. The Q-CatNet descriptor integrates enriched quantum-mechanical input features in its graph-based representation at node (v), edge (u), and global state ($\mathcal{G}$) levels. The node features comprise of fourteen carefully selected physicochemical properties (see **suplementary**) that are transformed into one-hot encoding forms of size $v_i \in \{0,1\}^{(1\times 230)}$, thereby capturing essential thermo- and electronic characteristics relevant to catalytic systems [17]. Meanwhile, the edge features are expanded beyond the conventional bond-length attribute used in the traditional CGCNN [3] to include bond angles (θ), differences in electronegativity ($\Delta\chi$), and three quantum-informed descriptors that capture the underlying physics of atomic environments. These descriptors are derived by approximating mathematical components of the computationally intensive Schrödinger equation, as represented in their encoded Coulomb matrix [18], Ewald sum matrix [19], and Sine matrix [20-21] forms. Furthermore, this study incorporates global state attributes to adequately simulate the electronic and periodic dynamics that are associated with the catalyst system. This integration ensures that message-passing operations within the graph network are continuously informed by system-wide context, facilitating the propagation of relevant information across all nodes and edges. Specifically, two global attributes are encoded as one-dimensional vectors with a combined dimension of $\mathcal{G} \in \mathbb{R}^{(1\times 1204)}$, including density of states (DOS) signal $g(E) \in \mathbb{R}^{(1\times 1024)}$ of the bulk system, sampled within $\pm 10\, eV$ of the Fermi energy, and the X-ray diffraction (XRD) intensity profile $I_{XRD} \in \mathbb{R}^{(1\times 180)}$ of the adsorbate-bulk system, evaluated over a Bragg angle range up to $2\theta = 180^{\circ}$. In general, the DOS signal provides vital insights into the catalyst's electronic structure, especially in the context of *d-band* theory, which plays a key role in determining how the catalyst interacts with the adsorbate [12]. On the other hand, X-ray Diffraction (XRD) reveals information about the combined adsorbate-bulk crystal structure, along with other critical details such as phase composition and defect chemistry [22]. Figure 1 presents the Q-CatNet

framework alongside three representative adsorption sites where a hydrogen atom may bind to the surface of the $TiPt$ binary alloy from the Materials Project.

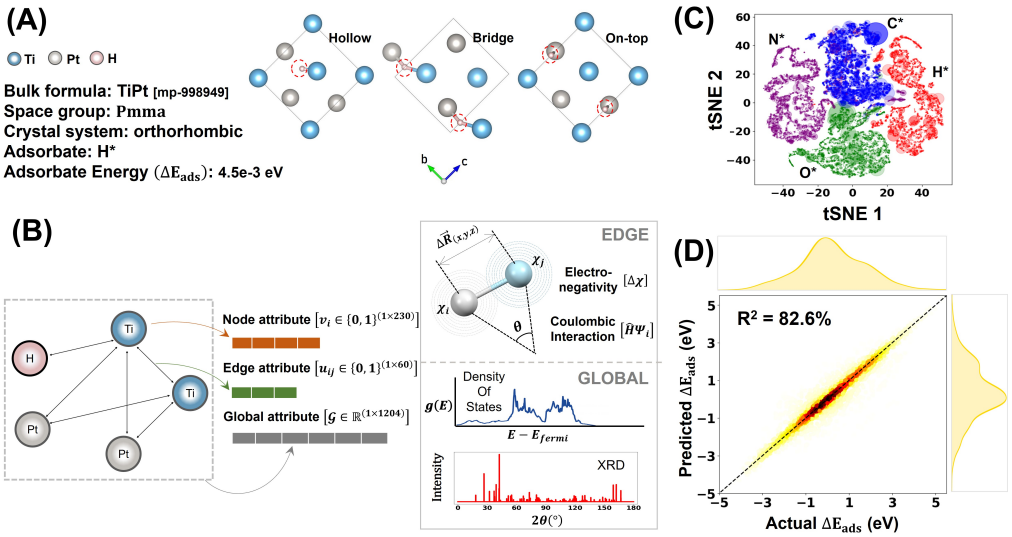

Figure 1: (A) A $TiPt$ bulk crystal, obtained from the Materials Project (MP) [5] database. To simulate enhanced binding stability in adsorbate–surface interactions, three adsorption sites are modeled: hollow, bridge, and on-top, prioritized in this order; (B) Q-CatNet graphical framework illustrating quantum-informed node, edge, and global attributes; (C) t-SNE projection of the adsorbate–surface dataset, showing distinct clustering by adsorbate species; (D) Regressive parity plot from the prediction of ΔE_{ads} with 82.6% in $overall$ R^2 accuracy.

