

<https://doi.org/10.1038/s41524-025-01917-7>

# Adaptive edge-aware graph convolutional with multi-task learning for simultaneous prediction of material properties

Check for updates

Yunhua Lu<sup>1,2</sup>, Mingyue Chen<sup>1</sup>, Qingwei Zhang<sup>1</sup> ✉, Junan Zhang<sup>1</sup>, Chao Zhang<sup>3</sup>, Shiai Xu<sup>3,4</sup> & Qiuyan Bi<sup>3</sup>

The targeted design of functional materials often requires the concurrent optimization of multiple interdependent properties. For boron-doped graphene (BDG), both the band gap and work function critically influence performance in electronic and catalytic applications, yet existing machine learning (ML) approaches typically focus on single-property prediction and rely on hand-crafted features, limiting their generality. Here we present an adaptive edge-aware graph convolutional neural network with multi-task learning (AEGCNN-MTL) for simultaneous prediction of multiple material properties. On a DFT-computed BDG dataset of 2613 structures, AEGCNN-MTL achieved high accuracy ( $R^2 = 0.9905$  for band gap and  $0.9778$  for work function), and under identical training budgets, outperformed representative single-task GNN baselines. When transferred to the QM9 benchmark, the framework delivered competitive performance across 12 diverse quantum chemical properties, demonstrating strong generalization capability. These results highlight the potential of AEGCNN-MTL as a scalable and accurate tool for high-throughput, multi-property screening and the data-driven discovery of multifunctional materials.

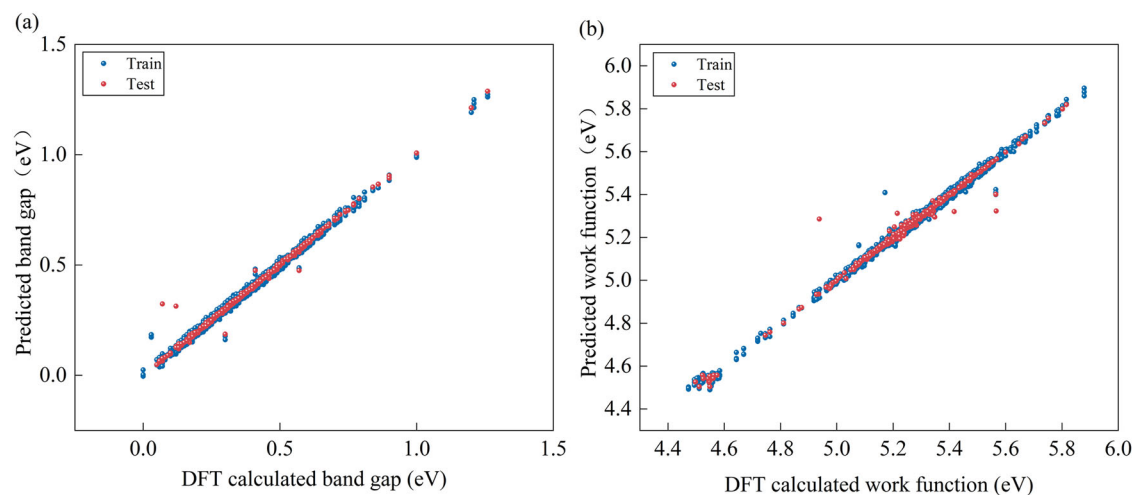
Graphene, a prototypical two-dimensional (2D) material, has attracted significant scientific interest since its isolation in 2004, owing to its exceptional electronic, mechanical, and thermal properties<sup>1–6</sup>. However, its intrinsic zero band gap and chemically inert surface severely limit its direct application in semiconducting and catalytic systems<sup>7,8</sup>. To address these limitations, a wide range of modification strategies such as heteroatom doping, vacancy engineering, and strain modulation have been extensively explored to modulate the intrinsic electronic and chemical properties of graphene<sup>9–11</sup>. Among these strategies, boron doping has emerged as an effective approach for modifying the electronic structure and physicochemical characteristics of graphene, resulting in BDG with enhanced catalytic activity and electronic tunability<sup>2,12–14</sup>. Notably, such structural modifications often lead to concurrent changes in multiple material properties. In particular, the band gap and work function are two critical electronic properties that govern the functionality of devices in fields such as nanoelectronics, photovoltaics, and catalysis<sup>15,16</sup>. For instance, Ganose et al. demonstrated that doping SnO<sub>2</sub> could simultaneously modulate its band gap and work function, thereby enhancing its transparent conducting performance in photovoltaic systems<sup>17</sup>. Both properties are strongly correlated with the material's electronic structure and bonding configuration, and can be simultaneously modulated by atomic-level structural perturbations.

Thus, a comprehensive understanding of how multiple material properties co-evolve under atomic-scale perturbations is essential for the targeted design of graphene-based functional materials.

DFT has been extensively employed to calculate material properties with high accuracy, such as band gap, formation energies, adsorption energies, and mechanical properties<sup>18–20</sup>. However, its high computational cost, particularly in large-scale or high-throughput screening scenarios, poses significant challenges to its scalability. ML has emerged as a powerful complementary paradigm that enhances the efficiency of materials discovery pipelines. ML models trained on data obtained from DFT simulations or experimental measurements have shown the ability to predict material properties at a fraction of the computational cost, with comparable accuracy<sup>7,21–23</sup>. However, conventional ML models often rely heavily on manually engineered descriptors, which are designed based on domain-specific heuristics and may fail to capture the full complexity of atomic interactions and structural diversity.

Graph neural networks (GNNs) have recently gained significant attention in materials science due to their ability to represent materials as graphs, in which atoms are treated as nodes and chemical bonds as edges<sup>24–26</sup>. This graph-based representation enables GNNs to effectively capture both structural and chemical information, facilitating the end-to-

<sup>1</sup>School of Artificial Intelligence, Chongqing University of Technology, Chongqing, China. <sup>2</sup>Salt Lake Chemical Engineering Research Complex, Qinghai University, Xining, China. <sup>3</sup>School of Chemical Engineering, Qinghai University, Xining, China. <sup>4</sup>School of Materials Science and Engineering, East China University of Science and Technology, Shanghai, China. ✉e-mail: [zhangqingwei@cqut.edu.cn](mailto:zhangqingwei@cqut.edu.cn)



**Fig. 1 | Regression scatter plots of band gap and work function predictions by the AEGCNN-MTL model on the BDG dataset.** **a** Predicted versus DFT-calculated band gap. Blue circles denote training samples; red circles denote test samples.

**b** Predicted versus DFT-calculated work function. Blue circles denote training samples; red circles denote test samples.

end modeling of complex relationships between atomic configurations and material properties<sup>27</sup>. A pioneering example is the Crystal Graph Convolutional Neural Network (CGCNN) proposed by Xie et al.<sup>28</sup>, which employed local neighbor-based message passing to accurately predict crystal properties such as formation energies, band gap, and elastic moduli, while maintaining model interpretability. However, CGCNN is limited in its ability to account for long-range interactions or global state information, which are essential for accurately modeling collective electronic correlation effects. MEGNet introduced global state vectors to encode external conditions such as temperature and pressure, thereby enhancing the model's transferability across diverse material classes<sup>29</sup>. Nevertheless, MEGNet relies on static and relatively simplistic edge features, which may be insufficient to model subtle interatomic interactions, thereby limiting its expressiveness. Due to the planar geometry of 2D materials, variations in bond angles and torsional deformations are inherently restricted, rendering the angular-dependent encodings in models like ALIGNN<sup>30</sup>, DimeNet++<sup>31</sup>, and GemNet<sup>32</sup> marginally informative. In addition, such models often incur substantial computational overhead due to the use of complex directional filters or hierarchical graph constructs. This overhead hinders scalability in high-throughput screening scenarios while yielding only limited performance improvements for low-dimensional systems with geometrically constrained topologies.

Despite the growing adoption of MTL in computational materials science, existing frameworks remain hindered by intrinsic challenges in scalability, cross-domain generalization, and structural representation. For instance, models such as MT-MD DimeNet++<sup>33</sup> and DeepFM<sup>34</sup>, although effective within specific property domains, rely on task-specific architectural assumptions or manually engineered features, while H-CLMP<sup>35</sup> employs a composition-driven contrastive learning strategy without incorporating explicit structural information. Although these methods have demonstrated success in certain applications, they are inherently constrained by limited structural awareness and lack the capacity to jointly model multiple interdependent properties across materials with diverse atomic topologies. This reveals an unmet need for a generalizable and structure-aware framework that can effectively learn from the coupled evolution of multiple properties across complex atomic environments. Although all tasks share the same molecular-graph input, multi-task learning remains challenging: band gap is governed by local bond topology, whereas work function reflects long-range surface charge, so a shared representation drifts to a compromise that hurts both. Without task-specific decoders or balanced gradient terms, the network also suffers from conflicting gradients and unstable convergence.

Misra et al. designed the Cross-stitch networks<sup>36</sup>, they mitigate this by allocating separate backbones each task, but parameter count grows linearly with the number of properties and becomes prohibitive for data sets such as QM9.

These highlight the necessity for a unified and robust GNN framework that can achieve accurate and scalable multi-property prediction across materials with diverse structural and compositional complexity. In this work, we propose AEGCNN-MTL, a novel architecture that effectively encodes edge-level chemical features and captures long-range atomic interactions, and supports the simultaneous prediction of multiple inter-related material properties. The model is evaluated on both a newly constructed BDG dataset and the widely used QM9 benchmark, demonstrating superior accuracy, generalization ability, and scalability in multi-property prediction tasks.

The main contributions of this study are summarized as follows:

1. We introduce a novel framework combining an Adaptive Edge-aware Graph Convolution and a Multi-head Attention Layer, which collaboratively models local chemical interactions and long-range atomic dependencies, thereby enhancing the representation of complex molecular and crystalline structures.
2. We design a multi-task collaborative learning structure that enhances training stability and generalization by combining shared latent representations, task-specific decoder architectures, and synchronization-aware regularization strategies, and demonstrate robust cross-domain generalization on QM9 through correlation-guided task grouping across 12 quantum-chemical properties.
3. We develop a high-quality BDG prediction dataset encompassing diverse doping concentrations and configurations, providing a comprehensive data foundation for advancing AI-driven performance modeling methods for two-dimensional materials.

