Global Environmental Science
www.sciltp.com/journals/ges

Article

Multi-Task Graph Convolutional Network Model with Improved Performance and Broad Applicability Domains for Identifying Aquatic Toxic Chemicals

Shengshe Ji, Wenjia Liu, Guijia Bi, Haobo Wang * and Jingwen Chen *

Key Laboratory of Industrial Ecology and Environmental Engineering (Ministry of Education), Dalian Key Laboratory on Chemicals Risk Control and Pollution Prevention Technology, School of Environmental Science and Technology, Dalian University of Technology, Dalian 116024, China
* Correspondence: wanghaobodut@mail.dlut.edu.cn (H.W.); jwchen@dlut.edu.cn (J.C.)

How To Cite: Ji, S.; Liu, W.; Bi, G.; et al. Multi-Task Graph Convolutional Network Model with Improved Performance and Broad Applicability Domains for Identifying Aquatic Toxic Chemicals. *Glob. Environ. Sci.* **2026**, *2*(3), 278–294. https://doi.org/10.53941/ges.2026.100019

Publication History

Received: 25 February 2026
Revised: 13 April 2026
Accepted: 23 April 2026
Published: 22 July 2026

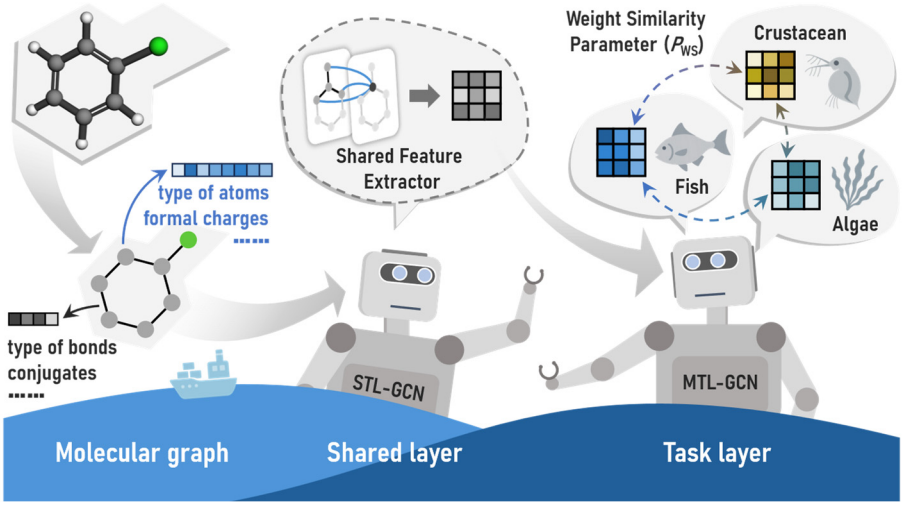
Keywords

aquatic toxicity;
multi-task learning;
graph convolutional network;
attention mechanism;
applicability domain

Abstract: High-throughput computational models are essential for regulatory screening of large chemical inventories for aquatic toxicity, while their performance is frequently constrained by limited datasets, inadequate exploitation of inter-endpoint correlations, and lack of applicability domain (AD) characterization. This study curated a regulation-oriented aquatic toxicity dataset comprising 17,514 chemicals across six ecotoxicological endpoints. An end-to-end multi-task learning graph convolutional network (MTL-GCN) was constructed to enable joint modeling of multiple toxicity endpoints. Relative to single-task learning (STL-GCN), the MTL-GCN improved the area under the receiver operating characteristic curve by 0.4~13.4% across the six endpoints. A weight similarity parameter (P_{WS}) was calculated, indicating that MTL effectively captures intrinsic correlations among endpoints. The integrated attention mechanism enhanced interpretability by identifying aquatic toxicophores. AD characterization method based on the structure-activity landscape (AD_{SAL}) analysis confirmed that MTL-GCN improves in-domain screening accuracy while maintaining comparable coverage to STL-GCN. When applied to 783,681 inventory chemicals, the combined MTL-GCN and AD_{SAL} framework achieved reliable screening with 94.7% of chemicals located in the AD of at least one endpoint. The constructed dataset and the MTL-GCN framework establish benchmarks for aquatic toxicity prediction and provide a robust method for large-scale ecological hazard assessment.

Highlights

- Aquatic toxicity datasets covering six endpoints expanded chemical coverage by 66.0% on average.
- MTL-GCN achieved high prediction accuracy and broad applicability domains.
- MTL-GCN with AD_{SAL} enabled reliable screening with 94.7% coverage.



1. Introduction

While promoting economic development and improving human life quality, chemicals are inevitably released into the environment and lead to environmental contamination. In particular, a substantial proportion of water pollution incidents are caused by chemical pollution [1–4]. Chemicals in water can lead to a series of adverse effects on human and ecological health [3–5]. Consequently, assessing aquatic toxicity of chemicals is essential for hazard identification and regulatory control. Major regulatory frameworks, such as the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) regulation [6], the US Toxic Substances Control Act (TSCA) [7], and China's Measures for the Environmental Management Registration of New Chemical Substances [8], mandate rigorous aquatic acute and chronic toxicity testing for chemicals. Consistent with the Globally Harmonized System of Classification and Labelling of Chemicals (GHS) [9], these regulations require the evaluation of acute and chronic toxicity using fish, crustaceans, and algae as representative organisms. Assessing these six endpoints reflects chemical toxicity across different aquatic trophic levels, ensuring both regulatory compliance and a systematic evaluation of aquatic hazards.

Traditionally, aquatic toxicity has been evaluated through *in vivo* bioassays employing organisms such as fish, crustaceans, and algae [10,11]. However, given the rapidly increasing number of chemicals, experimental methods are insufficient to meet the demands of sound chemical management [12,13]. Therefore, there is an urgent need to develop efficient and accurate screening models to identify aquatic toxic chemicals [13]. In recent years, machine learning (ML) has emerged as a powerful tool for diverse scientific and engineering domains, including materials science [14–17], climate science [18,19], and ecological research [20,21]. ML models have been widely employed to predict the bioactivities of chemicals, including their acute or chronic aquatic toxicity toward representative organisms such as fish, crustaceans, and algae [22–36].

Despite the advances, several methodological limitations remain. First, most existing ML models rely on predefined molecular descriptors, which may not adequately capture the complex structural features governing toxicological responses [37–40]. In addition, modeling approaches based on molecular descriptors require extensive feature engineering, which is time-consuming and endpoint-specific [22,27,31,40]. Furthermore, data scarcity for certain endpoints, such as fish chronic toxicity, hinders the development of robust models using classical ML methods [41,42]. Under the constraints, prediction models relying on handcrafted features and classical algorithms may struggle to learn the highly nonlinear structure-activity relationships underlying chemical toxicity [38,40].

Graph neural networks (GNNs), as an emerging deep learning (DL) algorithm, have been successfully employed in chemical toxicity screening, with graph convolutional networks (GCNs) being a representative example [25,43,44]. Distinct from classical ML methods, GCN operates directly on molecular graphs (MGs), representing atomic topological connectivity and neighborhood relationships. The GCN model takes MGs as inputs and learns the mapping between structures and toxicity endpoints, thereby achieving end-to-end learning. Consequently, GNNs automatically extract informative molecular representations, eliminating the need for manual feature engineering.

On the other hand, most existing aquatic toxicity screening models were constructed under a single-task learning (STL) paradigm, focusing on a single endpoint or species, such as fish acute toxicity [22,25,27,28,34]. Such a learning paradigm neglects the intrinsic correlations among toxicological responses across different species and endpoints. Multi-task learning (MTL) provides a principled modeling framework in which a unified model is constructed and optimized across multiple related tasks, allowing shared feature representations to be learned and potentially improving the robustness of models trained on limited data [45–47]. Accordingly, it is essential to construct a GCN model integrated with an MTL strategy (MTL-GCN) to screen the aquatic acute and chronic toxicity of chemicals across multiple species, including fish, crustaceans, and algae, thereby effectively supporting environmental chemical hazard assessment and serving regulatory needs.

While the efficacy of MTL in improving prediction accuracy, robustness, and generalization has been reported [26,45,48], systematic investigations into the underlying mechanisms responsible for improvements remain limited. To elucidate the underlying reasons for the effectiveness of MTL, a weight similarity parameter (P_{WS}) was designed to quantify the similarity of neuron weights within neural network layers. Furthermore, P_{WS} was employed to investigate the association between MTL modeling performance and key factors reflecting task-relatedness, including the number of shared molecules and their label consistency rate (R_{LC}).

Despite the high performance, DL models such as GCNs are often regarded as "black boxes" due to limited interpretability, making it difficult to understand the rationale underlying model predictions [49–53]. Models intended for regulatory purposes may require mandatory interpretability steps [54]. Attention mechanisms provide a promising avenue for enhancing the interpretability of DL-based toxicity models by enabling the identification of molecular substructures most relevant to toxicity screening [55–57]. Beyond interpretability, applicability domain (AD) characterization represents another critical approach for ensuring the reliability of model predictions, yet it has received limited attention in aquatic toxicity screening studies [22,30,34,58]. The absence of defined

ADs would reduce the reliability of model predictions, especially when models are used for chemical regulatory purposes [59–62].