The mapping model is constructed by integrating deep neural networks, specifically by combining CGCNN layers with fully connected linear layers. Fundamentally, the CGCNN layer performs nonlinear convolution and pooling operations to iteratively update atomic (node) feature vectors through message-passing interactions with neighboring atoms, as specified by the adjacency matrix of the adsorbate-bulk system. This representation is supplemented with additional quantum context by concatenating with the global-state attributes. Fully connected linear layers are then employed to model the complex relationships between the combined crystal features and their corresponding target properties. The training is performed by minimizing the difference between the predicted property $\Delta\hat{E}_{ads}$ and the DFT calculated property ΔE_{ads}, defined by a mean squared error cost function.

3 Results and discussion

The analysis shows that the model achieved an *overall* MAE of 0.294 eV across all adsorbate–surface combinations, indicating strong generalization capability in predicting ΔE_{ads} for unknown catalysts. Among the individual adsorbates, Hydrogen yielded the lowest MAE of 0.186 eV, highlighting its relatively high accuracy and model reliability in this subset. In contrast, Oxygen exhibited the highest coefficient of R² at 84.46%, despite a slightly higher MAE of 0.318 eV compared to Hydrogen's 75.07% R². This discrepancy arises from the narrower variance observed in the training distribution of Hydrogen-surface systems, which can limit the model's ability to explain variability in predictions. Conversely, the broader spread in data for Oxygen, Nitrogen, and Carbon systems enables the model to better capture diverse adsorbate behaviors, as reflected by their consistently high R² values, all exceeding 80% (additional context on data distribution in **supplementary**). Moreover, the predictive results are benchmarked against four baseline models spanning diverse graph-based and image-based deep learning approaches that are reliably applied in molecular and materials informatics. Specifically, Table 1 summarizes the MAE, RMSE, and R² scores for the six evaluated models considered in the study. The ablation begins with the CGCNN in its baseline form, incorporating only simple edge features derived from bond length. This is compared to an edge-enhanced CGCNN*, which integrates all considered edge attributes, including bond length, bond angle, electronegativity difference, Coulomb matrix, Ewald-sum matrix, and Sine matrix, while excluding global state attributes. From the investigation, it can be observed that incorporating enhanced edge features

yields a modest improvement in overall accuracy, reducing MAE from 0.422 eV (plain CGCNN) to 0.409 eV (edge-enhanced CGCNN*), an improvement of approximately 3%. These results suggest that edge-enhanced attributes alone may only offer limited gains unless combined with global state attributes. In fact, with the integration of edge and global features, as demonstrated in Q-CatNet, the predictive accuracy improves substantially, with MAE reductions of 30.3% and 28.1% compared to the plain CGCNN and edge-enhanced CGCNN*, respectively. This underscores the synergistic role of edge and global features in capturing complex adsorption-surface interactions. Beyond the ablation studies, Q-CatNet's performance is benchmarked against one image-based and three graph-based deep learning architectures. The chosen image-based baseline is the Fourier Transformed Crystal Property (FTCP) descriptor [23], which is an invertible representation that encodes atomic positions and types into the reciprocal (Fourier) space of a crystal lattice. The applied FTCP model yields an MAE of 0.544 eV, indicating that the Q-CatNet achieves a relative improvement of 46%. Besides, even the plain CGCNN surpasses FTCP by 22.4%, reinforcing the advantage of graph-based over image-based crystal representations. The current study attributes this performance gap to the inherent invariance of graph neural networks, with respect to rotation, translation, and permutation, properties that ensure physical consistency. In contrast, image-based approaches require discretizing atomic structures into voxel or grid formats, which can degrade precision and significantly increase input dimensionality. Finally, Q-CatNet is benchmarked against three reputable graph-based deep learning architectures: Graph Isomorphism Network Convolution (GINConv) [24], Neural Network Convolution (NNConv) [25], and the Continuous-filter Convolutional Neural Network for Molecules and Materials (SchNet) [14]. For a fair comparison, GINConv and NNConv were augmented with the same global attributes (i.e. $\mathcal{G} \in \mathbb{R}^{1\times 1204}$) used in Q-CatNet, and trained on the same dataset split for evaluating the Q-CatNet. The results show that Q-CatNet outperforms GINConv by 8.7%, NNConv by 26.3%, and SchNet by 38.2% in MAE accuracy, underlining its excellent representational capacity.