## Results

### Results on the BDG dataset

This subsection details the performance evaluation of the AEGCNN-MTL model on the constructed BDG dataset, which comprises 2613 samples of two-dimensional graphene supercell structures featuring diverse boron doping configurations. Each sample incorporates DFT calculated values for band gap and work function, serving as dual-task targets. Figure 1 illustrates the regression performance of the AEGCNN-MTL model on the BDG dataset for two key electronic properties. In both subplots, the x-axis represents the DFT-calculated reference values, and the y-axis denotes the

**Table 1 | Performance of GNN models on the BDG dataset**

Model	MAE↓ $E_{\text{gap}}$	RMSE↓ $E_{\text{gap}}$	$R^2$ ↑ $E_{\text{gap}}$	MAE↓ WF	RMSE↓ WF	$R^2$ ↑ WF
GATConv	0.0076	0.0172	0.9898	0.0155	0.0355	0.9683
GraphSAGE	0.0127	0.0237	0.9819	0.0299	0.0506	0.9354
GINConv	0.0123	0.0249	0.9800	0.0185	0.0324	0.9736
GCNII	0.0349	0.0459	0.9343	0.0501	0.0668	0.8939
SchNet	0.0097	0.0271	0.9640	0.0788	0.1034	0.7472
AEGCNN-MTL	<b>0.0031</b>	<b>0.0131</b>	<b>0.9905</b>	<b>0.0073</b>	<b>0.02700</b>	<b>0.9778</b>

Reported metrics are mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination ( $R^2$ ) for band gap ( $E_{\text{gap}}$ ) and work function (WF). The downward arrow indicates that a lower value denotes better performance, while the upward arrow indicates that a higher value denotes better performance. The best results for each metric are highlighted in bold.

**Table 2 | Prediction performance comparison of single-task and multi-task models on the BDG dataset**

Task	Type	MAE↓	RMSE↓	$R^2$ ↑
Band gap	Single-task	0.0049	0.0168	0.9870
Band gap	Multi-task	<b>0.0031</b>	<b>0.0131</b>	<b>0.9905</b>
Work function	Single-task	0.0084	0.0260	0.9799
Work function	Multi-task	<b>0.0073</b>	0.0270	0.9778

The bold values indicate the best performance among the compared methods for each property.

corresponding predictions made by the model. Blue and red markers indicate training and test samples, respectively. In Fig. 1a, the predicted band gap values show a well-aligned distribution along the identity line, demonstrating that the model can accurately approximate DFT-calculated band gaps across the full range of values. The dispersion of testing samples is limited, indicating effective generalization beyond the training set. The absence of outliers or skewed regions suggests that the model captures the influence of boron doping on the electronic band structure in a consistent manner. In Fig. 1b, a similarly strong correlation is observed for the prediction of work function. The data points follow the diagonal trend, with a slightly greater spread compared to the band gap subplot. Despite this, the alignment remains clear for both training and testing sets. This indicates that the model maintains reliable prediction performance even for surface-sensitive properties like the work function, which are often more challenging to model due to their sensitivity to local atomic environments.

To further assess the effectiveness of AEGCNN-MTL on the BDG dataset, we conducted comparative experiments against several widely used GNN baselines, including GATConv<sup>37</sup>, GraphSAGE<sup>38</sup>, GINConv<sup>39</sup>, GCNII<sup>40</sup>, and SchNet<sup>41</sup>. We did not include GemNet<sup>32</sup>, ALIGNN<sup>30</sup>, or equivariant Transformers<sup>42</sup> on the BDG benchmark because these architectures are primarily optimized for three-dimensional materials where angular or directional descriptors are essential. The chosen baselines span complementary paradigms and are widely used in both graph learning and materials informatics: GATConv represents attention-based message passing, GraphSAGE reflects inductive neighborhood sampling, GINConv is theoretically maximally expressive for distinguishing graph structures, GCNII addresses over-smoothing in deep GCNs, and SchNet is a chemistry-specific model that incorporates interatomic distances into message passing. All models were trained and evaluated under identical data splits, preprocessing, and optimization settings to ensure fairness. The results are summarized in Table 1. Although all baselines achieve good band gap accuracy, values of  $R^2$  greater than 0.90, their performance on work function scatters widely values of  $R^2$  from 0.75 to 0.97. GATConv attains an  $R^2$ -squared of 0.9898 for band gap prediction, whereas work function performance is much weaker for several baselines, with GCNII and SchNet reaching 0.8939 and 0.7472, respectively. AEGCNN-MTL simultaneously

delivers the lowest MAE for both properties, 0.0031 eV for band gap and 0.0073 eV for work function, and records the highest average  $R^2$  of 0.9842, demonstrating that the proposed multi-task architecture yields genuine gains over strong single-task baselines.

Furthermore, we evaluated the same architecture AEGCNN-MTL in two modes: single-task and multi-task. As shown in Table 2, for band gap the multi-task mode reduces MAE from 0.0049 to 0.0031 and RMSE from 0.0168 to 0.0131, with  $R^2$  increasing from 0.9870 to 0.9905. For work function, MAE decreases from 0.0084 to 0.0073, while RMSE changes from 0.0260 to 0.0270 and  $R^2$  from 0.9799 to 0.9778. These results provide a direct within-architecture comparison of single-task and multi-task training on the BDG dataset.

### Generalization evaluation on the QM9 dataset

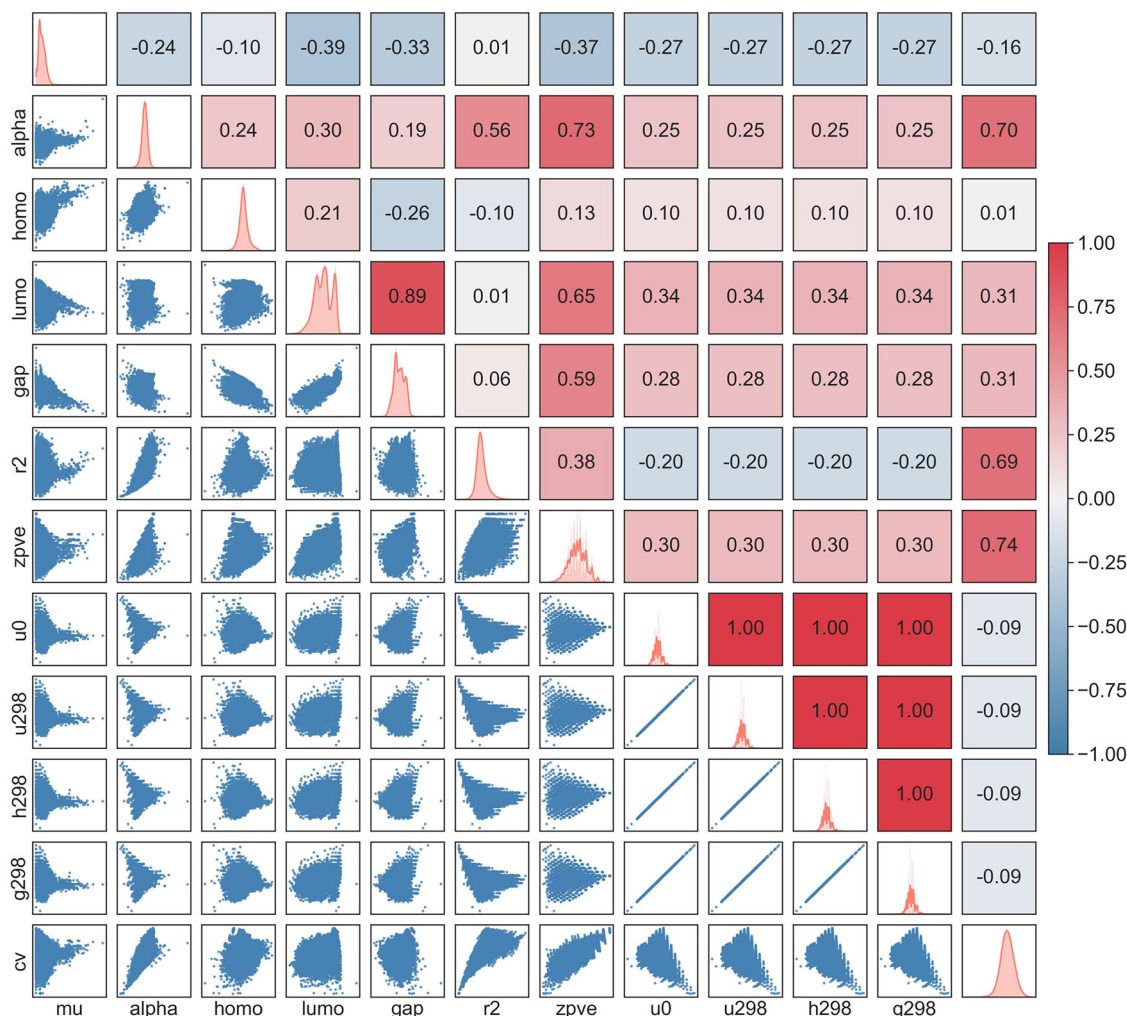
On the QM9 dataset, we evaluated the generalization capability of the proposed AEGCNN-MTL model. The 12 QM9 properties were divided into four groups according to their correlations and physical meanings (Fig. 2 for the correlation analysis and Table 3 for the grouping criteria).

All experiments adhere to the same training/test splits as applied to the QM9 dataset, with performance evaluated using the MAE metric. Baseline models included several classical graph neural networks such as SchNet<sup>41</sup>, MEGNet<sup>29</sup>, PhysNet<sup>43</sup>, DimeNet<sup>44</sup>, ET<sup>45</sup>, R-MAT<sup>26</sup>, and GeoT<sup>46</sup>, as shown in Table 4. The performance of AEGCNN-MTL within each group is analyzed as follows. Group A includes isotropic polarizability ( $\alpha$ ) and heat capacity at 298 K ( $C_v$ ), which are moderately correlated and primarily influenced by the overall molecular size and bonding flexibility. The model achieved MAEs of 0.052  $a_0^3$  for  $\alpha$  and 0.030 cal/mol-K for  $C_v$ , comparable to those reported by strong baselines such as DimeNet and GeoT.

These results suggest that, group B contains four thermodynamic properties, atomization energy at 0 K ( $U_0$ ), at 298 K ( $U$ ), enthalpy ( $H$ ), and Gibbs free energy ( $G$ ), which are closely related and show extremely high pairwise correlations (Pearson  $r > 0.99$ ). AEGCNN-MTL yielded MAEs of 10.3, 10.5, 10.2, and 10.7 meV for  $U_0$ ,  $U$ ,  $H$ , and  $G$ , respectively. These results are comparable to those of leading single-task models. Group C includes HOMO, LUMO, and the HOMO-LUMO energy gap ( $\Delta_e$ ), which describe the frontier orbital characteristics of molecules. Among them, HOMO and  $\Delta_e$  exhibit a particularly strong Pearson correlation ( $r = 0.89$ ). Within this group, AEGCNN-MTL achieved MAEs of 23.1 meV for HOMO and 32.1 meV for HOMO-LUMO energy gap, one of which is the lowest reported MAE across all baselines evaluated. As for LUMO, the model achieved an MAE of 22.0 meV, which outperforms the majority of baseline models, including SchNet (34.0 meV), MEGNet (31.0 meV), PhysNet (24.7 meV), and R-MAT (29.0 meV). These results demonstrate that, when the target properties are physically and statistically coupled, multi-task learning can lead to meaningful improvements in prediction accuracy through shared representation learning. Group D includes dipole moment ( $\mu$ ), electronic spatial extent ( $R^2$ ), and zero-point vibrational energy (ZPVE). These properties exhibit weak correlation with other tasks in the dataset, respectively. In preliminary trials, joint training of these tasks with others resulted in unstable convergence and reduced accuracy. As a result, AEGCNN-MTL employed its single-task mode for this group. Under this configuration, the model achieved MAEs of 0.053 D for  $\mu$ , 0.832  $a_0^3$  for  $R^2$ , and 1.90 meV for ZPVE. While these values are slightly higher than those of some baselines, the outcomes remain within an acceptable range. The relatively lower performance in Group D, compared to Groups B and C, is primarily attributed to the absence of substantial inter-task correlations, which constrained the model's ability to benefit from its multi-task learning architecture. This finding reinforces the importance of task-relatedness for effective shared representation learning and highlights the strength of AEGCNN-MTL in exploiting interdependent tasks for improved predictive performance.