Recently, an AD characterization method based on a structure-activity landscape (SAL) analysis, AD_{SAL}, has demonstrated superior performance over other AD methods in delineating the reliable prediction space of models for screening toxic chemicals [56,63–65]. It is expected that the AD_{SAL} method may also be effective in enhancing the reliability of aquatic toxicity prediction models within their defined ADs. Given that MTL enables the learning of shared feature representations across related endpoints, it holds the potential to expand AD coverage while maintaining high performance. Although the capability has not been previously confirmed, it is crucial to investigate the impact of MTL on AD coverage.

In this study, consistent with regulatory requirements, acute toxicity datasets for fish, crustaceans, and algae, as well as their corresponding chronic toxicity datasets, were curated, and an MTL-GCN screening model was constructed accordingly. The MTL-GCN model demonstrated performance superior to that reported in previous studies. Additionally, the MTL-GCN model coupled with AD_{SAL} exhibited expanded AD coverage compared with STL models while maintaining comparable prediction performance. The optimal model was subsequently employed to identify aquatic toxic chemicals from inventories comprising approximately 106 chemicals. Given its state-of-the-art performance and rigorous AD characterization, the proposed framework offers a robust tool to support chemical registration and regulatory decision-making, thereby facilitating sound chemical management.

2. Materials and Methods

2.1. Data Sets Collection and Data Preprocessing

According to relevant regulations [6–9] and test guidelines [66–70], aquatic toxicity data were collected for six endpoints, including (I) fish acute toxicity characterized by lethality measured as the median lethal concentration (LC_{50}, mg/L), (II) fish chronic toxicity involving the inhibition of early life-stage growth and development measured as the no-observed-effect concentration (*NOEC*, mg/L), (III) crustacean acute toxicity defined by immobilization using the median effective concentration (EC_{50}, mg/L), (IV) crustacean chronic toxicity focusing on reproduction (*NOEC*, mg/L), (V) algal acute toxicity characterized by the inhibition of average growth rate (ErC_{50}, mg/L), and (VI) algal chronic toxicity involving the inhibition of average growth rate (*NOEC*, mg/L).

To construct a comprehensive dataset, this study compiled data from peer-reviewed literature [24–32,36], publicly accessible databases, including ECOTOX [71], eChemPortal [72], EnviroTox [73], TOXRIC [74], as well as an aquatic toxicity list provided by the Japanese Ministry of the Environment (MoE) [75]. The integrated dataset comprises simplified molecular input line entry system (SMILES) codes, corresponding chemical toxicity labels, test species, and exposure durations. Summary information on the raw data is presented in Table 1.

Table 1. Summary of endpoints and integrated datasets.

Endpoint	Number of Raw Data	Number of Species	Source
Fish acute toxicity	93,643	623	[24,25,27–29,71–75]
Fish chronic toxicity	2261	72	[71–73,75]
Crustacean acute toxicity	13,086	112	[26,31,72,75]
Crustacean chronic toxicity	2163	11	[32,71–73,75]
Algal acute toxicity	9371	80	[29,33,71,72,75]
Algal chronic toxicity	5398	63	[71–73,75]

All records were systematically cleaned, standardized, and curated through a series of data processing steps prior to model development to ensure compliance with established regulations [6–9] and test guidelines [66–70]. Records with exposure durations inconsistent with relevant test guidelines [66–70] were excluded. For instance, fish acute toxicity data were restricted to 96-h LC_{50} measurements. In addition, according to GHS regulations [9], organisms were categorized into three groups, including fish, crustaceans, and algae. For each endpoint, only test species that satisfied the corresponding regulatory guidelines were retained. Detailed descriptions of the curation procedures and endpoint-specific exposure duration criteria are provided in the Supporting Information (SI).

Following the processing steps, continuous toxicity measurements were discretized into binary toxic/non-toxic labels in accordance with the GHS [9] to support the construction of high-throughput screening models. According to GHS [9], for the acute toxicity endpoints, compounds were classified as toxic when fish 96 h LC_{50}, crustacean 48 h EC_{50}, or algal 72 h ErC_{50} values were ≤100 mg/L, and as non-toxic when these values were >100 mg/L. For the chronic toxicity endpoints, compounds were classified as toxic when *NOEC* values for fish, crustaceans, or algae were ≤1 mg/L, and as non-toxic when these values were >1 mg/L.

Chemical structures were standardized through SMILES processing, including desalting, desolvation, inorganic molecule filtering, charge neutralization, and canonicalization [45,55,76–78]. Salts and solvents were removed because they may interfere with descriptor calculation and structural encoding [45,76–78]. Inorganic molecules that could not be converted into molecular graphs were excluded [45,55,78]. Charge neutralization

 https://doi.org/10.53941/ges.2026.100019

and canonical SMILES generation were subsequently applied to reduce inconsistencies arising from alternative ionic forms or equivalent structural representations and standardize structural representations for downstream modeling [45,76–78].

Furthermore, for compounds with multiple experimental records from disparate sources, all records were discretized individually and then integrated. Consistent labels were assigned directly. For inconsistent labels, a conservative consensus approach was applied, under which a compound was categorized as toxic if any record indicated toxicity and as non-toxic only if all records indicated non-toxicity. The data processing steps were carried out using the open-source package Data Integration and Processing for Computational Toxicology, namely DiPTox (https://pypi.org/project/diptox/, accessed on 30 November 2025).

2.2. Molecular Graphs (MGs) and MTL-GCN Architecture

The proposed MTL-GCN architecture comprises four sequential modules, including an MG encoder, a feature extractor, task-specific feature generators, and predictors. To facilitate effective MTL, the feature generator employs an attention mechanism to transform shared general features into task-specific toxicity features by assigning distinct weights to molecular substructures, thereby facilitating simultaneous prediction for six aquatic toxicity endpoints. In addition, the attention mechanism provides a basis for model interpretability by highlighting the molecular substructures that contribute most to each toxicity prediction.

Within the framework, each molecule is first represented as a graph $G = (h, e)$ (Figure 1A), where h and e denote the node and edge representations encoded from atomic and bond attributes, respectively. A detailed description of the features is provided in the SI. MG construction and feature encoding procedures were implemented using the Deep Graph Library package [79]. The proposed algorithm was adapted from a general architecture [80]. And the corresponding configuration of architecture are summarized in Table S10.

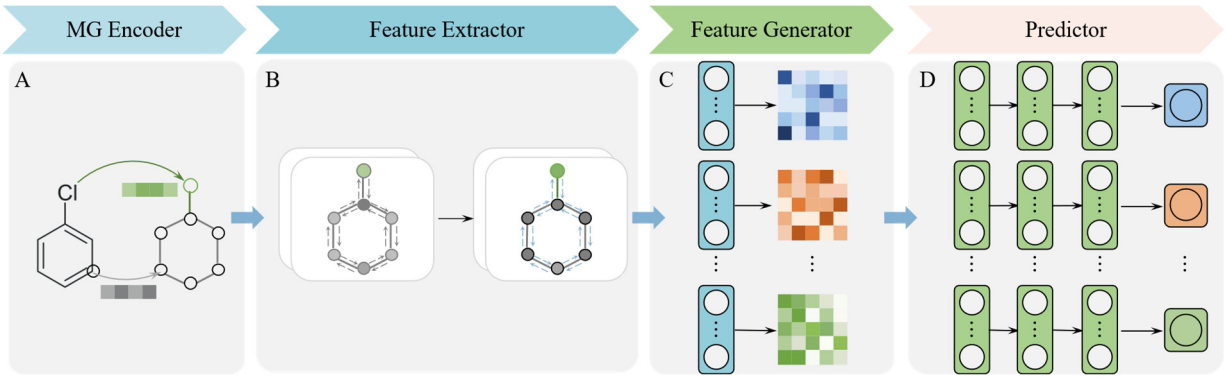


Figure 1. Workflow of MTL-GCN Architecture, which encompasses four sequential modules: (**A**) MG encoder, (**B**) feature extractor, (**C**) feature generator, and (**D**) predictor.

Subsequently, the MGs serve as inputs to feature extractor (Figure 1B). The shared feature extractor consisted of an optimized number of GCN layers, utilizing ReLU as the activation function. This module functions as a shared representation learning module across the six aquatic toxicity tasks, thereby enabling effective knowledge transfer among endpoints. Specifically, the feature extractor learns local atomic representations through iterative neighborhood aggregation, in which information from adjacent atoms and their associated bonds is progressively integrated. The representation of each atom is updated by aggregating features from its local neighborhood according to the following update rule:

$$m_v^{t+1} = \sum_{w \in N_v} M_t(h_v^t, h_w^t, e_{v,w}^t) \quad (1)$$

where m_v^{t+1} denotes the aggregated message vector for target node v at iteration $t + 1$; h_v^t denotes the feature representation of target node v after t iterations; h_w^t denotes the feature representation of neighboring node w after t iterations; $e_{v,w}^t$ represents the features of the edge connecting node pairs v and w after t iterations. Additionally, N_v represents the set of neighboring nodes connected to node v; and $M_t(\cdot)$ represents a function that passes the neighbor features to the target node.