Table 1: Comparative performance evaluation of Q-CatNet, its ablated CGCNN variants, and baseline models for adsorbate energy prediction. Results are reported as MAE (eV), RMSE (eV), and R^2 (%).

Adsorbate species	Metric	Plain CGCNN	CGCNN*	FTCP	GINConv	NNConv	SchNet	Q-CatNet
Carbon	MAE	0.483	0.485	0.585	0.350	0.460	0.568	**0.350**
	RMSE	0.925	0.95	0.951	0.715	0.946	1.035	**0.791**
	R^2	73.05	71.57	69.41	83.53	71.14	66.41	**80.30**
Hydrogen	MAE	0.290	0.258	0.406	0.244	0.276	0.322	**0.186**
	RMSE	0.604	0.532	0.700	0.501	0.553	0.631	**0.425**
	R^2	49.60	60.91	34.15	60.71	52.27	33.68	**75.07**
Nitrogen	MAE	0.46	0.463	0.623	0.361	0.446	0.483	**0.328**
	RMSE	0.867	0.915	1.038	0.811	0.876	0.87	**0.746**
	R^2	77.31	74.71	69.03	79.76	76.41	77.27	**83.18**
Oxygen	MAE	0.463	0.438	0.57	0.338	0.42	0.542	**0.318**
	RMSE	0.882	0.835	0.923	0.699	0.821	0.994	**0.706**
	R^2	75.71	78.27	70.18	83.42	77.15	65.32	**84.46**
Overall	MAE	0.422	0.409	0.544	0.322	0.399	0.476	**0.294**
	RMSE	0.826	0.821	0.909	0.689	0.81	0.893	**0.679**
	R^2	74.18	74.51	67.76	81.14	73.93	68.69	**82.56**

4 Conclusion

In this work, we introduced Q-CatNet, a Quantum-informed Catalysis Network model designed to accelerate heterogeneous catalyst simulation and discovery by enhancing adsorbate energy prediction. Comparative benchmarking against baseline models further validated the robustness of Q-CatNet. Notably, Q-CatNet outperformed the image-based FTCP representation by 46%, and surpassed established graph-based models with improvements spanning 8% to 38% in terms of prediction accuracy. These results reinforce the strength of graph-based architectures over image-based ones for materials informatics tasks, especially due to their inherent invariance to rotation, translation, and permutation as well as their ability to efficiently capture periodic boundary conditions. Overall, Q-CatNet establishes itself as a powerful forward-design tool, offering improved accuracy, physical consistency, and scalability for catalysis research. By providing a reliable framework for predicting adsorbate energies, it paves the way for more efficient screening, simulation, and rational discovery of heterogeneous catalysts.

Data and code availability

The dataset used in the study can be downloaded from the Materials Project and Open Catalyst (2020) data platforms. Link to source codes can be found at https://github.com/chenebuah/Q-CatNet

Acknowledgements

This research was supported by the National Research Council of Canada (NRC) through its Artificial Intelligence for Design Program led by the Digital Technologies Research Centre.

References

[1] Jain, A., Hautier, G., Moore, C.J., et al. A high-throughput infrastructure for density functional theory calculations. *Comput. Mater. Sci.*, **50**, 2011, 2295–310. https://doi.org/10.1016/j.commatsci.2011.02.023

[2] Amis, E.J., Xiang, X-D., & Zhao, J-C. Combinatorial Materials Science: What's New Since Edison? *MRS Bulletin*, **27**, 4, 2002, 295-300. https://doi.org/10.1557/mrs2002.96

[3] Xie, T. & Grossman, J.C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.*, 145301, 2018. https://doi.org/10.1103/PhysRevLett.120.14530

[4] Chen, C., Ye, W., Zuo, Y, et al. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chem. Mater.*, **31**, 9, 2019, 3564–3572. https://doi.org/10.1021/acs.chemmater.9b01294

[5] Jain, A. Ong, S.P., Hautier, G. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *Apl. Mater.*, **1**, 011002, 2013. https://doi.org/10.1063/1.4812323

[6] Choudhary, K., Garitty, K.F., Reid, A.C.E, et al. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Comput. Mater.*, **6**, 173, 2020. https://doi.org/10.1038/s41524-020-00440-1