### Ablation Studies

To ascertain the individual contributions of key modules within the AEGCNN-MTL architecture, a series of ablation experiments were



**Fig. 2 | Correlation and distribution analysis of molecular properties in the QM9 dataset.** Upper triangle: heat map of Pearson coefficients, where red indicates positive correlation, blue indicates negative correlation, and near-white indicates weak correlation; cell numbers show the coefficient values. Diagonal panels: red histograms and density curves show the marginal distributions. Lower triangle: blue scatter plots show pairwise relationships. The color bar on the right encodes correlation strength from negative one to positive one.

**Table 3 | Task grouping of 12 molecular properties in the QM9 dataset**

Task Group	Property	Unit	Property description	Group description
A	$\alpha$	$\text{a}_0^3$	Isotropic polarizability	Polarizability and Heat Capacity Properties
	$C_v$	$\frac{\text{cal}}{\text{molK}}$	Heat capacity at 298 K	
B	u0	eV	Atomization energy at 0 K	Thermochemical Properties
	u298	eV	Atomization energy at 298 K	
	h298	eV	Atomization enthalpy at 298 K	
	g298	eV	Atomization Free energy at 298 K	
C	$\epsilon_{LUMO}$	eV	LUMO energy	Frontier Orbital Energies and Energy Gap
	$\epsilon_{HOMO}$	eV	HOMO energy	
	$\Delta\epsilon$	eV	HOMO-LUMO gap	
D	$\mu$	D	Dipole moment	Independent Properties
	$R^2$	$\text{a}_0^3$	Electronic spatial extent	
	ZPVE	eV	Zero-point vibrational energy	

The properties are categorized into four task groups (A-D) according to Pearson correlation analysis and their underlying physical meanings.

**Table 4 | Comparison of MAE for 12 molecular property prediction tasks on the QM9 dataset**

properties	unit	SchNet	MEGNet	Physnet	DimeNet	ET	R-MAT	GeoT	Ours
$\mu$	D	0.033	0.040	0.0529	0.0286	0.011	0.110	0.0297	0.053
$\alpha$	$a_0^3$	0.235	0.083	0.0615	0.0469	0.059	0.082	0.052	0.052
$\epsilon_{HOMO}$	meV	41.0	38.0	32.9	27.8	20.3	31	25	23.1
$\epsilon_{LUMO}$	meV	34.0	31.0	24.7	19.7	17.5	29	20.2	22.0
$\Delta_\epsilon$	meV	63.0	61.0	42.5	34.8	36.1	48	43	32.1
$R^2$	$a_0^3$	0.073	0.265	0.765	0.331	0.033	0.676	0.30	0.832
ZPVE	meV	1.70	1.40	1.39	1.29	1.84	2.23	1.70	1.90
$U_0$	meV	14.0	9.0	8.15	8.02	6.16	12.0	11.1	10.3
U	meV	19.0	10.0	8.34	7.89	6.38	10.0	11.7	10.5
H	meV	14.0	10.0	8.42	8.11	6.16	10.0	11.3	10.2
G	meV	14.0	10.0	9.40	8.98	7.62	10.0	11.7	10.7
$C_v$	$\frac{cal}{molK}$	0.0330	0.030	0.0280	0.0249	0.026	0.036	0.0276	0.030

The results of the proposed AEGCNN-MTL model (Ours) are compared against several representative baselines, including SchNet, MEGNet, PhysNet, DimeNet, ET, R-MAT, and GeoT. Reported MAEs are computed using consistent training/test splits of QM9, with bold numbers highlighting the best performance for each property.

**Table 5 | Ablation model configurations and module composition**

Model	GCN	Mean Pool	Set2set	AEGC	MHA	FFn
M1	✓	✓				shared
M2	✓		✓			shared
M3			✓	✓		shared
M4			✓	✓	✓	shared
AEGCNN-MTL			✓	✓	✓	specific

The table summarizes different ablation variants (M1–M4) compared with the full AEGCNN-MTL model. Each variant selectively removes or replaces key components, including Graph Convolutional Network (GCN), Mean Pooling, Set2Set pooling, Adaptive Edge-aware Graph Convolution (AEGC), Multi-Head Attention (MHA), and feed-forward networks (FFn). Variants employ shared FFns across tasks, while the complete AEGCNN-MTL adopts task-specific FFns to enhance multi-task learning capacity.

**Table 6 | Performance comparison of ablation models on band gap and work function prediction tasks**

Model	MAE↓ $E_{gap}$	RMSE↓ $E_{gap}$	$R^2$ ↑ $E_{gap}$	MAE↓ WF	RMSE↓ WF	$R^2$ ↑ WF
M1	0.0142	0.0264	0.9771	0.0184	0.0353	0.9687
M2	0.0102	0.0252	0.9764	0.0129	0.0305	0.9774
M3	0.0086	0.0211	0.9827	0.0126	0.0325	0.9739
M4	0.0040	0.0170	0.9888	0.0075	0.0272	0.9795
AEGCNN-MTL	<b>0.0031</b>	<b>0.0131</b>	<b>0.9905</b>	<b>0.0073</b>	<b>0.0270</b>	0.9778

The table reports the predictive accuracy of ablation variants (M1–M4) and the full AEGCNN-MTL model. Metrics include mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination ( $R^2$ ) for both band gap ( $E_{gap}$ ) and work function (WF). Bold numbers indicate the best performance.

conducted on the BDG dataset. These experiments systematically constructed and compared five sub-models (M1–M4 and the full model) to evaluate the roles of the AEGC, MHA, and task-specific modeling components (Set2Set pooling and independent Feedforward Networks) in multi-task crystal property prediction.

Table 5 summarizes the module composition of each model. Model M1 serves as the baseline architecture, employing traditional GCN with mean

pooling and a shared feedforward network. Model M2 leverages Set2Set pooling to generate richer and more expressive graph-level representations by adaptively aggregating node features across the entire graph structure. Model M3 integrates the AEGC module, which strengthens the network's capability to capture local structural information by incorporating edge-weighted interactions that reflect the underlying chemical environments. Model M4 employs a distance-modulated attention mechanism to effectively capture long-range dependencies within the crystal structure, enabling the model to learn global relational patterns critical for property prediction. Finally, the complete AEGCNN-MTL framework integrates task-specific FFns, allowing the model to generate personalized predictions tailored to each target property within the multi-task learning setting.

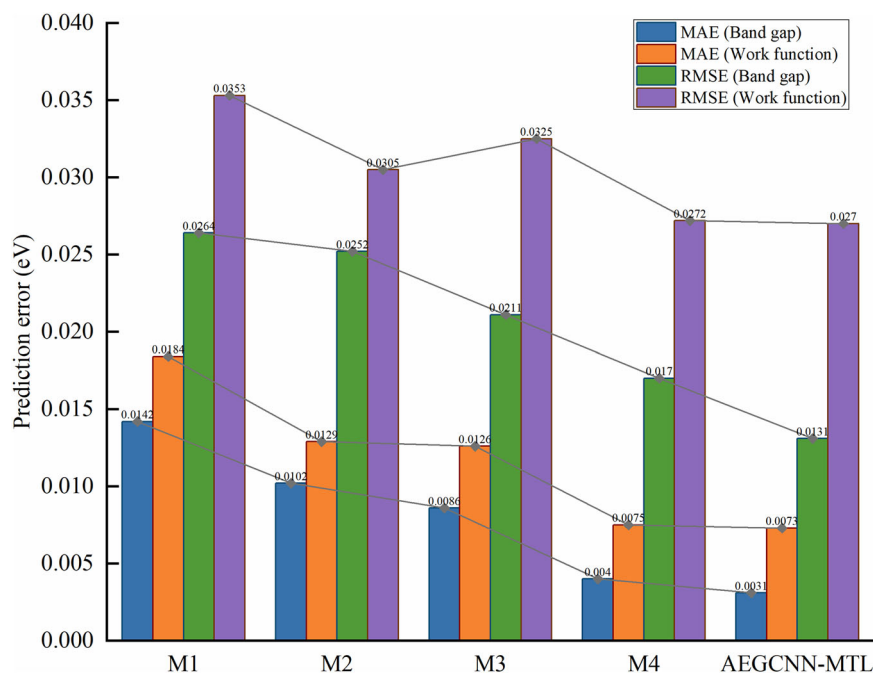
Table 6 presents the performance results for the  $E_{gap}$  and WF regression tasks, evaluated using MAE, Root RMSE, and  $R^2$  metrics. The incorporation of Set2Set pooling (M2) significantly reduced the MAE by approximately 29% for  $E_{gap}$  and 26% for WF compared to M1, indicating the effectiveness of Set2Set pooling in capturing graph-level structural variations. The introduction of the AEGC (M3) led to further improvements, with MAE reduced by an additional 16% for  $E_{gap}$  and 5% for WF compared to M2, highlighting the benefit of edge-weighted interactions in local structural modeling. The attention mechanism (M4) resulted in noticeable enhancements in  $R^2$ , especially for the band gap task, reflecting the advantage of modeling global dependencies for handling long-range structural effects. The full AEGCNN-MTL model, equipped with task-specific decoders, delivered the most exceptional overall performance, with MAE values reduced to 0.0031 for  $E_{gap}$  and 0.0073 for WF.

Figure 3 presents the ablation analysis of AEGCNN-MTL on the BDG dataset, showcasing the progressive impact of individual architectural modules on the prediction performance of two key electronic properties: band gap and work function. The figure illustrates both MAE and RMSE values across five model configurations M1 to M4, and the complete AEGCNN-MTL using bar and line plots. Specifically, blue and orange bars represent MAE for band gap and work function, respectively, while green and purple bars depict their corresponding RMSE values. A clear downward trend in both MAE and RMSE is observed as the model evolves from M1 to the complete version, confirming that each module contributes to reducing prediction error.