Subsequently, the features of the center nodes v were updated at iteration $t + 1$ through the function $U_t(\cdot)$, incorporating both molecular structural characteristics and local chemical information. The resulting atomic representation encoded not only local node features but also the structural context of the atom-centered neighborhoods.

$$h_v^{t+1} = U_t(h_v^t, m_v^{t+1}) \quad (2)$$

To capture task-specific characteristics, distinct task-specific attention layers were introduced for the six aquatic toxicity tasks, serving as task-specific feature generators (Figure 1C). The layers assign importance weights to individual atoms, thereby transforming shared molecular representations into task-aware features that

emphasize substructures most relevant to each endpoint while supporting model interpretability. The attention weights ($\omega_{v,k}$) are computed as follows:

$$\omega_{v,k} = sigmoid(W_k \cdot h_v + b_k) \quad (3)$$

where $\omega_{v,k}$ denotes the attention weight of node v for a specific toxicity task k, quantifying the contribution of the local substructure centered at atom v to that task; h_v represents the final-layer latent representation of node v obtained from the feature extractor; the $sigmoid(\cdot)$ activation function constrains the $\omega_{v,k}$ of each node between 0 and 1; W_k and b_k denote the trainable matrix and the bias vector within the feature generator module for task k. This attention formulation was designed to simultaneously support task-specific feature generation for MTL and enhance model interpretability. In addition, the number of attention layers was varied during hyperparameter tuning to evaluate the effect of alternative attention configurations on model performance. Subsequently, these features are aggregated to generate a task-specific feature vector (F_k):

$$F_k = \sum_{v=1}^{N} (\omega_{v,k} \cdot h_v) \quad (4)$$

Following the attention layer, the resulting task-adapted molecular representations are propagated through a predictor to yield final toxicity predictions (Figure 1D). The predictor consisted of FC layers selected through parameter tuning. ReLU was used as the activation function. Collectively, the GCN layers were designed to extract shared features directly from MGs. The task-specific attention mechanism facilitates MTL, thereby enhancing model interpretability. Finally, the FC layers serve as predictors. To further elucidate how MTL-GCN captures task-general versus task-specific knowledge, a weight similarity parameter (P_{WS}) was introduced:

$$P_{WS}(k_1, k_2) = \frac{W_{k_1} \cdot W_{k_2}}{\|W_{k_1}\| \|W_{k_2}\|} \quad (5)$$

where P_{WS} (k_1, k_2) denotes the similarity between neuron weight matrices of task k_1 (W_{k1}) and task k_2 (W_{k2}); $||W||$ represents the Euclidean norm of the flattened weight vectors. P_{WS} was calculated for the task-specific feature generators to quantitatively characterize the extent of knowledge sharing among toxicity endpoints.

For comparison, STL-GCN models were constructed for each of the six toxicity endpoints, as detailed in the SI. In addition, STL models were also developed based on three classical ML algorithms, including Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), combined with Avalon, PubChem, Molecular ACCess System (MACCS) keys and Extended-Connectivity fingerprints. In total, 13 STL-ML models were developed for each endpoint. The optimal model for each endpoint was selected as the baseline for comparison with the MTL-GCN model.

2.3. Model Training and Evaluation

For each endpoint, the datasets were randomly split into training, test, and validation sets at a ratio of 8:1:1. Specifically, the training, test, and validation sets were used for model fitting, hyperparameter tuning, and final evaluation, respectively. To ensure statistical robustness, ten independent runs were conducted with different random seeds for data splitting and model initialization. Model performance was evaluated on the validation sets, and the reported results are presented as the mean ± standard deviation across the independent runs. For the classification tasks, model performance was quantified using the area under the receiver operating characteristic curve (A_{ROC}) and balanced accuracy (R_{BA}).

Model parameters were optimized using Adam. The optimal hyperparameter configuration of MTL-GCN was selected by grid search according to model performance on the test set, including the numbers of GCN layers, the hidden units of GCN layers, the numbers of attention layers, the hidden units of attention layers, the numbers of FC layers, and the hidden units of FC layers, dropout settings, learning rate, weight decay, and batch size. The parameter configurations for all models were provided in Tables S10–S13. An additional scaffold-based splitting strategy was adopted for model evaluation, and the corresponding results were provided in Table S8.

2.4. AD Characterization

An AD characterization method $AD_{SAL}\{\rho_{s,q} \geq \rho_{s,T}, I_{A,q} \leq I_{A,T}\}$ [64] was adopted, where the SAL represents variations in activity with molecular similarity distributions of compounds; $\rho_{s,q}$ and $I_{A,q}$ stand for weighted molecular similarity density of a query compound represented by the subscript q in chemical space of the training set, and weighted inconsistency in molecular activities, respectively; $\rho_{s,T}$ and $I_{A,T}$ are the corresponding thresholds. Therefore, a higher $\rho_{s,q}$ indicates that the query compound shares greater structural similarity with the training compounds, resulting in a relative reliable prediction. In parallel, a higher $I_{A,q}$ indicates that the query compound is more likely located within activity cliffs. These activity cliffs are formed by molecules with similar structures but conflicting activity, corresponding to larger prediction errors and lower model reliability. The AD_{SAL} method has been proved to outperform the ADs in previous studies [64,76,81]. $\rho_{s,q}$ and $I_{A,q}$ were calculated as follows:

$$\rho_{s,q} = \sum_{t \in T} w_{q,t} \quad (6)$$

where the subscripts q and t represent a query compound and a training compound, respectively; T represents the entire training set; $w_{q,t}$ stands for a weight function for the pair of compounds q and t. The weight function $w_{q,t}$, which is associated with a parameter (α) for modulating

 https://doi.org/10.53941/ges.2026.100019

weighted contributions, was assessed for calculating ρ_s. I_A was calculated as follows:

$$I_{A,q} = \frac{\sum_{t \in T} w_{q,t} \cdot S_{WD,t}}{\sum_{t \in T} w_{q,t}} \quad (7)$$

where $S_{WD,t}$ stands for weighted local discontinuity scores (S_{WD}) for the training compound (t). Various $\rho_{s,T}$ and $I_{A,T}$ values were defined to evaluate their effects on the number of compounds (n) kept within the AD and on the model performance. More details on the AD_{SAL} calculation are provided in the SI.

On this basis, AD_{SAL} was employed to characterize the AD of the proposed MTL-GCN model. For consistency, the same AD characterization was conducted for the STL-GCN models. A comparative analysis of the AD coverage and prediction reliability was performed to evaluate whether the MTL framework enhances AD coverage relative to STL models. In addition, AD_{SAL} was further compared with a leverage-based AD method for the MTL-GCN and the corresponding results are presented in Table S18.

3. Results and Discussion

3.1. Aquatic Toxicity Data Sets

The total number of raw toxicity records initially collected is provided in Figure 2A,B, which present the number of unique chemicals retained after DiPTox curation, together with corresponding statistics from representative prior studies [24–32,36] and public databases [71–75]. The results demonstrate that the endpoint-specific aquatic toxicity datasets constructed in this study exhibit substantially broader coverage than previously reported datasets [24–32,36]. Relative to the largest datasets currently available [24–32,36,71–75], the number of unique chemicals increased by an average of 66.0% across all endpoints, with the most pronounced expansion observed for fish chronic toxicity (140.9% increase). The results indicate a substantial expansion of chemical space coverage within the modeling datasets following standardized data processing.

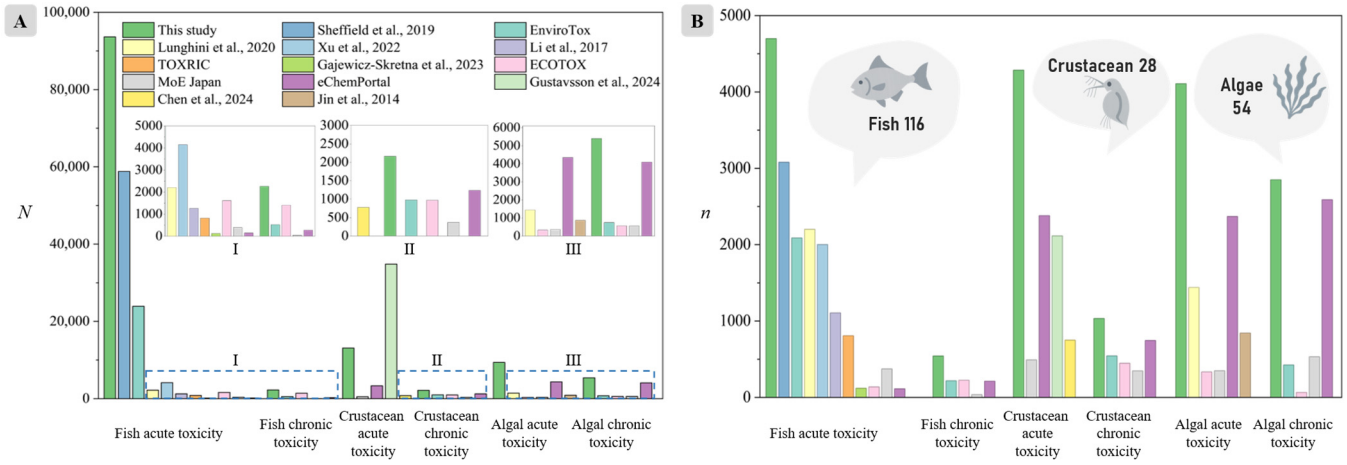


Figure 2. Construction and comparative analysis of the aquatic toxicity datasets: (**A**) comparison of the number of raw toxicity records (N) collected in this study with those reported in prior studies [25–29,31,33,82] or databases [70,71,73–75] for corresponding endpoints; (**B**) the number of unique chemicals (n) retained after DiPTox processing, together with the number of species covered.