[7] Kirklin, S., Saal, J., Meredig, B., et al. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.*, **1**, 15010, 2015. https://doi.org/10.1038/npjcompumats.2015.10

[8] Price, C.C., Singh, A., Frey, N.C., et al. Efficient catalyst screening using graph neural networks to predict strain effects on adsorption energy. *Sci. Adv.*, **8**, eabq5944, 2022. https://doi.org/10.1126/sciadv.abq5944

[9] Jiao, Z., Liu, Y., & Wang, Z. Application of graph neural network in computational heterogeneous catalysis. *J. Chem. Phys.*, **161**, 171001, 2024. https://doi.org/10.1063/5.0227821

[10] Chen, Z., Li, H., Xu, H., et al. The structure–property relationship of metallocene-based ethylene oligomerization catalysts using DFT and graph neural networks. *Dalton Trans.*, **54**, 2025, 4069-4081. https://doi.org/10.1039/D4DT03065F

[11] Aguilar Bejarano, E.A., Figueredo, G., Woodward, et al. HCat-GNet: An Interpretable Graph Neural Network for Catalysis Optimization. *ChemRxiv*, 2024. https://doi.org/10.26434/chemrxiv-2024-zjnkd

[12] Wang, Y., Qiu, W., Song, E., et al. Adsorption-energy-based activity descriptors for electrocatalysts in energy storage applications, *Natl. Sci. Rev.*, **5**, 3, 2018, 327–341. https://doi.org/10.1093/nsr/nwx119

[13] Chanussot, L., Das, A., Goyal, S., et al. Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catal.*, **11**, 10, 2021, 6059–6072. https://doi.org/10.1021/acscatal.0c04525

[14] Schütt, K.T., Kindermans, P-J, Sauceda, H.E., et al. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *Adv. Neural Inf. Process. Syst.*, **30**, 2017, 992-1002. https://arxiv.org/abs/1706.08566.

[15] Gasteiger, J., Giri, S., Margraf, J.T., et al. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. *Machine Learning for Molecules Workshop, NeurIPS*, 2020. https://doi.org/10.48550/arXiv.2011.14115

[16] Bogojeski, M., Vogt-Maranto, L., Tuckerman, M.E., et al. Quantum chemical accuracy from density functional approximations via machine learning, *Nat Commun.*, **11**, 2020, 5223. https://doi.org/10.1038/s41467-020-19093-1

[17] Chenebuah, E.T., Nganbe, M. & Tchagang, A. An evolutionary variational autoencoder for perovskite discovery. *Front. Mater.*, **10**, 1233961, 2023. https://doi.org/10.3389/fmats.2023.1233961

[18] Rupp, M., Tkatchenko, A., Müller, K-R, et al. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.*, **108**, 058301, 2012. https://doi.org/10.1103/PhysRevLett.108.058301

[19] Toukmaji, A.Y., & Board, J.A. Jr., Ewald summation techniques in perspective: a survey. *Comput. Phys. Commun.*, **95**, 1996, 73–92.

[20] Faber, F., Lindmaa, A., von Lilienfeld, O.A., et al. Crystal structure representations for machine learning models of formation energies *IJQC* **115**, 2015, 1094–101. https://doi.org/10.1002/qua.24917

[21] Himanen, L., Jäger, M.O.J, Morooka E.V, et al., DScribe: library of descriptors for machine learning in materials science. *Comput. Phys. Commun.*, **247**, 106949, 2020. https://doi.org/10.1016/j.cpc.2019.106949

[22] Chenebuah, E.T., Nganbe, M. & Tchagang, A.B. A deep generative modeling architecture for designing lattice-constrained perovskite materials. *npj Comput. Mater.* **10**, 198, 2024. https://doi.org/10.1038/s41524-024-01381-9

[23] Ren. Z., Tian, S.I.P., Noh, J., et al. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter*, **5**, 1, 2022, 314–35. https://doi.org/10.1016/j.matt.2021.11.032

[24] Xu, K., Hu, W., Leskovec, J., et al. How Powerful are Graph Neural Networks?. *Computer Vision and Pattern Recognition*, 2019. https://doi.org/10.48550/arXiv.1810.00826

[25] Gilmer, J., Schoenholz, S.S., Riley, P.F., et al. Neural Message Passing for Quantum Chemistry. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017. https://doi.org/10.48550/arXiv.1704.01212

[26] Larsen, A.H., Mortensen, J.J., Blomqvist, J., et al. The Atomic Simulation Environment—a Python library for working with atoms. *J. Phys. Condens. Matter.*, **29**, 27, 273002, 2017. https://doi.org/10.1088/1361-648X/aa680e

[27] Ong, S.P., Richards, W.D., Jain, A., et al. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.*, **68**, 2013, 314–319. https://doi.org/10.1016/j.commatsci.2012.10.028

[28] De Graef, M., & McHenry, M. E. Structure of Materials: An Introduction to Crystallography, Diffraction and Symmetry. *Cambridge University Press*, 2nd edn, **268**, 2012.