## Discussion

The results of this study demonstrate that the proposed AEGCNN-MTL framework can achieve accurate and robust prediction of multiple material properties by combining adaptive edge-aware convolution, distance-modulated attention, and task-specific decoders within a

**Fig. 3 | Ablation results of AEGCNN-MTL on the BDG dataset.** Prediction errors for band gap and work function across different models: M1–M4 and the AEGCNN-MTL. Blue bars represent the mean absolute error (MAE) of band gap, orange bars represent the MAE of work function, green bars represent the root mean square error (RMSE) of band gap, and purple bars represent the RMSE of work function. Error values consistently decrease as modules are added.



unified multi-task architecture. On the proprietary BDG dataset, the model achieves state-of-the-art prediction performance for both band gap and work function, consistently surpassing representative GNN baselines. Moreover, a direct comparison of single-task and multi-task training within the same architecture reveals that shared representation learning provides a marked advantage for band gap prediction, while maintaining competitive accuracy for the more surface-sensitive work function. These findings suggest that multi-task learning can be particularly beneficial when multiple properties share common structural descriptors, whereas its gains may be limited when the target properties are governed by distinct local environments.

The evaluation on the QM9 dataset provides further insights into the scope of multi-task learning. For property groups with strong physical and statistical correlations, such as the thermodynamic energy set ( $U_0$ ,  $U$ ,  $H$ ,  $G$ ) and the frontier orbital set (HOMO, LUMO, HOMO-LUMO energy gap), the model benefits significantly from parameter sharing and yields lower prediction errors than strong single-task baselines. This indicates that the proposed architecture can effectively capture latent physical regularities underlying related properties and exploit them for joint learning. In contrast, properties with weak coupling, such as dipole moment ( $\mu$ ), electronic spatial extent ( $R^2$ ), and zero-point vibrational energy (ZPVE), exhibit limited benefits from joint optimization. In these cases, single-task training proves more stable and accurate, underscoring the importance of task-relatedness as a prerequisite for successful multi-task modeling. Importantly, the ability of AEGCNN-MTL to flexibly accommodate both joint and independent training modes within the same framework mitigates the risk of negative transfer and enhances its adaptability across diverse property sets.

The ablation studies further elucidate the contributions of individual architectural components. The introduction of Set2Set pooling improves graph-level representation quality by adaptively aggregating structural information across nodes. The adaptive edge-aware convolution strengthens local chemical environment modeling through edge-weighted interactions, while the distance-modulated attention mechanism enhances the capture of long-range structural dependencies. Finally, the task-specific decoders ensure that property-specific features are retained during multi-task optimization, preventing overly generic representations. The progressive improvement observed across these modules confirms that the superior performance of the full AEGCNN-

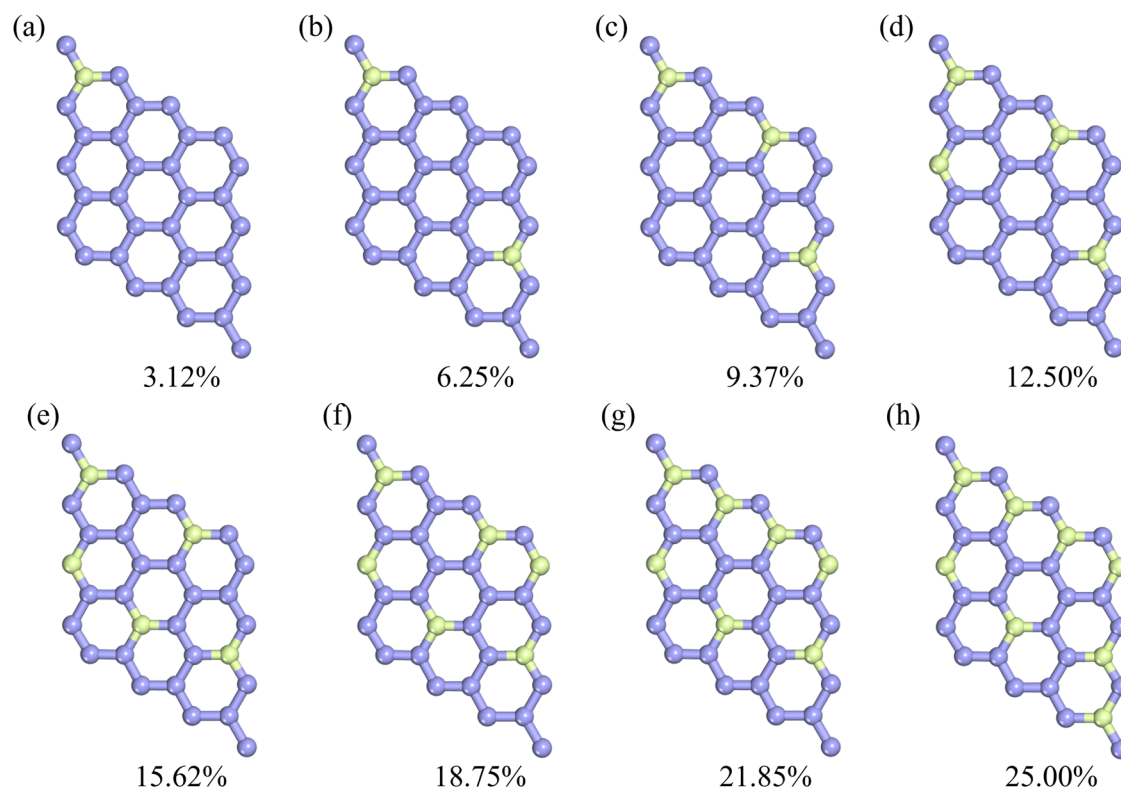
MTL model arises from the synergistic integration of local, global, and task-specific learning mechanisms.

In this work, we proposed an AEGCNN-MTL to tackle the challenge of simultaneously predicting multiple material properties. The model introduces edge feature encoding and a dynamic attention-based weighting mechanism to effectively integrate topological and relational information during message propagation. In addition, AEGCNN-MTL incorporates task-specific Set2Set pooling and a gated attention fusion module to better capture inter-task dependencies, demonstrating strong predictive performance and interpretability. Comprehensive evaluations demonstrate the efficacy of AEGCNN-MTL. On our newly constructed BDG dataset, the model achieves state-of-the-art accuracy for the simultaneous prediction of band gap ( $R^2 = 0.9905$ ) and work function ( $R^2 = 0.9778$ ), significantly outperforming single-task models and highlighting the benefits of shared representation learning and task co-regularization. More importantly, AEGCNN-MTL exhibits exceptional generalization capability. When evaluated on the diverse QM9 benchmark encompassing 12 quantum chemical properties, it delivers performance comparable to or surpassing specialized state-of-the-art single-task GNNs across most tasks. Given the modularity and scalability of AEGCNN-MTL, it can be as a powerful tool for high-throughput screening and accelerated discovery of materials with tailored multi-functional properties. Future work will focus on enriching the edge encoding with explicit domain knowledge, exploring the model's potential for inverse materials design, and integrating it with generative frameworks to enable the autonomous discovery of novel multi-property-optimized materials.

## Methods

### Construction of Boron-doped graphene dataset

To systematically investigate the effects of varying doping concentrations and configurations on the electronic properties of BDG, we constructed a high-quality dataset using high-throughput DFT calculations. The dataset encompasses a broad range of representative doping scenarios. All configurations were constructed using a  $4 \times 4 \times 1$  graphene supercell containing 32 carbon atoms, which was used as the structural template for boron substitution. As shown in Fig. 4, a diverse set of doping configurations was generated by substituting 1 to 8 carbon atoms with boron in unique, non-redundant combinations, resulting in doping concentrations ranging from approximately 3.1% to 25.0%. For each concentration level, multiple



**Fig. 4 | Representative atomic structures of BDG with varying boron concentrations.** Purple spheres denote carbon atoms and green spheres denote boron atoms; all panels show a  $4 \times 4 \times 1$  graphene supercell. **a** 3.12% **(b)** 6.25% **(c)** 9.37% **(d)** 12.50% **(e)** 15.62% **(f)** 18.75% **(g)** 21.88% **(h)** 25.00%.

spatially distinct configurations were considered, resulting in 747 unique BDG structures. (a–h) Atomic models illustrate BDG configurations with different numbers of boron atoms (highlighted in green) substituted into a  $4 \times 4 \times 1$  graphene supercell (carbon atoms shown in purple). These structures exemplify the diversity of doping configurations and concentrations explored in this study.

To leverage the inherent high spatial symmetry of the graphene lattice, we further applied a symmetry expansion strategy by employing mirror operations to generate equivalent configurations. As illustrated in Fig. 5, this symmetry expansion substantially improves the dataset's coverage of the configurational space. In Fig. 5a, original BDG configuration with multiple boron atoms (green) is substituted into a  $4 \times 4 \times 1$  graphene supercell (carbon atoms in purple). In Fig. 5b–d, new configurations obtained by applying different mirror symmetry operations (black dashed lines) to the original structure, illustrating how symmetry expansion increases the diversity of equivalent doping arrangements in the dataset. After eliminating redundant structures, we finalized a comprehensive BDG dataset comprising 2613 samples.

### Computational methods and parameter settings

All first-principles calculations were performed using the Vienna Ab initio Simulation Package (VASP) within the framework of DFT. To avoid the spurious interaction between graphene layers, a vacuum thickness in the supercell was set to 15.0 Å. The planar lattice constant of graphene supercell was fixed at  $4a_0$ , where  $a_0 = 2.468$  Å for the calculated lattice constant of graphene. The interaction between ions and valence electrons was described by the projector-augmented wave method, and the exchange–correlation energy was treated with the Perdew Burke Ernzerhof (PBE) functional under the generalized gradient approximation. The plane-wave basis set was truncated with a kinetic energy cutoff of 450 eV, which was validated through convergence testing. The electronic self-consistent field (SCF) calculations were performed with an energy convergence criterion of  $1 \times 10^{-6}$  eV, while geometry optimization was conducted until the residual

atomic forces were less than 0.01 eV/Å. Although the PBE functional is known to underestimate absolute band gap values, it has been shown to reliably reproduce the doping-dependent band gap trends in boron-doped graphene systems, serving as a good approximation for electronic property modulation in such materials<sup>47</sup>.

The electronic band gap ( $E_{\text{gap}}$ ) was computed for all structures as the energy difference between the conduction band minimum (CBM) and valence band maximum (VBM), extracted from the calculated electronic band structures:

$$E_{\text{gap}} = E_{\text{CBM}} - E_{\text{VBM}} \quad (1)$$

where  $E_{\text{CBM}}$  and  $E_{\text{VBM}}$  denote the energies of the valence band maximum and conduction band minimum, respectively.