In addition to the increase in unique chemicals, the constructed datasets demonstrate broad species coverage, encompassing 116 fish species, 28 crustacean species and 54 algal species. Overall, the constructed datasets contain larger data counts compared to existing datasets [24–32,36,71–75] and cover a wide number of species, supporting the construction of robust screening models for regulatory purposes. The datasets are provided in Tables S1–S6. The specific counts of toxic and nontoxic compounds for each endpoint are summarized in the Supplementary Materials.

To further evaluate the representativeness of the constructed datasets, their chemical communities were systematically compared with those of reference datasets, which were derived from prior studies [26,27] and public databases [71,72] with the largest number of chemicals. Here, chemical communities refer to structurally coherent clusters of compounds, and the computational procedures and visualization details are described in the Supplementary Materials.

The constructed datasets consistently encompass a larger number of distinct chemical communities than the reference datasets [26,27,71,72] across all endpoints (Figure 3). Notably, fish chronic toxicity has historically been characterized by relatively limited data availability. In this study, the constructed fish chronic toxicity dataset comprises 36 distinct communities, compared with 30 identified in the reference datasets, representing a 20.0% increase. Overall, the constructed datasets demonstrate a broad and diverse coverage of chemical space.

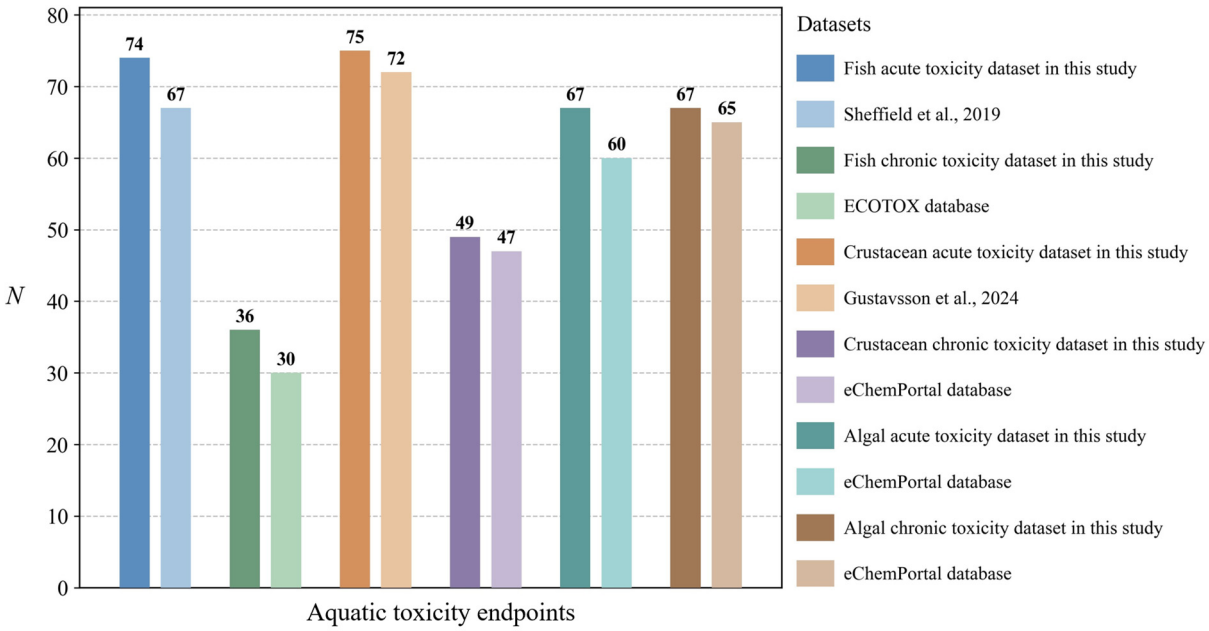


Figure 3. Comparison of chemical space coverage for six aquatic toxicity endpoints. The number of chemical communities (*N*) identified in this study is compared with those reported in representative datasets from prior studies and public databases [26,27,71,72].

Moreover, the chemical communities represented in the reference datasets [26,27,71,72] are entirely embedded within the chemical space of the constructed datasets, while additional communities are uniquely captured in the present compilation (Figure S1). The results suggest that the constructed datasets not only retain previously characterized chemical domains but also extend coverage to structurally underrepresented regions. To quantify chemical diversity, pairwise Tanimoto similarity distributions were calculated for the constructed datasets, and the results are shown in Figure S2. Across all endpoints, the similarity values fell within a low range, that cases with potential structural redundancy were relatively limited. The specific computational details for calculating these similarity metrics are comprehensively provided in the Supplementary Materials.

The distributions of molecular descriptors relevant to aquatic toxicity [82–84], including octanol/water partition coefficient (log*K*ow), molecular weight, atomic polarizability (APol), and topological polar surface area, are shown in Figure S3. The relatively close mean values and the overlapping distribution regions indicate that the descriptor distributions of the datasets for the six endpoints were generally concentrated and consistent. This indicates that the datasets for different endpoints showed relatively similar distributions of these physicochemical descriptors, which may facilitate knowledge sharing in the MTL framework. The detailed calculation methods for distributions of molecular descriptors are provided in the Supplementary Materials.

By incorporating additional and previously underrepresented structural domains, the constructed datasets enhance chemical space representativeness and provide a robust data foundation for screening model development and validation. Consequently, the datasets constitute a comprehensive dataset for regulatory-grade modeling to support ecological hazard assessment.

3.2. Performance of the MTL-GCN Model

Performance metrics across the validation sets for the six aquatic toxicity tasks are summarized in Table 2. Additional evaluation metrics, including the area under the precision-recall (PR) curve (A_{PR}), precision (R_P), recall (R_{TP}), and Matthews correlation coefficient (R_{MCC}), are provided in Table S7. In addition, a scaffold-based splitting strategy was also examined, and the corresponding results are provided in Table S8. In terms of A_{ROC}, scaffold-based splitting led to consistently lower MTL-GCN performance than random splitting across all six endpoints, with decreases ranging from 0.7∼9.6%. Notably, random splitting is more widely adopted in previous aquatic toxicity modeling studies [25,27,32,80,85]. Moreover, the use of the random splitting protocol enables a direct and consistent comparison of model performance with previously reported models [25,32,85], which were also evaluated under random splitting protocol. Therefore, all subsequent analyses and applications in this study were conducted based on the random splitting protocol.

Overall, the proposed MTL-GCN achieved consistently superior performance relative to the corresponding STL models under the present evaluation framework, and previous best models (Tables 2 and S7) [25,32,85]. In terms of A_{ROC}, MTL-GCN achieved improvements over the STL-GCN across multiple endpoints, with the maximum

 https://doi.org/10.53941/ges.2026.100019

gain reaching 13.4% and an average gain of 4.4%. To mitigate the impact of randomness arising from dataset splitting, paired *t* tests based on ten independent runs were conducted. The results confirmed that the observed performance improvements were statistically significant across all six endpoints (Table S9, $p < 0.001$). The results indicate that joint optimization across related toxicity endpoints enables effective knowledge sharing, thereby enhancing prediction performance compared with independently trained STL models.

Table 2. Classification performance comparison among MTL-GCN, STL models, and previous models on validation sets.

Endpoints	The Constructed Models							Previous Best Models			
	MTL-GCN			STL-GCN		Classical ML					
	N	$A_{ROC\text{-}val}$	$R_{BA\text{-}val}$	$A_{ROC\text{-}val}$	$R_{BA\text{-}val}$	$A_{ROC\text{-}val}$	$R_{BA\text{-}val}$	*N*	$A_{ROC\text{-}val}$	$R_{BA\text{-}val}$	Ref.
Fish acute toxicity	4697	0.860 ± 0.006	0.790 ± 0.014	0.802 ± 0.010	0.738 ± 0.009	0.776 ± 0.006	0.552 ± 0.002	1500	0.856	0.781	[25]
Fish chronic toxicity	542	0.806 ± 0.011	0.720 ± 0.029	0.711 ± 0.033	0.655 ± 0.061	0.771 ± 0.009	0.722 ± 0.004	/	/	/	/
Crustacean acute toxicity	4285	0.809 ± 0.028	0.722 ± 0.029	0.806 ± 0.028	0.767 ± 0.020	0.792 ± 0.003	0.515 ± 0.007	660	0.800	/	[85]
Crustacean chronic toxicity	1033	0.846 ± 0.039	0.753 ± 0.039	0.838 ± 0.039	0.784 ± 0.027	0.802 ± 0.003	0.688 ± 0.009	404	0.830	0.837	[32]
Algal acute toxicity	4107	0.815 ± 0.020	0.742 ± 0.022	0.807 ± 0.020	0.738 ± 0.028	0.774 ± 0.012	0.522 ± 0.007	555	0.810	0.757	[32]
Algal chronic toxicity	2850	0.778 ± 0.016	0.715 ± 0.021	0.753 ± 0.016	0.700 ± 0.023	0.751 ± 0.002	0.643 ± 0.007	/	/	/	/

N: number of total compounds of the endpoint for specific source; $A_{ROC\text{-}val}$: area under the receiver operating characteristic curve on validation set; $R_{BA\text{-}val}$: balanced accuracy on validation set; Ref.: references.