A Technical Appendices and Supplementary Material

A.1 Data Distribution and Preprocessing

The Q-CatNet model was trained on a curated dataset derived from the Open Catalyst 2020 (OC20) [3] and Materials Project (MP) [5] repositories. The OC20 dataset primarily describes adsorbate interactions with bulk surfaces of stable materials, many of which are also represented in the MP database. For this study, we restricted the data to adsorbate systems containing carbon (C), hydrogen (H), nitrogen (N), and oxygen (O), chosen for their central roles in energy and environmental applications. After curation and cleaning, the final dataset comprised 37,433 unique samples with a nearly uniform distribution across adsorbate species: 24.63% C, 26.04% H, 24.65% N, and 24.68% O. As such, Figure 2 presents combined box-and-violin plots depicting the statistical distributions of adsorption energies for each species. H–surface systems show a narrower distribution clustered near the optimal adsorption energy, reflecting more stable and consistent interactions. In contrast, N– and O–surface systems exhibit broader distributions, with nitrogen skewed toward higher ΔE_{ads} values and oxygen extending toward more negative extremes.

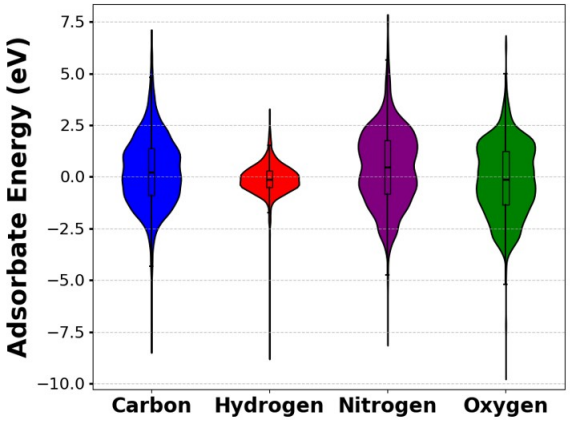

Figure 2: Combined box-and-violin plots illustrating the distribution of binding energies for C, H, N, and O adsorbates across the training dataset.

A.2 Simulation Framework and Imported Libraries

The quantum-informed edge descriptors, namely: Coulomb matrix, Ewald sum matrix, and Sine matrix, were computed using their respective modules from the `Dscribe` [21] package. In this work, these edge features were calculated only for neighboring atoms located within a spherical cut-off radius defined by covalent radii and identified through the adjacency matrix. To achieve this, the `NeighborList` module from the Atomic Simulation Environment (ASE) [26] was employed with non-self interaction enabled. To prioritize preferred adsorbate binding configurations (hollow, bridge, or on-top sites), the `AdsorbateSiteFinder` module from the `Pymatgen` [27] adsorption library was utilized, and applied to the conventional unit cell of each bulk substrate. The conventional cell, determined via the `SpacegroupAnalyzer` module, was chosen over the primitive cell to ensure better alignment with experimental conventions and to facilitate the construction of periodic slab models for density functional theory (DFT) calculations, which are critical for simulating extended catalytic surfaces. To compute and integrate global-state features into the model, XRD powder diffraction patterns of the bulk conventional crystal structures were generated using the `XRDCalculator` [28] module from `Pymatgen`, evaluated across a Bragg's angle range of 0° to 180°. For electronic structure features, Density of States (DOS) signals were sampled from the Materials Project database using `MPRester.get_bandstructure` and `MPRester.get_dos`. Data were extracted for energy levels within ±10 eV of the Fermi energy, followed by smooth one-dimensional cubic interpolation using `SciPy` to obtain a $\mathbb{R}^{(1\times1024)}$ vector length. All contributing edge and global attributes were normalized to a range between 0 and 1. We further develop an ablated version of Q-CatNet for predicting Density of States (DOS) signals, serving as a surrogate model for invertible predictions of novel compounds. Details of this approach are provided in the following subsections. Figure 3

summarizes the feature attributes used in study, as well as the graphical framework used to design the Q-CatNet learning model