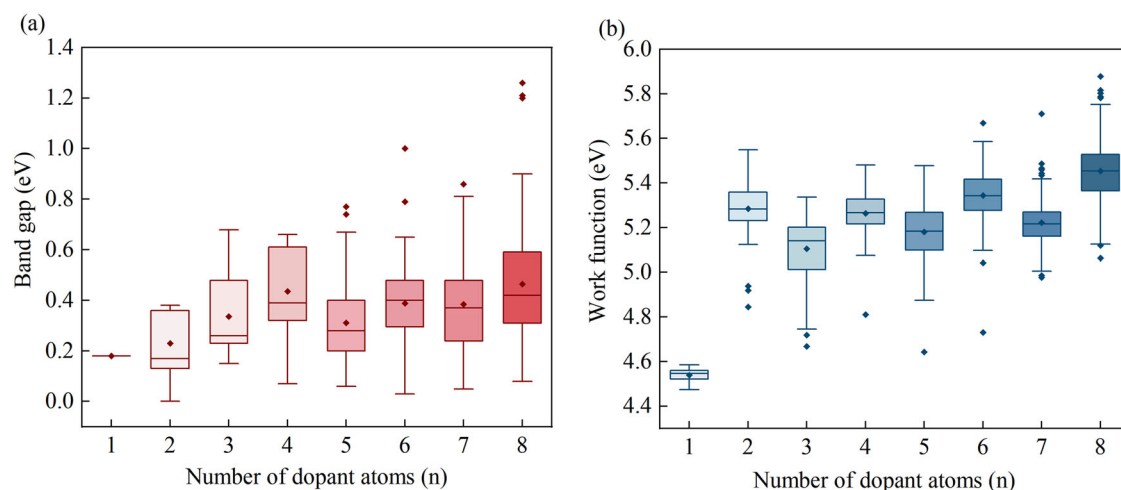
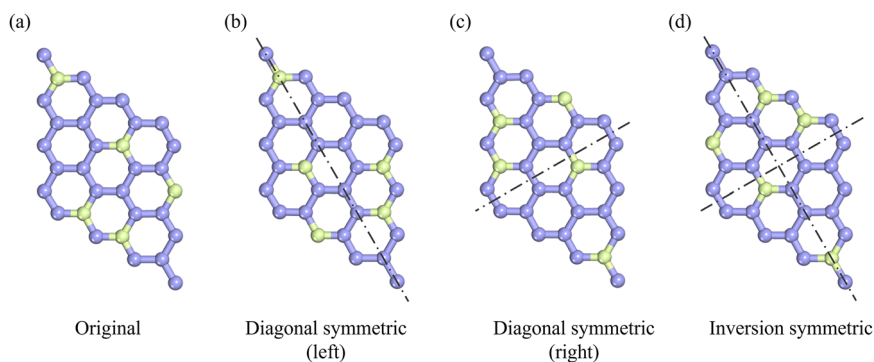
The work function ( $\Phi$ ) was computed via the vacuum slab method as the difference between the vacuum electrostatic potential and the Fermi level:

$$\Phi = V_{\text{vacuum}} - E_{\text{Fermi}} \quad (2)$$

where  $V_{\text{vacuum}}$  is the average vacuum potential in the region perpendicular to the graphene surface (obtained through planar averaging), and  $E_{\text{Fermi}}$  is the Fermi level from the SCF calculation.

The calculated band gap and work functions across different doping concentrations agree well with previous reports<sup>2–4</sup>. Figure 6 presents the relationships between the number of dopant atoms and the calculated properties. Figure 6a reveals a general increasing trend in band gap with rising boron concentration, suggesting a strong correlation between doping level and electronic structure modulation. Notably, configurations containing 4 and 8 dopants exhibit the broadest band gap distributions, with maximum values exceeding 1.2 eV, indicating a pronounced sensitivity to atomic configuration at higher doping levels. Figure 6b shows a monotonic increase in work function from approximately 4.45 eV to over 5.5 eV with

**Fig. 5 | Generation of symmetry-equivalent boron-doped graphene (5 boron atoms) configurations via lattice symmetry operations.** Purple spheres denote carbon atoms and green spheres denote boron atoms; black dashed lines indicate symmetry axes. **a** Original configuration. **b** Diagonal symmetric (left). **c** Diagonal symmetric (right). **d** Inversion-symmetric configuration.



**Fig. 6 | Boxplot of band gap and work function under different boron doping numbers.** **a** Boxplots of band gap values for BDG with 1-8 boron atoms. Red boxes indicate the distribution range, with central marks denoting the median; **b** Boxplots

of work function values for BDG with 1-8 boron atoms. Blue boxes represent the distributions, with central marks denoting the median.

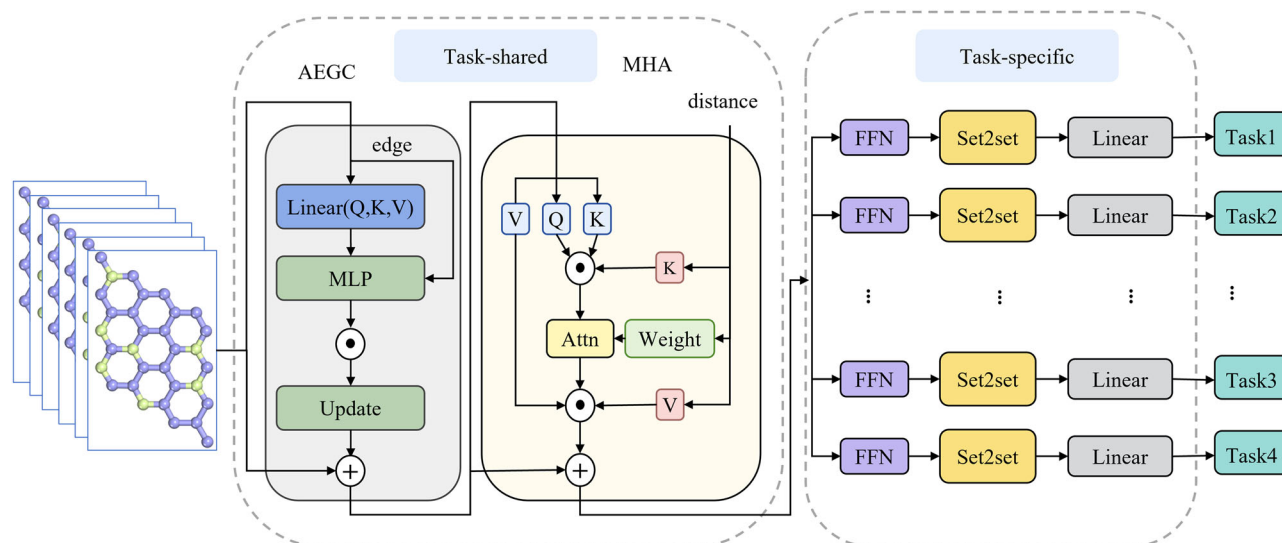
increasing boron content, consistent with the p-type doping effect and associated upward shift in surface potential. The observed variability in both band gap and work function arises from the diversity of local atomic environments, which modulate  $\pi$ -electron delocalization and influence the surface electrostatic potential. While both the band gap and work function originate from the electronic structure, they represent fundamentally different physical quantities. The band gap is a bulk property defined by the energy difference between the conduction and valence-band edges, whereas the work function is surface-sensitive and defined relative to the vacuum electrostatic potential. This distinction is reflected in the dataset: the band gap exhibits large variance at intermediate doping levels, while the work function increases nearly monotonically with boron concentration. The contrasting statistical behaviors demonstrate that their simultaneous prediction is a nontrivial multi-task challenge, rather than a simple outcome of their shared electronic origin.

### QM9 dataset

The QM9 dataset is a widely used benchmark in quantum chemical modeling and machine learning, comprising approximately 134,000 small organic molecules containing up to nine heavy atoms (C, N, O, and F). All molecular structures and their quantum chemical properties were calculated using DFT methods<sup>5</sup>. The dataset offers comprehensive molecular property data, including structural parameters, orbital energies, thermodynamic functions, and electronic responses, covering over 20 physicochemical properties. It serves as a crucial foundation for research in molecular representation, property prediction, transfer learning, and multi-task learning<sup>6,48,49</sup>. In this work, we select 12 representative quantum chemical

properties from QM9 as prediction targets for multi-task learning. To better understand the interdependencies among prediction targets and to inform subsequent model design, we applied Pearson correlation analysis to evaluate the statistical relationships between quantum chemical properties in the QM9 dataset.

As illustrated in Fig. 2, the upper triangle presents pairwise Pearson correlation coefficients, color-coded from blue (strongly negative) to red (strongly positive), indicating the degree of linear association between each property pair. The diagonal plots show the marginal distributions of each property, while the lower triangle displays scatter plots that capture nonlinear trends and distributional patterns. This tri-level visualization facilitates a detailed assessment of both global statistical dependencies and local variances within the dataset. Several notable correlation patterns are evident. Thermodynamic properties, including  $u_0$ ,  $u_{298}$ ,  $h_{298}$ , and  $g_{298}$ , exhibit near-perfect linear correlations ( $r \approx 1.00$ ), reflecting their shared physical basis in atomization energy computations. Likewise, strong correlations are observed among frontier orbital features LUMO, HOMO, and the energy gap  $\Delta\epsilon$ , with the LUMO,  $\Delta\epsilon$  pair showing the highest correlation ( $r = 0.89$ ). In contrast, properties such as dipole moment ( $\mu$ ), heat capacity ( $C_v$ ), and electronic spatial extent ( $R^2$ ) display weak or negligible correlations with other variables ( $|r| < 0.3$ ), suggesting limited information redundancy. These observations indicate that highly correlated tasks are suitable for grouped multi-task training, leveraging shared representations to enhance learning efficiency. Conversely, tasks exhibiting weak correlations or large disparities in numerical scale may introduce interference or imbalance when trained jointly, and are thus better handled independently.



**Fig. 7 | Overall architecture and modular decomposition of the AEGCNN-MTL model.** The model is composed of three major parts. The adaptive edge-aware graph convolution module (left) projects edge features through linear layers and a multilayer perceptron, followed by feature update with residual addition. The task-shared distance-modulated multi-head attention module (middle) computes query,

key, and value interactions weighted by interatomic distances to capture both short- and long-range dependencies. The task-specific decoders (right) include feed-forward networks, Set2Set pooling layers, and linear heads to generate predictions for individual target properties.

Guided by this correlation analysis and further informed by the physical interpretation, units, and statistical ranges of each property, we divided the 12 prediction targets into four semantic groups, as outlined in Table 3. Each task group reflects a specific category of physical or chemical properties to facilitate multi-task modeling and analysis. Group A consists of  $\alpha$  (isotropic polarizability) and  $C_v$  (heat capacity), which are both response properties despite their weak coupling. Group B encompasses thermodynamic quantities ( $u_0$ ,  $u_{298}$ ,  $h_{298}$ ,  $g_{298}$ ), which form a strongly interdependent set. Group C includes orbital energy-related features (HOMO, LUMO,  $\Delta\epsilon$ ), and Group D comprises tasks such as  $\mu$ ,  $R^2$ , and ZPVE, which demonstrate minimal inter-task correlation and exhibit considerable differences in physical nature and data magnitude. Consequently, Group D tasks were treated as standalone single-task models to avoid performance degradation due to negative transfer. This grouping strategy ensures semantic coherence within task clusters and provides a structured basis for heterogeneous multi-task learning.