Model stability was further evaluated using the standard deviation of A_{ROC} across repeated runs (Table 2). MTL-GCN exhibited a comparable maximum standard deviation (0.039) and a lower average standard deviation (0.020) than STL-GCN (maximum: 0.039; average: 0.024), suggesting improved robustness to training variability. Collectively, the results demonstrate the effectiveness of the MTL framework in enhancing model accuracy, robustness, and generalization.

Across all endpoints, GCN-based models employing MGs as input representations outperformed the constructed classical ML models based on descriptors. For several endpoints, the performance gains were substantial. For instance, in predicting fish acute toxicity, the MTL-GCN achieved an A_{ROC} of 0.860, compared with 0.802 for the STL-GCN and 0.776 for the classical ML model. The results highlight the advantage of MTL-GCN architectures in directly encoding molecular topology and local chemical features, enabling more effective extraction of task-relevant structural features than descriptor-based approaches that rely on predefined feature engineering.

Notably, fish chronic toxicity is characterized by limited data availability, which poses challenges for model development. To the best of our knowledge, no prior classification models have been reported for fish and algal chronic endpoints, leaving a methodological gap. Under the STL framework in this study, prediction performance for fish chronic and algal chronic toxicity remained moderate, with A_{ROC} values of 0.771 and 0.753, respectively. Significantly, the MTL-GCN achieved markedly improved performance and outperformed all corresponding STL models on the data-scarce endpoints, increasing the A_{ROC} to 0.806 (4.5% gain) for fish chronic toxicity and to 0.778 (3.3% gain) for algal chronic toxicity.

The improved performance relative to STL models indicates that the MTL framework effectively transfers shared chemical and biological knowledge from data-abundant and mechanistically related endpoints, such as acute toxicity tasks. By jointly learning across multiple aquatic toxicity endpoints, the MTL model constructs more informative and stable representations, thereby compensating for data sparsity in individual chronic toxicity tasks. The constructed MTL-GCN provides a practical tool that can directly support regulatory screening and chemical registration processes, where chronic toxicity data are often required but experimentally expensive and time-consuming to obtain.

To further investigate the mechanistic basis underlying the performance gains of MTL, the weight similarity parameter (P_{WS}) was calculated to quantify inter-task similarity in the learned feature generator. As shown in Figure 4A, relatively high P_{WS} values were observed between specific endpoint pairs. For instance, algal acute toxicity and crustacean acute toxicity exhibited the highest P_{WS} of 0.385, coinciding with notable performance improvements of MTL-GCN over STL-GCN for these tasks (Table 2). As illustrated in Figure 4B, the two endpoints share 2416 compounds, indicating substantial overlap in chemical space. The concordance between high structural overlap and elevated P_{WS} supports the interpretation that MTL enhances learning by exploiting shared molecular information across related tasks.

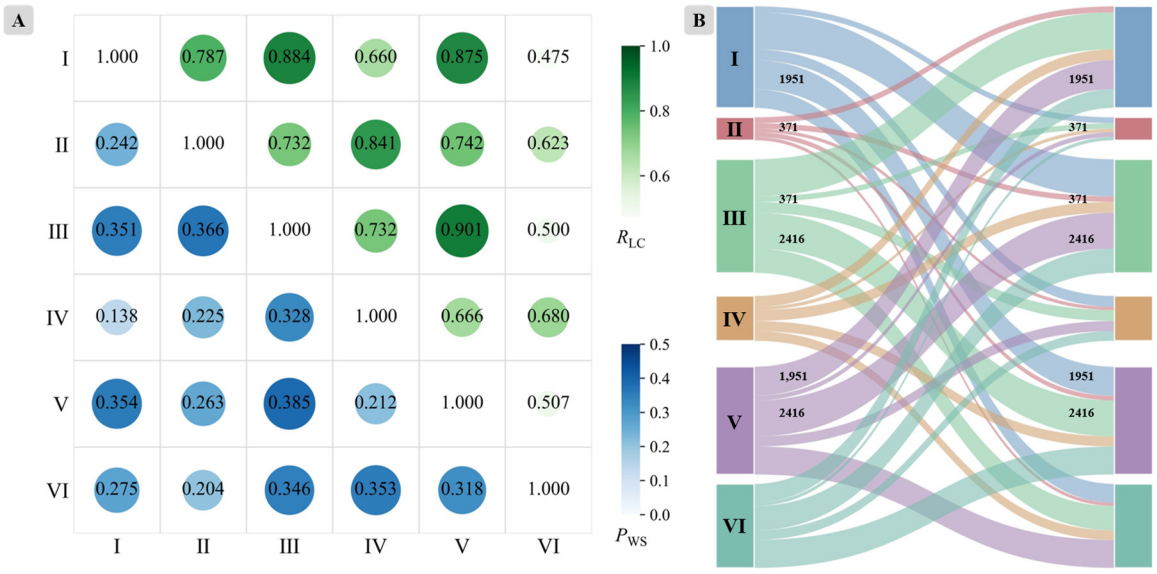


Figure 4. Visualization of the weight similarity parameter (P_{WS}), label consistency rate (R_{LC}) and dataset relatedness: (**A**) Heatmaps of the P_{WS} values for feature generator weights and R_{LC} values across six aquatic toxicity endpoints, including (I) fish acute toxicity, (II) fish chronic toxicity, (III) crustacean acute toxicity, (IV) crustacean chronic toxicity, (V) algal acute toxicity, and (VI) algal chronic toxicity; (**B**) Sankey diagram illustrating the overlap of shared compounds across the six aquatic toxicity endpoints.

In addition, fish chronic toxicity and crustacean acute toxicity exhibited a relatively high P_{WS} value (0.366), accompanied by significant performance improvements of MTL-GCN over STL models. Specifically, for the data-scarce fish chronic toxicity endpoint, the MTL-GCN achieved a 13.4% improvement in A_{ROC} over the STL-GCN despite sharing only 371 compounds with crustacean acute toxicity. The observation suggests that the number of shared compounds alone does not fully determine P_{WS} values and the effectiveness of knowledge transfer in the MTL framework.

To further elucidate the task affinity of related tasks, label consistency rate (R_{LC}) among shared compounds was calculated as an indicator of task affinity. A high R_{LC} indicates high task affinity between two endpoints, meaning that they tend to assign similar toxicity labels to the same chemicals. Specifically, a R_{LC} of 0.732 was observed between fish chronic and crustacean acute toxicity. This indicates that although these two endpoints share a limited number of chemicals, the shared compounds were still assigned highly concordant toxicity labels across the two endpoints, suggesting relatively high task affinity between the corresponding tasks. The comprehensive pairwise data including the P_{WS}, R_{LC} and the shared compounds are summarized in Table S14 to facilitate evaluation.

Collectively, the findings suggest that P_{WS}, shared chemicals and R_{LC} reflect the ability of the model not only to learn shared molecular structural features but also to capture the underlying biological correlations between endpoints. These provide mechanistic insight into the effectiveness of the MTL framework, demonstrating that the observed performance improvements arise not merely from increased data volume, but from the model's ability to integrate shared chemical structures and biologically meaningful cross-endpoint correlations.

3.3. Model Interpretation

To elucidate the model decision-making process, attention weights were visualized for fish acute toxicity (Figure 5), given that the endpoint comprises the largest dataset and exhibits the best prediction performance among the six endpoints. GCN models, including MTL-GCN and STL-GCN, assigned high attention weights to well-established toxicophores relevant to fish acute toxicity (Figure 5A,B). Prominent structural features included aliphatic hydroxyl groups, phenolic hydroxyl groups, and aldehyde groups, which are consistent with established mechanistic knowledge of aquatic toxicity [86–90].