(A)

ATTRIBUT - ES	FEATURES	VECTOR LENGTH/BINS
Node $[v_i]$	Atomic number	100
	Group number	18
	Row number	9
	Block	4
	Pauling electronegativity	10
	Covalent radius	10
	Valence	9
	Ionization energy (log scale)	10
	Electron affinity	10
	Molar volume (log scale)	10
	Average ionic radius	10
	Polarizability	10
	Specific heat (log scale)	10
	Thermal conductivity (log scale)	10

ATTRIBUT - ES	FEATURES	VECTOR LENGTH/BINS
Edge $[u_{ij}]$	Bond length	10
	Bond angle	10
	Electronegativity difference	10
	Coulomb interaction	10
	Ewald-sum interaction	10
	Electrostatic sine interaction	10
Global $[\mathcal{G}]$	Density of states	1024
	XRD pattern	180

(B)

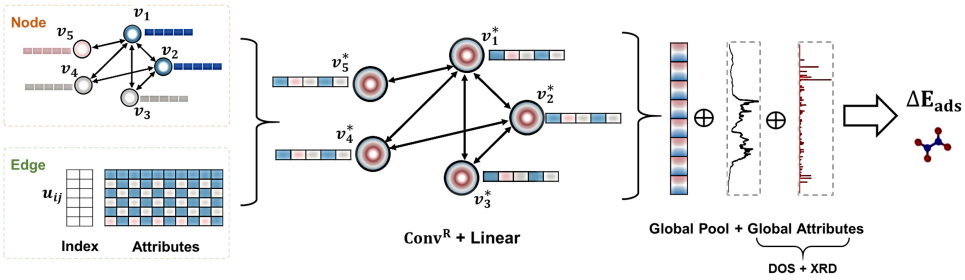

Figure 3: (A) Complete list of all contributing feature attributes, categorized as node, edge, and global features, as applied in the present study, along with their corresponding discretized vector lengths; (B) Schematic representation of the Q-CatNet learning framework, which maps a crystal graph composed of nodes, edges, and global descriptors, to the adsorption energy (ΔE_{ads}) of a catalyst system. Message-passing is performed among atoms within the conventional unit cell using R convolutional layers, followed by fully connected linear layers.

A.3 Generalization Performance and Data-Efficiency of Q-CatNet

A key future application of Q-CatNet lies in its ability to estimate ΔE_{ads} in previously unexplored regimes, leveraging limited input from first-principles calculations and/or experimental data on heterogeneous catalysis systems. To evaluate this capability, we conducted a predictive assessment using a 90/10 train–test split, where each adsorbate contributed 10% of its total dataset to ensure a balanced and uniform distribution across the test set. Parity plots in Figure 4 illustrate regression performance on test set. The model achieved R^2 values of 80.3%, 75.07%, 83.18%, and 84.46% for carbon, hydrogen, nitrogen, and oxygen systems, respectively. These results confirm predictive accuracy, with most data points closely aligned along the diagonal, indicating strong agreement between predicted and true values. To further examine generalization, we evaluated the effect of progressively reducing the training sample size, while preserving proportional representation of adsorbates in the test batches. Results are summarized in Figure 5, with predictive performance reported as MAE (Fig. 5(A)) and R^2 (Fig. 5(B)). On average, MAE improved by 5.5 meV per 1% increase in training size, reflecting strong stability. Among individual adsorbates, hydrogen achieved the best improvement, at 3.4 meV per 1% increase in training data size. In terms of R^2, performance improved at an average rate of 0.446% per 1% increase in training data size, with nitrogen and oxygen demonstrating the greatest resilience (drops of only 0.412% and 0.425%, respectively). These findings highlight that Q-CatNet captures fundamental adsorption–surface interaction patterns rather than overfitting to the training data, underscoring its robustness and suitability for extrapolation to unseen catalyst systems.