AEGCNN-MTL was designed to address the challenges of predicting multiple properties in molecular and crystalline systems. The architecture of the proposed AEGCNN-MTL model encompasses feature embedding, graph structure encoding, multi-task decoders, and a joint optimization mechanism. As illustrated in Fig. 7, the AEGCNN-MTL framework follows a “shared encoder + task-specific decoder” design principle and is composed of three functional stages: feature embedding, task-shared representation learning, and task-specific prediction heads.

In the first stage, atomic graphs constructed from molecular or crystalline structures are processed through a multi-channel embedding module that encodes atomic number, degree, spatial position, and bond type into learnable feature vectors. These embeddings form the input to the task-shared backbone. The second stage comprises two complementary modules: an Adaptive Edge-aware Graph Convolution module and a Multi-head Attention Layer module. The AEGC module dynamically integrates edge features and node features using an attention mechanism, enabling refined local message passing sensitive to chemical bonding environments. Following this, the MHA module captures long-range dependencies via an attention mechanism, which is further modulated by interatomic distances encoded through radial basis functions. Together, these two components enable joint modeling of local and global structural contexts, yielding expressive graph-level representations. Finally, the third stage employs task-

specific decoders to address the heterogeneity among different prediction targets. Each property is processed through a dedicated pathway comprising a feedforward network, a Set2Set pooling layer, and a linear output head. This design ensures that the model can disentangle shared representations while accommodating task-specific feature distributions. To mitigate optimization conflicts across tasks, a synchronized loss function is introduced to balance task-specific gradients and stabilize multi-task training.

### Adaptive Edge-aware graph convolution

In the AEGCNN-MTL model, each molecular graph is constructed based on atomic connectivity, where nodes represent atoms and edges denote interatomic bonds. The encoding strategy for nodes and edges adopted in AEGCNN-MTL is illustrated in Fig. 8. Node features consist of three key components: atomic type ( $z$ ), degree ( $d$ ), and positional encoding (PE). These features are transformed into  $ddd$ -dimensional embeddings using multilayer perceptrons (MLP), as defined below:

$$h_z = \text{Emb}_z(z), h_d = \text{Emb}_d(d), h_{pe} = \text{MLP}(PE) \quad (3)$$

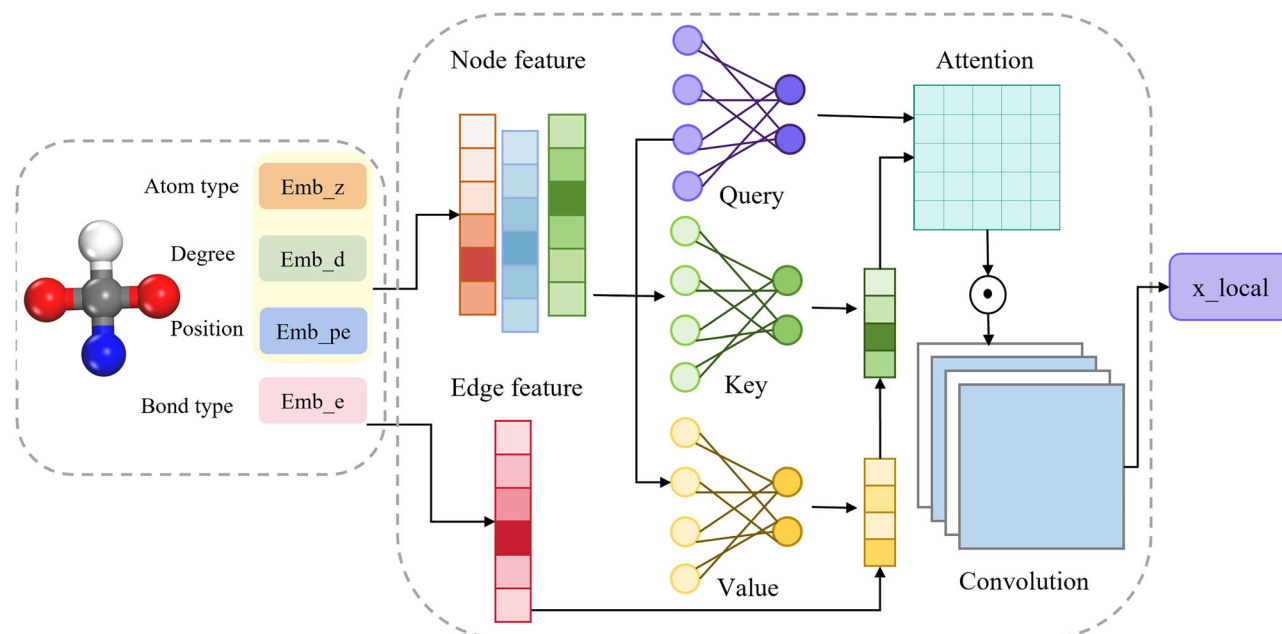
where  $\text{Emb}_z$  and  $\text{Emb}_d$  denote the embedding functions for atomic type and degree, respectively, and MLP represents a multilayer perceptron applied to the positional encoding.

The initial representation of node  $i$  was obtained by summing these embeddings:

$$x_i^{(0)} = h_z + h_d + h_{pe} \quad (4)$$

We introduce the AEGC layer to enhance the model’s capacity to capture fine-grained atomic interactions and bond-specific dependencies within molecular graphs. As a core component of the GNN backbone, AEGC integrates node-level semantics with edge-level structural cues. This fusion enables the generation of localized, context-aware representations critical for downstream property prediction.

The AEGC module enables unified message passing by incorporating both node and edge features, as illustrated in Fig. 8. Distinctively, the AEGC layer incorporates an attention mechanism that adaptively adjusts the weights assigned to messages exchanged between neighboring atoms. In this process, node features are projected into query (Q), key (K), and value (V)



**Fig. 8 | Architecture of the AEGC module.** On the left, four types of input descriptors are shown: atom type, degree, position, and bond type. These are converted into embeddings through corresponding embedding layers (orange, green, blue, and red rectangular blocks). The embedded node features are projected into three branches, represented by purple, green, and yellow networks, which produce

query, key, and value representations. Edge features are integrated into the key and value branches. The resulting query, key, and value are combined within the attention module, represented by the blue grid, to calculate attention weights. These weights are applied to the value branch to perform weighted message passing, shown as stacked blue blocks labeled as convolution.

vectors, while edge features are embedded via a learnable MLP and incorporated into the key and value branches. This dynamic attention weighting enables the model to better capture variations in chemical environments and to learn context-sensitive local topological representations. By explicitly modeling edge-aware interactions, the AEGC module enhances the network's representational capacity for capturing bond-order-sensitive patterns and structural anisotropy.

Specifically, for each edge  $(i, j)$  in the graph, node features were first projected into Q, K, and V vectors:

$$q_i = W_q x_i, k_j = W_k x_j, v_j = W_v x_j \quad (5)$$

If an edge feature  $e_{ij}$  existed, it was further embedded via an MLP to enrich the key and value representations:

$$\tilde{k}_j = k_j + MLP_{edge}(e_{ij}), \tilde{v}_j = v_j + MLP_{edge}(e_{ij}) \quad (6)$$

The attention weight  $\alpha_{ij}$  between nodes  $i$  and  $j$  was computed using scaled dot-product attention, followed by normalization across all neighbors of node:

$$\alpha_{ij} = softmax\left(\frac{q_i \cdot \tilde{k}_j}{\sqrt{d}}\right) \quad (7)$$

The aggregated message for node  $i$  was a weighted sum of its neighbors' values:

$$m_i = \sum_{j \in N(i)} \alpha_{ij} \cdot \tilde{v}_j \quad (8)$$

The final representation of node  $i$  is updated via a nonlinear transformation, residual connection, and batch normalization:

$$x'_i = BN(x_i + MLP_{update}(m_i)) \quad (9)$$

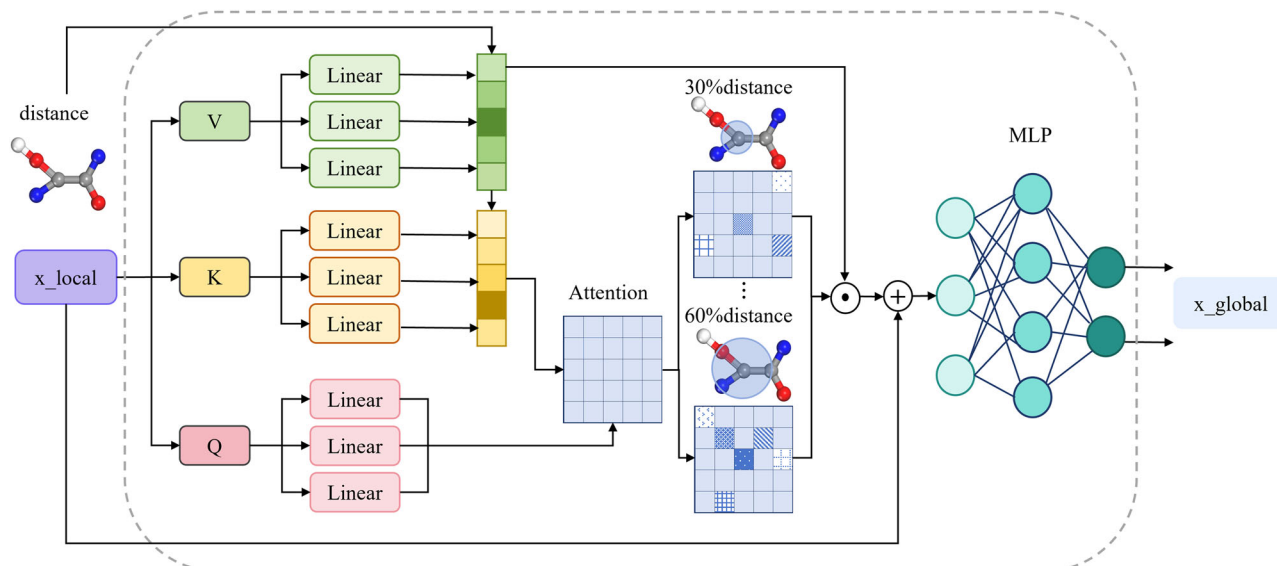
To model the evolution of edge features, AEGC included an edge update path. The updated edge embedding depends on both node features and the original edge attributes:

$$e'_{ij} = MLP_{e-update}\left([x_i || x_j || e_{ij}]\right) \quad (10)$$

Through this design, the AEGC module enables structure-aware modeling via node-edge interactions, enhancing expressive power and physical consistency on complex crystal structures and heterogeneous molecular graphs, thereby supporting accurate property prediction in multi-task learning.

### Multi-head attention layer

To capture long-range and higher-order interactions, we introduce a multi-head attention layer informed by interatomic geometric distances. The input node representations are projected into multiple attention heads, each of which linearly transforms the features into separate Q, K, and V vectors. This layer was constructed based on a Transformer framework<sup>50</sup>. This module addresses the receptive field constraints of traditional graph neural networks, which rely primarily on local neighborhood aggregation. By explicitly incorporating interatomic distance information, the model is able to capture global interactions between distant atoms, which is particularly important for modeling phenomena such as conjugated systems, long-range polarization, and electronic delocalization.