Specifically, the high attention weights on aliphatic hydroxyl groups suggest that the model successfully captures the mechanism of non-polar narcosis. This mechanism acts as baseline toxicity, primarily disrupting membrane integrity [86]. Phenolic compounds, due to their hydrogen-bonding capability, can interact more strongly with biological macromolecules, resulting in enhanced toxicity compared to baseline toxicity [86,87]. Aldehyde groups, as reactive electrophiles, are capable of forming covalent bonds with nucleophilic sites in proteins and other biomolecules, thereby inducing toxic effects [87–89]. The accurate identification of these toxicophores demonstrates that the GCN models with the MG inputs effectively capture meaningful structure-toxicity relationships.

 https://doi.org/10.53941/ges.2026.100019

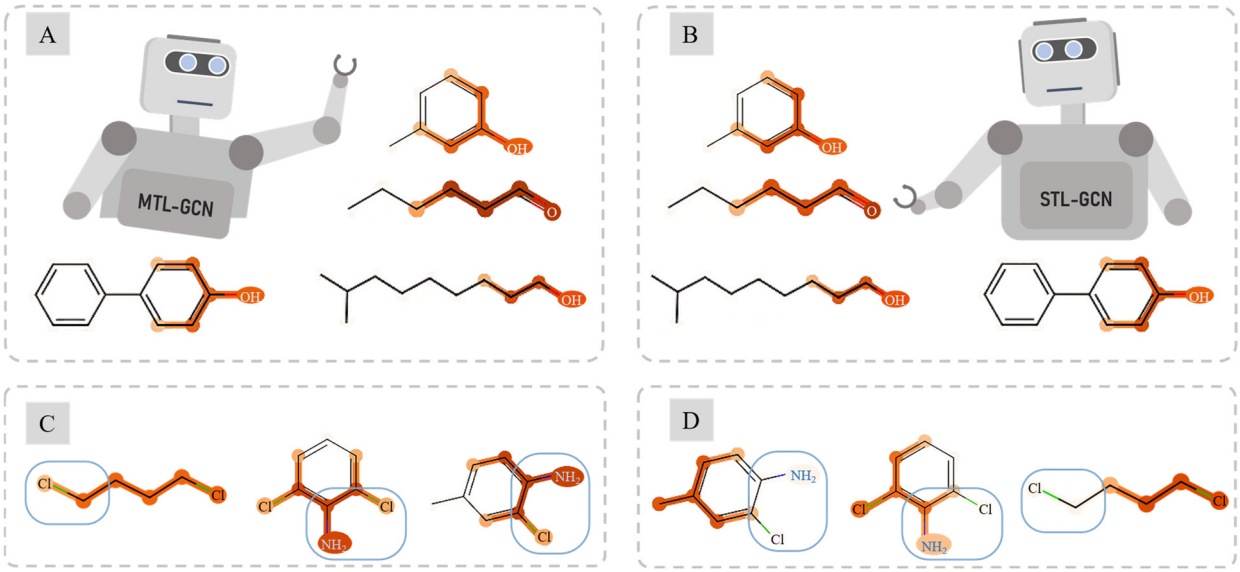


Figure 5. Visualization of attention weights for interpretability analysis of fish acute toxicity: (**A**) and (**B**) illustrate the consistent attention weights assigned by MTL-GCN and STL-GCN to representative toxicophores; (**C**) and (**D**) highlight representative examples of differences in attention distribution between the MTL-GCN and STL-GCN models for structurally complex toxicophores; blue circles indicate key substructures where the MTL-GCN demonstrates superior recognition capabilities.

Overall, the incorporation of an attention mechanism confers interpretability to the GCN architecture by highlighting structurally and mechanistically plausible determinants of toxicity. The enhanced transparency strengthens the reliability in the model's predictions and supports its application in chemical hazard assessment and regulatory decision-making contexts.

Notably, discernible disparities in attention weight distribution are observed in the representative examples in Figure 5C,D. The MTL-GCN assigns high attention weights to amino groups in anilines and to chlorine substituents, whereas STL-GCN attributes comparatively lower weights to the features. Aromatic amines, similar to phenols, can act through polar narcosis mechanisms mediated by hydrogen-bonding interactions, and thus constitute mechanistically relevant structural alerts [87]. Chlorine substituents can increase hydrophobicity and membrane partitioning, and in certain contexts can shift the toxic mode of action toward more potent mechanisms, such as uncoupling [89].

To further assess whether high attention weights were consistently assigned to the specific substructures, SMILES arbitrary target specification (SMARTS) matching was performed to identify molecules containing these substructures. The atomic attention weights for these substructures were extracted and compared between the MTL-GCN and STL-GCN. For instance, the mean atomic attention weight of aryl chlorine substituents in the MTL-GCN model was 3.7% higher than that in the STL-GCN. Similarly, for aromatic amines, the average attention weight in the MTL-GCN model was 80.7% higher than that in the STL-GCN. The attention weights for these and other analyzed substructures are provided in Table S15. Overall, the MTL-GCN generally yielded higher attention weights than the STL-GCN for these substructures. The analytical scripts are available in the GitHub repository.

The enhanced sensitivity of MTL-GCN to the structurally complex and context-dependent toxicophores indicates that MTL facilitates the extraction of more comprehensive and transferable molecular representations compared with STL modeling. Consequently, MTL-GCN demonstrates improved mechanistic coherence, which is advantageous for reliable chemical hazard evaluation.

3.4. Applicability Domain (AD) Characterization

The effects of varying thresholds on validation performance and AD coverage are illustrated in Figure 6, using fish acute toxicity as an example. As depicted, progressively stricter AD thresholds based on AD_{SAL} characterization lead to a gradual reduction in the number of validation compounds (n) retained within the AD, accompanied by an increase in prediction performance (A_{ROC}) for compounds within the AD. The results demonstrate the effectiveness of AD_{SAL} in delineating model reliability boundaries.

A joint analysis of AD_{SAL} for MTL-GCN and STL-GCN reveals AD coverage differences between the two modeling strategies (Figure 6, Tables S16 and S17). For fish acute toxicity, the A_{ROC} of MTL-GCN within the AD ranges from 0.854 to 0.945, whereas that of STL-GCN ranges from 0.882 to 0.929. In terms of AD coverage, the n of chemicals retained within the AD spans 164~478 for MTL-GCN, compared with 103~187 for STL-GCN. Notably, when the n within the AD was set to 175, MTL-GCN

yielded an A_{ROC} of 0.908, whereas STL-GCN only reached 0.884. The results demonstrate that MTL enables the model to expand its AD coverage when simultaneously enhancing in-domain performance. Consequently, MTL-GCN provides superior prediction reliability and an extensive reach in chemical space.

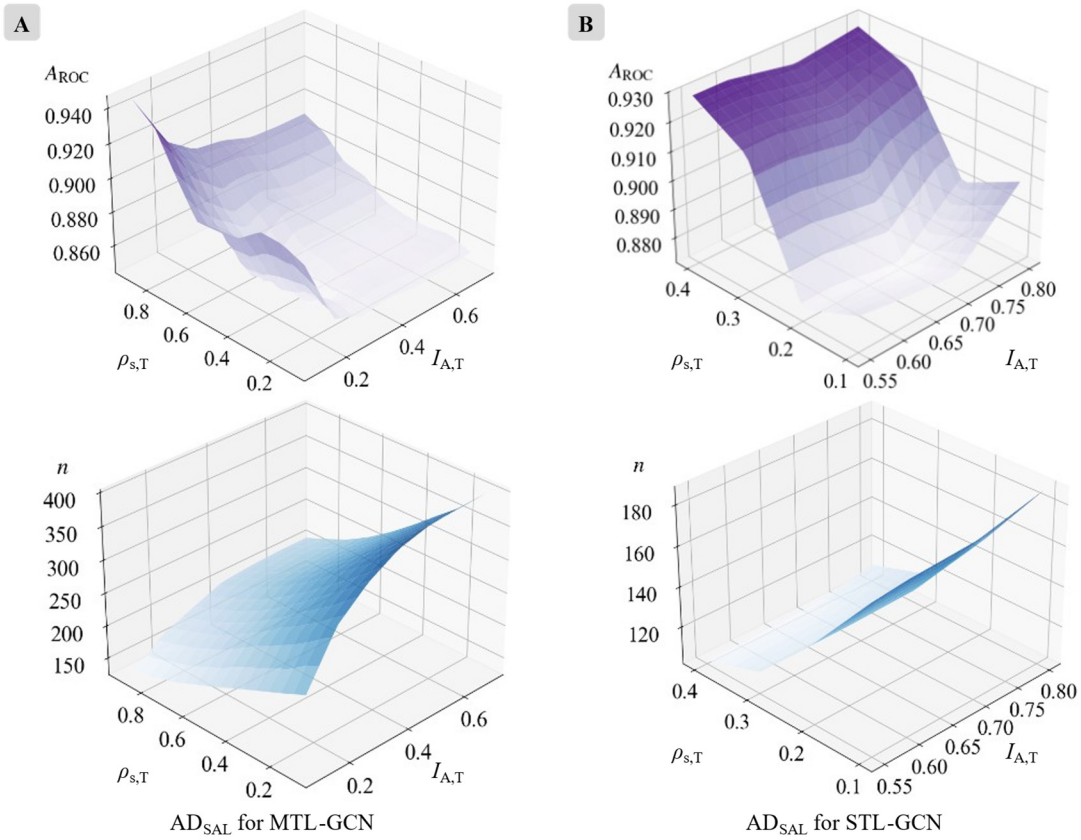


Figure 6. Effects of varying $\rho_{s,T}$ and $I_{A,T}$ values on the performance (A_{ROC}) and the number of compounds (n) within the ADs of the GNN models, where AD_{SAL} was characterized for (**A**) MTL-GCN and (**B**) STL-GCN.