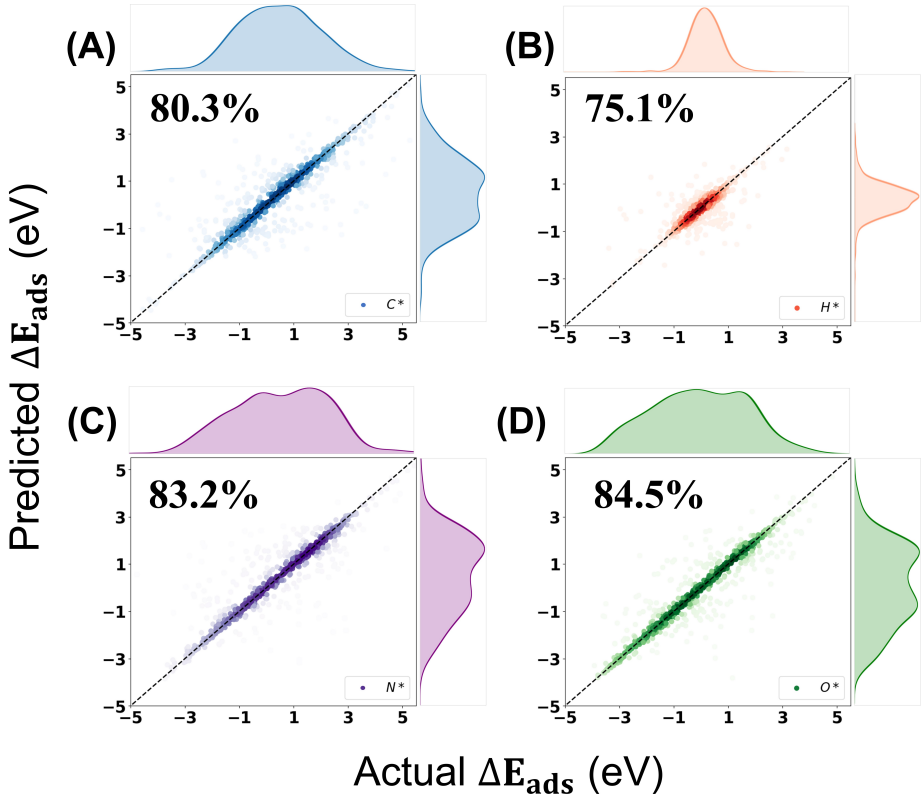

Figure 4: Parity plots from test set evaluation showing the comparison between Q-CatNet-predicted adsorption energies and DFT-calculated values for four adsorbate–surface catalyst systems. The model achieves R^2 values of 80.3%, 75.07%, 83.18%, and 84.46% for Carbon, Hydrogen, Nitrogen, and Oxygen systems, respectively, indicating strong predictive accuracy.

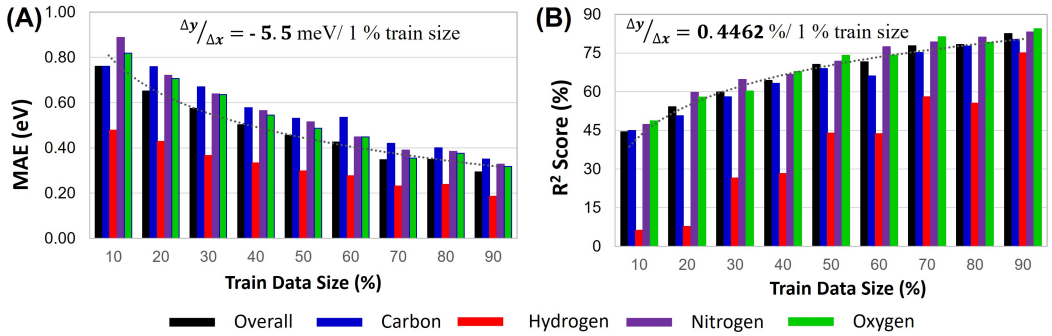

Figure 5: Effect of increasing training data proportion on adsorbate energy prediction performance using Q-CatNet. The results reveal a gradual decline in predictive accuracy, highlighting the model's robustness and its ability to generalize underlying patterns rather than memorize the training data. (A) Mean Absolute Error (MAE) decreases with an overall slope of 5.5 meV per 1% increase in train size. (B) Coefficient of determination (R^2) increases with an overall slope of 0.446% per 1% increase in train size.