**Fig. 9 | Architecture of the multi-head attention layer.** The module converts local atomic features into query, key, and value representations using parallel linear layers, illustrated as rectangular boxes. a grid denotes the attention matrix that computes interaction weights. Distance cues modulate the heads (molecular icons with percentage labels indicate different spatial ranges only, not specific values). Outputs are aggregated by residual addition (circles with plus signs) and passed to a multi-layer perceptron (network of connected circles) to form the global feature.

As illustrated in Fig. 9, the MHA module takes the node-level representations  $x_{\text{local}}$  as input and first projects them into Q, K, and V vectors through three independent linear transformations. These vectors are computed in parallel across attention heads, enabling the model to learn diverse interaction patterns from different relational perspectives. Simultaneously, the Euclidean distances between atom pairs are embedded using radial basis functions (RBF) to produce continuous, multi-scale geometric descriptors. These RBF-based embeddings are incorporated into the attention mechanism to modulate interaction strengths based on interatomic proximity. Within each attention head, distance-aware key and value vectors are used to compute attention weights, which are further modulated by a cosine-based cutoff function. This mechanism enables the model to prioritize short-range interactions while preserving physically meaningful long-range dependencies. The resulting weighted messages are aggregated and fused across heads via concatenation or summation. To ensure training stability and information retention, the output is further processed via residual connections and a feedforward MLP, followed by layer normalization.

Specifically, for each node input feature  $x_i$ , linear transformations were applied to generate the Q, K, and V vectors representations, as previously defined in Eq. (5). Interatomic Euclidean distances  $d_{ij}$  were expanded using radial basis functions (RBF), yielding a rich multi-scale representation:

$$f_{ij} = \text{RBF}(d_{ij}) \quad (11)$$

The extended RBF vectors were further mapped into the key vectors and value vectors by nonlinear transformations:

$$\Delta k_{ij} = \text{MLP}_k(f_{ij}), \Delta v_{ij} = \text{MLP}_v(f_{ij}) \quad (12)$$

The attention weight  $\alpha_{ij}$  was modulated by a distance-decay function and a learnable scalar coefficient  $\alpha$ , resulting in:

$$\alpha_{ij}^{(h)} = \text{SiLU} \left( \frac{q_i^{(h)} \cdot (k_j^{(h)} + \Delta k_{ij}^{(h)})}{\sqrt{d_h}} \right) \cdot C(d_{ij}) \cdot \alpha \quad (13)$$

where  $C(d_{ij})$  is a cosine-shaped distance cutoff function used to suppress the influence of long-range interactions, and  $\alpha$  is a learnable parameter that controls the sensitivity to distance decay. The output of each attention head was computed as the weighted aggregation of values from neighboring nodes:

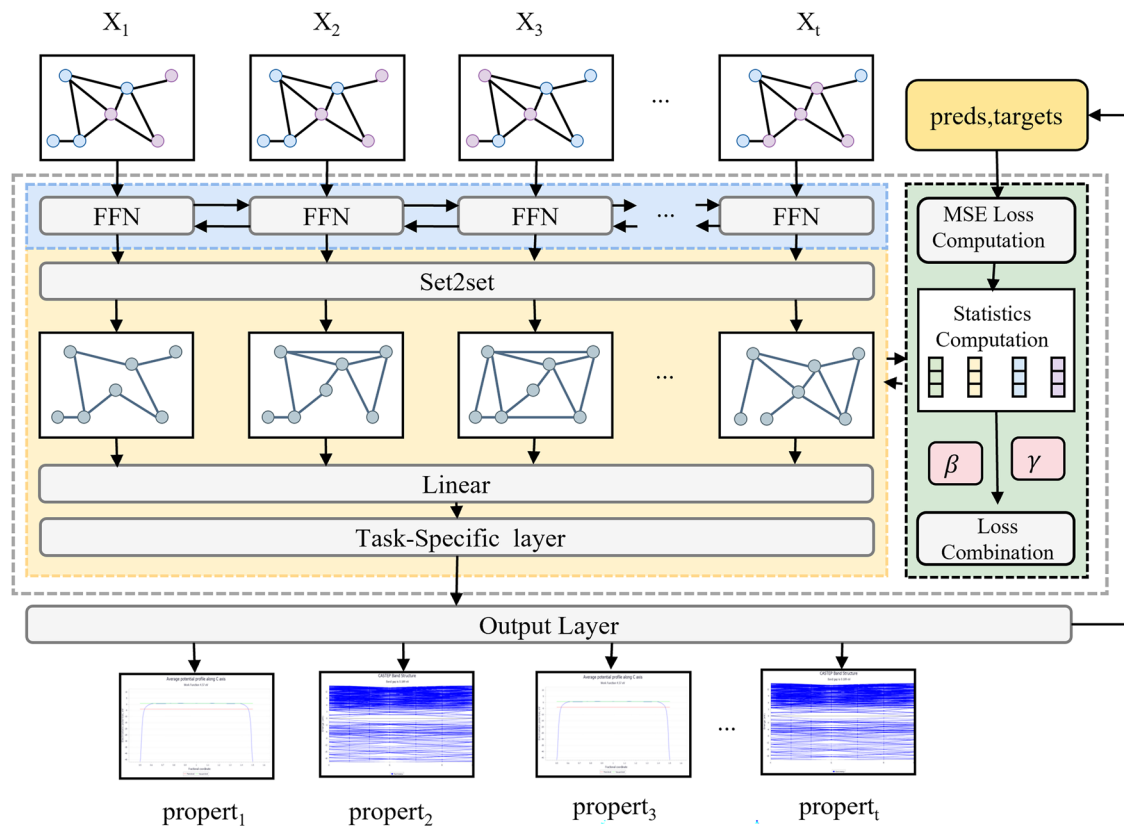
$$m_i^{(h)} = \sum_{j \in N(i)} \alpha_{ij}^{(h)} \cdot (v_j^{(h)} + \Delta v_{ij}^{(h)}) \quad (14)$$

The outputs from all attention heads were concatenated and linearly projected to obtain the final node embedding:

$$y_i = W_o \cdot \text{Concat}(m_i^{(1)}, \dots, m_i^{(n)}) \quad (15)$$

### Task-specific decoders and synchronous loss optimization

To address the inherent heterogeneity across different physicochemical property prediction tasks, AEGCNN-MTL adopts a modular decoding strategy coupled with a synchronized loss optimization mechanism. This design enables the model to effectively decouple shared and task-specific representations and to adapt to property-specific learning objectives. Importantly, even when tasks share the same input modality, they can depend on different aspects of the structure-property relationship. The band gap reflects bulk electronic structure and is sensitive to local bonding environments and the periodic potential, whereas the work function is defined as the difference between the vacuum electrostatic level and the Fermi level and is strongly influenced by surface dipoles and near-surface



**Fig. 10 | Architecture of the multi-task decoding and loss coordination module.** Each task input is first processed by a feed-forward network, shown as rectangular boxes, and then aggregated using Set2Set pooling. The pooled representations are passed through linear layers and task-specific decoders to produce predictions for each property at the output layer. On the right, the loss coordination module

combines predictions and targets. Mean squared error loss is calculated, statistical information is computed, and weighting factors are introduced, denoted as beta and gamma blocks, to balance different tasks. These components together form the coordinated loss that guides training.

charge redistribution. Consequently, their predictive signals may arise from partially distinct cues. Using task-specific decoders helps retain task-dependent information while still enabling efficient sharing of common representations.

As illustrated in Fig. 10, the diagram depicts the decoding pipeline and the synchronized loss optimization strategy adopted in the AEGCNN-MTL framework. The left section shows the task-specific prediction branches, which include a feedforward network (FFN), Set2Set pooling, property-specific projection layers, and output heads corresponding to each target property. The right section displays the loss computation module, which comprises mean squared error (MSE) loss evaluation, task-wise statistics estimation, and adaptive loss weighting based on learnable coefficients  $\beta$  and  $\gamma$ .

In the left module, task-specific decoders receive shared graph representations  $x_{global}$  as input. Each decoder comprises a stack of feedforward networks, which perform task-dependent transformation of node features:

$$x_i^{(l+1,t)} = FF_t^l(x_i^{(l,t)} + x_i^{global}) \quad (16)$$

After stacking multiple layers of feedforward networks, each task-specific output layer used Set2Set pooling to generate task-specific graph-level representations, followed by a linear layer and task-specific output head:

$$g_t = Linear\left( Set2Set\left( \left( \left\{ x_i^{(L,t)} \right\}_{i=1}^N \right) \right) \right), \hat{y}_t = Head_t(g_t) \quad (17)$$

This architecture ensures that both shared and task-specific patterns are effectively captured, thereby enhancing the adaptability and generalization of the model in multi-task learning scenarios.

To promote balanced and coordinated learning, we designed a multi-task synchronous loss function with three components:

$$L_{total} = \bar{L} + \beta \cdot D + \gamma \cdot N \quad (18)$$

where: Task-averaged loss:  $\bar{L} = \frac{1}{T} \sum_{t=1}^T L_t$  evaluates the average loss over all tasks and reflects overall performance; Deviation penalty:  $D = \frac{1}{T} \sum_{t=1}^T |L_t - \bar{L}|$  penalizes large discrepancies in learning progress across tasks and suppresses task-wise instability; Non-descent penalty:  $N = \frac{1}{T} \sum_{t=1}^T ReLU(L_t - L_t^{prev})$  discourages loss rebounds over epochs to promote stable convergence; Dynamic weights:  $\beta = softplus(\log \beta')$ ,  $\gamma = softplus(\log \gamma')$  are learnable non-negative parameters that adaptively balance the contributions of the three terms.

By simultaneously regulating task equilibrium, synchronizing gradient behavior, and leveraging dynamic adaptability, this loss function effectively mitigates inter-task interference and negative transfer during optimization, thereby improving collaborative generalization across all tasks.

**Hyperparameter Settings**

To comprehensively evaluate the effectiveness of the proposed model in material property prediction tasks, we performed systematic experiments on two distinct datasets: BDG and QM9, which vary in scale and task

**Table 7 | Hyperparameter Settings for BDG and QM9 Experiments**

Hyperparameter	BDG Value	QM9 Value
Training Set Size	2000	110,000
Validation Set Size	500	10,000
Batch Size	64	128
Learning Rate	$2 \times 10^{-4}$	$5 \times 10^{-4}$
Model Dimension	128	256
Number of Attention Heads	4	8
Number of Net Layers	4	4-12
Optimizer	AdamW	AdamW
Learning Rate Scheduler	ReduceLROnPlateau	Warmup + CosineAnnealing
Weight Decay	$1 \times 10^{-6}$	$5 \times 10^{-4}$
Total Epochs	500	1000

The table lists the detailed hyperparameter configurations used in training the proposed AEGCNN-MTL model on the BDG and QM9 datasets. Distinct settings were adopted for BDG and QM9 to account for differences in dataset scale and property distribution.

characteristics. All models were developed using the PyTorch Geometric library and trained a single NVIDIA RTX3090 GPU.