The observed improvement can be attributed to the ability of the MTL framework to capture fundamental and generalizable molecular representations. By learning from multiple related toxicity endpoints, the MTL-GCN extracts shared structural and biochemical features common across tasks. The comprehensive training chemical space corresponds to the union of datasets from all endpoints, resulting in a substantially larger and more chemically diverse representation manifold than those of any STL models. Consequently, compounds are more likely to reside within the expanded AD of the MTL-GCN model, increasing the number of reliably predicted chemicals while preserving high prediction accuracy.

The appropriate thresholds for AD_{SAL} of the MTL-GCN model, together with the corresponding performance and AD coverage, are summarized in Table 3. As shown in Table S16, fish chronic toxicity reached an in-domain A_{ROC} of 1.000 under $AD_{SAL}\{\rho_{s,q} \geq 0.100, I_{A,q} \leq 0.900\}$. Furthermore, Table 3 indicates that the selected thresholds resulted in a high AD coverage, retaining approximately 83.0% of the compounds in the external validation set for this endpoint. Achieving this performance on this proportion of the validation data demonstrates that the AD_{SAL} filters out unreliable predictions. As the AD becomes more stringent, the number of retained chemicals continues to decrease while the A_{ROC} remains at 1.000. Therefore, the selected threshold represents a trade-off between model performance and coverage.

Table 3. Performance of the MTL-GCN model with AD_{SAL} characterization.

Endpoints	The Selected AD_{SAL} Threshold		Entire Validation Set		Validation Set in ADs	
	$\rho_{s,T}$	$I_{A,T}$	N	A_{ROC}	n	A_{ROC}
Fish acute toxicity	0.300	0.400	458	0.860	322	0.874
Fish chronic toxicity	0.100	0.900	53	0.806	44	1.000
Crustacean acute toxicity	0.100	0.900	451	0.809	396	0.893
Crustacean chronic toxicity	0.100	0.900	104	0.846	91	0.856
Algal acute toxicity	0.100	0.300	403	0.815	288	0.828
Algal chronic toxicity	0.500	0.800	282	0.778	149	0.812

$\rho_{s,T}$: threshold of weighted molecular similarity density; $I_{A,T}$: threshold of weighted inconsistency in molecular activities; N: the number of compounds in validation set; n: the number of compounds in AD_{SAL}; A_{ROC}: area under the receiver operating characteristic curve.

 https://doi.org/10.53941/ges.2026.100019

Under comparable in-domain coverage, the MTL-GCN with AD_{SAL} generally yielded better in-domain performance than that with the leverage-based AD method, indicating that AD_{SAL} provided a more effective characterization of reliable prediction regions for the present datasets (Tables S16 and S18). The selection of the thresholds followed a criterion that prioritized maximizing AD coverage while ensuring that in-domain performance consistently exceeded out-of-domain performance. This balance ensures both the breadth and reliability of the aquatic toxicity predictions. Ultimately, the integration of MTL-GCN with AD_{SAL} establishes an accurate and reliable framework for chemical hazard evaluation and risk management.

3.5. Application of the MTL-GCN Model

The constructed model was employed to identify aquatic toxic chemicals across six endpoints for a large-scale chemical inventory. The inventory, curated from previous research [63], initially comprised approximately 106 chemicals, from which 783,681 molecules were retained following DiPTox processing. Detailed information on the inventory screening procedure is provided in the Supplementary Materials. Subsequently, the constructed MTL-GCN coupled with AD_{SAL} was applied to screen the chemicals. The prediction outcomes and AD coverage statistics are summarized in Table S19.

Across the six endpoints, the proportion of chemicals falling within the AD ranged from 23.4% to 83.6%, with a median of 66.9%. The fish acute toxicity endpoint exhibited the highest coverage (83.6%), while the remaining endpoints also demonstrated substantial inclusion rates within the AD of the MTL-GCN model. Notably, a total of 742,384 chemicals were located within the AD of at least one endpoint, corresponding to an overall inventory coverage of 94.7%. The results indicate that the MTL framework effectively enlarges the chemical space encompassed by the model, thereby enhancing its applicability to structurally diverse chemicals. The broad AD coverage improves the capacity to prioritize potentially aquatic toxic chemicals for environmental hazard management.

To further explore the molecular structural basis of chemicals within the AD that exhibited toxicity across all six aquatic endpoints, these specific compounds were analyzed. A total of 49,810 chemicals were predicted to be toxic across all six endpoints by MTL-GCN with AD_{SAL}. Clustering analysis and chemical space visualization were conducted to characterize their structural distribution (Figure 7), with methodological details provided in the Supplementary Materials. Uniform manifold approximation and projection (UMAP) provided an instructive visualization tool to illustrate the structural distribution of the chemicals predicted to be toxic across all six aquatic endpoints.

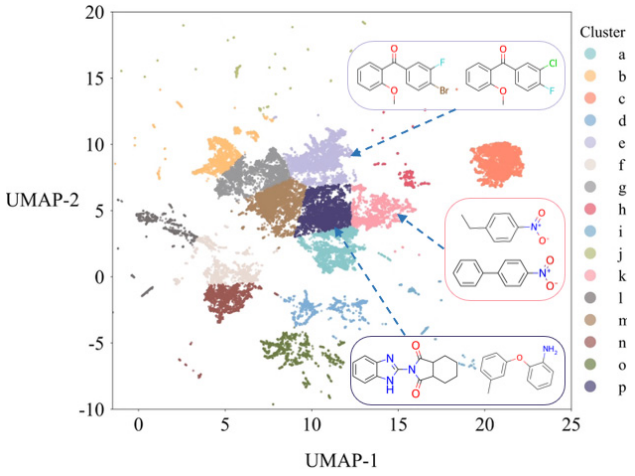


Figure 7. Chemical space visualization and structural clustering of chemicals identified to be toxic across six aquatic endpoints, which is visualized using UMAP, with representative molecules from selected clusters highlighted. UMAP-1 and UMAP-2 represent the two-dimensional embedding of the high-dimensional molecular fingerprints.

Subsequently, a structural analysis was performed for the three largest clusters. Substructure matching based on SMILES arbitrary target specification was employed to quantify the prevalence of key functional groups within each cluster. Detailed information on cluster assignments and predicted toxicity profiles is provided in Table S20. Cluster c, comprising 6069 chemicals, is predominantly characterized by nitrobenzene derivatives, which accounted for 97.5% of the cluster. Nitrobenzene is a well-documented toxic pollutant and was the principal contaminant in the Songhua River pollution incident [91]. Experimental studies have demonstrated that nitrobenzene exhibits acute toxicity toward aquatic organisms, including algae, crustaceans, and fish [91].

Cluster e, containing 5813 chemicals, witnesses a significant presence of halogenated benzenes (27.9%), a class of chemicals widely reported to induce toxic effects on aquatic organisms [11,92,93]. In cluster *p* (5010 chemicals), aromatic amines constitute the dominant structural category (52.9%). The chemicals are known to induce both acute and chronic toxicity in fish and algae [94–96]. Notably, the prominence of aromatic amines is consistent with the model interpretability analysis, where MTL-GCN attention mechanisms assigned high weights to substructures corresponding to aromatic amine functionalities. This concordance between mechanistic interpretability and large-scale screening results provides additional evidence of the model's capacity to capture chemically meaningful toxicophores.

Overall, the convergence of high AD coverage and consistency with established toxicological knowledge [11,91–96] supports the robustness and practical utility of the MTL-GCN framework. The MTL-GCN model demonstrates strong potential as a high-throughput screening method

for large-scale chemical prioritization and environmental risk management.

The predicted toxic chemicals outside all ADs are listed in Table S21. Their presence reflects the objective trade-off between prediction reliability and AD coverage in large-scale screening, highlighting the practical advantage of the AD_{SAL} method in improving screening reliability. Consequently, for large scale aquatic hazard screening, integrating AD_{SAL} with MTL-GCN effectively manages this trade off and ensures reliable results. Additional supporting information is available in the Supplementary Materials.

4. Conclusions

This study curated a comprehensive aquatic toxicity dataset encompassing 17,514 chemicals across six endpoints, covering an unprecedented chemical space to support large-scale regulatory screening. Using molecular graphs as inputs, an end-to-end GCN was developed within an MTL framework to jointly model multiple toxicity endpoints. By extracting shared molecular representations, the MTL-GCN improved the A_{ROC} by 0.4~13.4% compared with STL models. The proposed weight similarity parameter (P_{WS}) enabled quantitative analysis of knowledge transfer across endpoints, demonstrating that MTL effectively captures intrinsic correlations between different endpoints. Incorporation of an attention mechanism further enhanced interpretability, with substructure-level visualization confirming that MTL-GCN accurately identifies meaningful aquatic toxicophores. The AD characterization method based on the SALs analysis demonstrated that MTL-GCN maintains equivalent chemical coverage while improving prediction accuracy compared to STL models. In screening a million-scale chemical inventory, the combined MTL-GCN and AD_{SAL} framework achieved 94.7% coverage. Overall, the constructed dataset and the proposed MTL-GCN framework provide a robust, reliable, and interpretable benchmark for multi-endpoint aquatic toxicity screening, offering methodological guidance and practical decision support for large-scale ecological hazard assessment of chemicals.