A.4 Extending Q-CatNet to General Materials Property Prediction

The applicability of Q-CatNet is extended beyond catalytic systems to encompass other essential DFT-derived properties that quantify material stability and functionality. Accordingly, Table 2 presents test set results demonstrating the predictive capabilities of Q-CatNet for a range of properties, including formation energy (E_f), band gap (E_g), energy above the convex hull (E_{hull}), total DFT-computed

energy (E_{tot}), and one-dimensional interpolated Density of States (DOS) signals. These predictions were performed on approximately 52,000 newly extracted compounds from the Materials Project database. The results were generated using an ablated version of Q-CatNet, which incorporates all edge-computed features (as described in the main article) and exclusive XRD-derived global-state attributes. All evaluations were consistently conducted using a 90/10 train-test split. Figure 6 further presents the corresponding parity plots, demonstrating that Q-CatNet achieves high predictive accuracy across all considered properties. Notably, for band gap prediction, Q-CatNet attains a MAE score of 0.097 eV, a substantial improvement over the original CGCNN benchmark, which reported an MAE of 0.388 eV[3] when trained on ~16,000 Materials Project structures. These results collectively highlight Q-CatNet's versatility in learning diverse material property landscapes, underscoring its potential as a unified framework for accelerating large-scale materials discovery.

Table 2: Predictive performance of an ablated version of the Q-CatNet across various target properties. Scalar outputs from DFT are evaluated using regression metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and coefficient of determination (R^2). Vector-valued DFT properties, are assessed using similarity metrics including Jaccard index, Cosine similarity, and Pearson correlation coefficient.

Scalar Target Property	**MAE**	**RMSE**	**R² (%)**
Formation energy, E_f (eV/atom)	0.038	0.131	97.19
Energy above the convex hull, E_{hull} (eV/atom)	0.024	0.104	61.30
Energy band gap, E_g (eV)	0.097	0.390	85.25
DFT total energy, E_{tot} (eV)	0.559	2.949	94.31
Vector target property	**Jaccard**	**Cosine**	**Pearson**
Normalized Density of State (DOS) signal	0.761	0.922	0.825

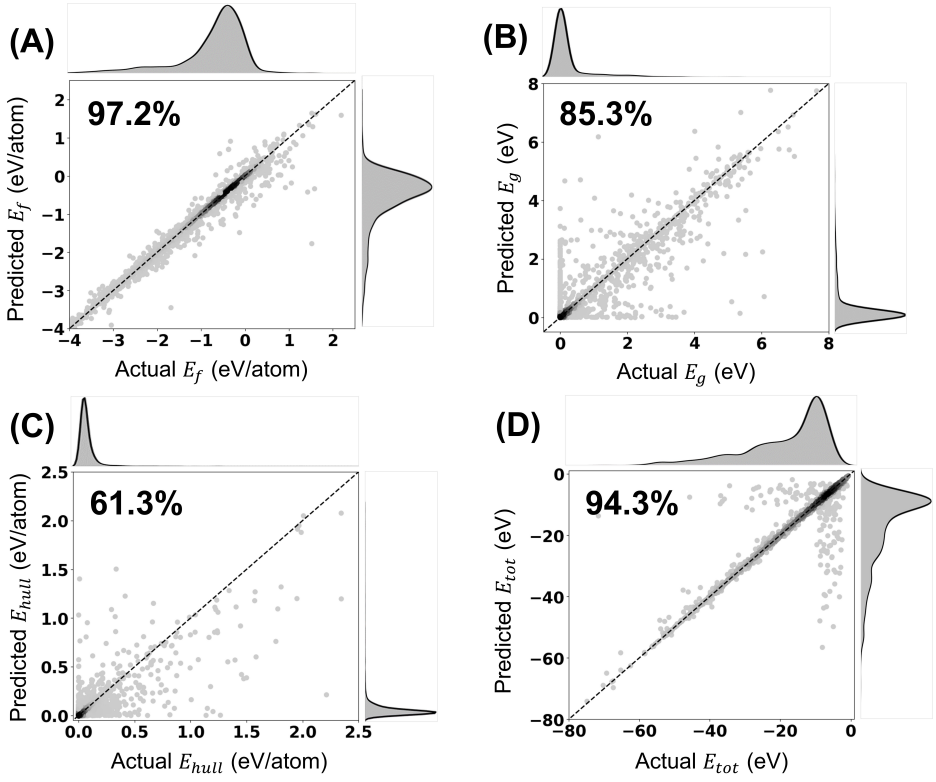

Figure 6: Parity plots from test set evaluation comparing Q-CatNet-predicted scalar properties: (A) formation energy (E_f), (B) band gap (E_g), (C) energy above the convex hull (E_{hull}), and (D) total energy (E_{tot}) against DFT reference values the Materials Project.