The BDG dataset, comprising 2613 samples, was employed to evaluate the model's performance under small-sample conditions, with a focus on the concurrent prediction of two pivotal electronic properties  $E_{\text{gap}}$  and WF in two-dimensional boron-doped graphene materials. Given the limited sample size, a lightweight model architecture was selected to mitigate the risk of overfitting. The specific hyperparameter configurations for the BDG dataset are delineated in Table 7.

In contrast, the QM9 dataset served as a large-scale benchmark for molecular property modeling. In this study, tasks were grouped based on Pearson correlation coefficients among properties, and the neural network depths were tailored for each task group to reflect the varying degrees of modeling complexity and representational needs. The principal hyperparameter settings for the QM9 dataset are summarized in Table 7. All listed hyperparameters were used for experiments on the BDG and QM9 datasets. "Model Dimension" refers to the feature dimension of the neural network layers. In the QM9 multi-task experiments, the number of network layers was set to 6 for task groups A and D, 12 for task group B, and 4 for task group C. All experiments used the AdamW optimizer with the specified learning rate scheduling and weight decay strategies.

### Data availability

The QM9 dataset used in this study is publicly available and can be accessed from the Quantum Machine repository ([https://figshare.com/collections/Quantum\\_chemistry\\_structures\\_and\\_properties\\_of\\_134\\_kilo\\_molecules/978904](https://figshare.com/collections/Quantum_chemistry_structures_and_properties_of_134_kilo_molecules/978904)). The boron-doped graphene (BDG) dataset generated during the current study is not publicly available due to project-specific restrictions, but can be obtained from the corresponding author upon reasonable request.

### Code availability

The code used for data processing and model training in this study is not publicly available but can be obtained from the corresponding author upon reasonable request.

Received: 11 August 2025; Accepted: 5 December 2025;

Published online: 24 December 2025

### References

- Novoselov, K. S. et al. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).

- Bhushan, R. et al. Microwave graphitic nitrogen/boron ultradoping of graphene. *npj 2D Mater. Appl.* **8**, 19 (2024).
- Chen, X., Yao, Y., Yao, H., Yang, F. & Ni, J. Topological  $p + ip$  superconductivity in doped graphene-like single-sheet materials  $\text{BC}_3$ . *Phys. Rev. B* **92**, 174503 (2015).
- Dieb, M., Hou, Z. & Tsuda, K. Structure prediction of boron-doped graphene by machine learning. *J. Chem. Phys.* **148**, 241716 (2018).
- Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 140022 (2014).
- Nandi, S., Vegge, T. & Bhowmik, A. MultiXC-QM9: Large dataset of molecular and reaction energies from multi-level quantum chemical methods. *Sci. Data* **10**, 783 (2023).
- Jiang, X. et al. Interpretable machine learning applications: A promising prospect of AI for materials. *Adv. Funct. Mater.* **35**, 2507734 (2025).
- Koepfli, S. M. et al. Controlling photothermoelectric directional photocurrents in graphene with over 400 GHz bandwidth. *Nat. Commun.* **15**, 7351 (2024).
- Chen, W. et al. Heteroatom-doped flash graphene. *ACS Nano* **16**, 6646–6656 (2022).
- Thiemann, F. L., Rowe, P. N., Zen, A., Müller, E. A. & Michaelides, A. Defect-dependent corrugation in graphene. *Nano Lett.* **21**, 7231–7237 (2021).
- Guo, M., Qian, Y., Qi, H., Bi, K. & Chen, Y. Experimental measurements on the thermal conductivity of strained monolayer graphene. *Carbon* **157**, 185–190 (2020).
- Yu, X. et al. Boron-doped graphene for electrocatalytic  $\text{N}_2$  reduction. *Joule* **2**, 1610–1622 (2018).
- Chowdhury, S., Jiang, Y., Muthukaruppan, S. & Balasubramanian, R. Effect of boron doping level on the photocatalytic activity of graphene aerogels. *Carbon* **128**, 237–248 (2018).
- Liu, J., Liang, T., Tu, R., Lai, W. & Liu, Y. Redistribution of  $\pi$  and  $\sigma$  electrons in boron-doped graphene from DFT investigation. *Appl. Surf. Sci.* **481**, 344–352 (2019).
- Belfar, A. Simulation study of the a-Si:H/nc-Si:H solar cells performance sensitivity to the TCO work function, the band gap and the thickness of i-a-Si:H absorber layer. *Sol. Energy* **114**, 408–417 (2015).
- Zhang, Q. et al. Bandgap renormalization and work function tuning in  $\text{MoSe}_2/\text{hBN}/\text{Ru}(0001)$  heterostructures. *Nat. Commun.* **7**, 13843 (2016).
- Ganose, A. M. & Scanlon, D. O. Band gap and work function tailoring of  $\text{SnO}_2$  for improved transparent conducting ability in photovoltaics. *J. Mater. Chem. C* **4**, 1467–1475 (2016).
- Formalik, F., Shi, K., Joodaki, F., Wang, X. & Snurr, R. Q. Exploring the structural, dynamic, and functional properties of metal–organic frameworks through molecular modeling. *Adv. Funct. Mater.* **34**, 2308130 (2024).
- Lejaeghere, K. et al. Reproducibility in density functional theory calculations of solids. *Science* **351**, aad3000 (2016).
- Zhang, Y. et al. Screening dual variable-valence metal oxides doped calcium-based material for calcium looping thermochemical energy storage and  $\text{CO}_2$  capture with DFT calculation. *J. Energy Chem.* **107**, 170–182 (2025).
- Fedik, N. et al. Extending machine learning beyond interatomic potentials for predicting molecular properties. *Nat. Rev. Chem.* **6**, 735–755 (2022).
- Huang, G., Huang, F. & Dong, W. Machine learning in energy storage material discovery and performance prediction. *Chem. Eng. J.* **492**, 152294 (2024).
- Wang, C. et al. A guided review of machine learning in the design and application for pore nanoarchitectonics of carbon materials. *Mater. Sci. Eng. R. Rep.* **165**, 101010 (2025).

24. Reiser, P. et al. Graph neural networks for materials science and chemistry. *Commun. Mater.* **3**, 93 (2022).
25. Fung, V., Zhang, J., Juarez, E. & Sumpter, B. G. Benchmarking graph neural networks for materials chemistry. *npj Comput. Mater.* **7**, 84 (2021).
26. Omeo, S. S. et al. Scalable deeper graph neural networks for high-performance materials property prediction. *Patterns* **3**, 100491 (2022).
27. Ciano, G., Rossi, A., Bianchini, M. & Scarselli, F. On inductive–transductive learning with graph neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 9335498 (2022).
28. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
29. Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
30. Choudhary, K. & DeCost, B. Atomistic line graph neural network for improved materials property predictions. *npj Comput. Mater.* **7**, 185 (2021).
31. Gasteiger, J., Giri, S., Margraf, J. T. & Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115*. <https://doi.org/10.48550/arXiv.2011.14115>.
32. Gasteiger, J., Becker, F. & Günnemann, S. GemNet: Universal directional graph neural networks for molecules. *Adv. Neural Inf. Process. Syst.* **34**, 13268–13280 (2021).
33. Liang, C. et al. Multi-task mixture density graph neural networks for predicting catalyst performance. *Adv. Funct. Mater.* **34**, 2404392 (2024).
34. Feng, M. et al. Finding the optimal CO<sub>2</sub> adsorption material: Prediction of multi-properties of metal–organic frameworks (MOFs) based on DeepFM. *Sep. Purif. Technol.* **302**, 122111 (2022).
35. Kong, S., Guevarra, D., Gomes, C. P. & Gregoire, J. M. Materials representation and transfer learning for multi-property prediction. *APL Mater.* **8**, 021409 (2021).
36. Misra, I., Shrivastava, A., Gupta, A. & Hebert, M. Cross-stitch networks for multi-task learning. *In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 3994–4003 (2016).
37. Veličković, P. et al. Graph Attention Networks. *ICLR*. <https://doi.org/10.48550/arXiv.1710.10903> (2018).
38. Hamilton, W., Ying, Z. & Leskovec, J. Inductive representation learning on large graphs. *Adv. Neural Inf. Process. Syst.* **30**, 1024–1034 (2017).
39. Xu, K., Hu, W., Leskovec, J. & Jegelka, S. How powerful are graph neural networks? *ICLR*. <https://doi.org/10.48550/arXiv.1810.00826>.
40. Chen, M., Wei, Z., Huang, Z., Ding, B. & Li, Y. Simple and Deep Graph Convolutional Networks (GCNII). *ICML*. <https://doi.org/10.48550/arXiv.2007.02133>.
41. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
42. Thölke, P. & De Fabritiis, G. TorchMD-NET: Equivariant transformers for neural-network-based molecular potentials. *arXiv 2202.02541*. <https://doi.org/10.48550/arXiv.2202.02541>.
43. Unke, O. T. & Meuwly, M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).
44. Gasteiger, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. *Proc. 8th Int. Conf. Learn. Represent. (ICLR)*. <https://doi.org/10.48550/arXiv.2003.03123>.
45. Thölke, P. & Fabritiis, G. D. TorchMD-NET: Equivariant transformers for neural network-based molecular potentials. *Mach. Learn. Sci. Technol.* **3**, 015022 (2022).
46. Kwak, B. et al. GeoT: a geometry-aware transformer for reliable molecular property prediction and chemically interpretable representation learning. *ACS Omega* **8**, 39759–39769 (2023).
47. Rani, P. & Jindal, V. K. Designing band gap of graphene by B and N dopant atoms. *RSC Adv.* **3**, 802–812 (2013).
48. Sheshanarayana, R. & You, F. Knowledge distillation for molecular property prediction: A scalability analysis. *Adv. Sci.* <https://doi.org/10.1002/adv.202503271> (2018).
49. Chang, J. & Zhu, S. MGNN: Moment graph neural network for universal molecular potentials. *npj Comput. Mater.* **11**, 55 (2025).
50. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst.* <https://doi.org/10.48550/arXiv.1706.03762>.

## Acknowledgements

This research is supported by the key project of science and technology research program of Chongqing Education Commission of China (KJZD-K202501109), the National Natural Science Foundation of China (U22A20434), Scientific research foundation of Ministry of Industry and Information Technology of the People's Republic of China (TC220A04A-43).

## Author contributions

Y.L. contributed to data curation, preprocessing, and feature engineering. M.C. performed model implementation, experiments, and result analysis. Q.Z. conceived and supervised the study, designed the methodology, and was responsible for project administration and funding acquisition. J.Z. assisted in algorithm development and experimental validation. C.Z. contributed to data interpretation and visualization. S.X. provided guidance on material science applications and contributed to the theoretical framework. Q.B. assisted with manuscript editing and literature review. All authors discussed the results, contributed to the writing of the manuscript, and approved the final version for submission.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Qingwei Zhang.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025