Supplementary Materials

The additional data and information can be downloaded at: https://media.sciltp.com/articles/others/260722103127027O/GES-26020135-Supplementary-Materials.7z. The MTL-GCN model, STL-models, and the data are available free of charge and stored at https://github.com/JssXmy/MTL-GCN-AqTox. Curation of Aquatic Toxicity Data; detailed information on dataset processing; detailed description of the features used in GCN; implementation details of single-task learning GCN models; details on AD_{SAL} characterization about calculation; summary of endpoints and integrated datasets; construction of Tanimoto similarity distributions; methods for physicochemical property calculation; details on method chemical clustering and visualization; calculation of endpoint label consistency rate; detailed processing and screening of chemicals inventory; details on visualization of inventory chemicals with six endpoints toxicity; visualization of the chemical space distribution: Figure S1. Visualization of the chemical space distribution for six aquatic toxicity endpoints through principal component analysis; visualization of Tanimoto similarity distribution: Figure S2. Pairwise Tanimoto similarity distributions for the constructed datasets across six aquatic toxicity endpoints; and visualization of physiochemical property distribution: Figure S3. Physicochemical property distributions of descriptors relevant to aquatic toxicity for the constructed datasets across six endpoints. Related aquatic toxicity datasets; validation performance of MTL-GCN based on random splitting and scaffold splitting; statistical comparisons of MTL-GCN and STL-GCN performance under ten random seeds; hyperparameters of all models; comparisons of atomic attention weights between the MTL-GCN and STL-GCN models; results of AD_{SAL} and leverage-based AD; results of endpoint R_{LC}, corresponding with P_{WS} and number of shared chemicals; the screening summary of the chemical inventory and the identified chemicals within the inventory predicted to simultaneously exhibit all six aquatic toxicity endpoints, along with their corresponding molecular cluster information (XLSX). Reference [97] are cited in the supplementary materials.

Author Contributions

S.J.: Methodology, Investigation, Data curation, Machine learning modeling, Writing—original draft, Visualization; W.L.: Conceptualization, Visualization, Writing—reviewing and editing; G.B.: Visualization; H.W.: Conceptualization, Methodology, Visualization, Writing—reviewing and editing; J.C.: Conceptualization, Project administration, Funding acquisition, Writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (22136001, 225B2603).

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Relevant data are available upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-assisted Technologies

No AI tools were utilized for this paper.

References

1. Schwarzenbach, R.P.; Escher, B.I.; Fenner, K.; et al. The Challenge of Micropollutants in Aquatic Systems. *Science* **2006**, *313*, 1072–1077.
2. Wang, M.; Bodirsky, B.L.; Rijneveld, R.; et al. A Triple Increase in Global River Basins with Water Scarcity Due to Future Pollution. *Nat. Commun.* **2024**, *15*, 880.
3. Singh, P.K.; Kumar, U.; Kumar, I.; et al. Critical Review on Toxic Contaminants in Surface Water Ecosystem: Sources, Monitoring, and Its Impact on Human Health. *Environ. Sci. Pollut. Res.* **2024**, *31*, 56428–56462.
4. Lin, L.; Yang, H.; Xu, X. Effects of Water Pollution on Human Health and Disease Heterogeneity: A Review. *Front. Environ. Sci.* **2022**, *10*, 880246.
5. Münzel, T.; Hahad, O.; Lelieveld, J.; et al. Soil and Water Pollution and Cardiovascular Disease. *Nat. Rev. Cardiol.* **2025**, *22*, 71–89.
6. European Parliament; Council of the European Union. Regulation (EC) No. 1907/2006 Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH); Luxembourg, 2006. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006R1907 (accessed on 15 July 2025).
7. U.S. Environmental Protection Agency. United States Code, 2006 Edition, Supplement 5, Title 15—Commerce and Trade Chapter 53—Toxic Substances Control; Washington, DC, USA, 2006. Available online: https://www.govinfo.gov/app/details/USCODE-2011-title15/USCODE-2011-title15-chap53 (accessed on 15 July 2025).
8. Ministry of Ecology and Environment of the People's Republic of China. Measures for the Environmental Management Registration of New Chemical Substances (Mee Order No. 12); Beijing, 2020. Available online: https://www.gov.cn/gongbao/content/2020/content_5530350.htm (accessed on 15 July 2025).
9. United Nations. Globally Harmonized System of Classification and Labelling of Chemicals (GHS Rev. 10, 2023); New York and Geneva, 2023. Available online: https://unece.org/transport/standards/transport/dangerous-goods/ghs-rev10-2023 (accessed on 17 July 2025).
10. Zhou, N.; Sui, S.; Liu, H.; et al. Determining High Priority Disinfection Byproducts Based on Experimental Aquatic Toxicity Data and Predictive Models: Virtual Screening and in Vivo Study. *Sci. Total Environ.* **2024**, *951*, 175489.
11. Wang, Y.; Liu, H.; Yang, X.; et al. Aquatic Toxicity and Aquatic Ecological Risk Assessment of Wastewater-Derived Halogenated Phenolic Disinfection Byproducts. *Sci. Total Environ.* **2022**, *809*, 151089.
12. Hartung, T.; Rovida, C. Chemical Regulators Have Overreached. *Nature* **2009**, *460*, 1080–1081.
13. Krewski, D.; Andersen, M.E.; Tyshenko, M.G.; et al. Toxicity Testing in the 21st Century: Progress in the Past Decade and Future Perspectives. *Arch. Toxicol.* **2020**, *94*, 1–58.
14. Merchant, A.; Batzner, S.; Schoenholz, S.S.; et al. Scaling Deep Learning for Materials Discovery. *Nature* **2023**, *624*, 80–85.
15. Rao, Z.; Tung, P.-Y.; Xie, R.; et al. Machine Learning–Enabled High-Entropy Alloy Discovery. *Science* **2022**, *378*, 78–85.
16. Anand, G.; Koniusz, P.; Kumar, A.; et al. Graph Neural Networks-Enhanced Relation Prediction for Ecotoxicology (GRAPE). *J. Hazard. Mater.* **2024**, *472*, 134456.
17. Allah, M.A.-A.; Toor, I.U.; Shams, A.; et al. Application of Machine Learning and Deep Learning Techniques for Corrosion and Cracks Detection in Nuclear Power Plants: A Review. *Arab. J. Sci. Eng.* **2025**, *50*, 3017–3045.
18. Bi, K.; Xie, L.; Zhang, H.; et al. Accurate Medium-Range Global Weather Forecasting with 3D Neural Networks. *Nature* **2023**, *619*, 533–538.
19. Reichstein, M.; Benson, V.; Blunk, J.; et al. Early Warning of Complex Climate Risk with Integrated Artificial Intelligence. *Nat. Commun.* **2025**, *16*, 2564.
20. Gohr, C.; Rodríguez, G.; Belomestnykh, S.; et al. Artificial Intelligence in Sustainable Development Research. *Nat. Sustain.* **2025**, *8*, 970–978.
21. Tuia, D.; Kellenberger, B.; Beery, S.; et al. Perspectives in Machine Learning for Wildlife Conservation. *Nat. Commun.* **2022**, *13*, 792.
22. Kavitha, R.; Guru, D.S. Quantitative Structure-Activity Relationship Modeling for the Prediction of Fish Toxicity Lethal Concentration on Fathead Minnow. In Proceedings of the Pattern Recognition ICPR 2024, Kolkata, India, 1–5 December 2024.
23. Cao, Q.; Liu, L.; Yang, H.; et al. *In Silico* Estimation of Chemical Aquatic Toxicity on Crustaceans Using Chemical Category Methods. *Environ. Sci. Process. Impacts* **2018**, *20*, 1234–1243.
24. Gajewicz-Skretna, A.; Wyrzykowska, E.; Gromelski, M. Quantitative Multi-Species Toxicity Modeling: Does a Multi-Species, Machine Learning Model Provide Better Performance Than a Single-Species Model for the Evaluation of Acute Aquatic Toxicity by Organic Pollutants? *Sci. Total Environ.* **2023**, *861*, 160590.
25. Xu, M.; Yang, H.; Liu, G.; et al. *In Silico* Prediction of Chemical Aquatic Toxicity by Multiple Machine Learning and Deep Learning Approaches. *J. Appl. Toxicol.* **2022**, *42*, 1766–1776.
26. Gustavsson, M.; Käll, S.; Svedberg, P.; et al. Transformers Enable Accurate Prediction of Acute and Chronic Chemical Toxicity in Aquatic Organisms. *Sci. Adv.* **2024**, *10*, eadk6669.
27. Sheffield, T.Y.; Judson, R.S. Ensemble QSAR Modeling to Predict Multispecies Fish Toxicity Lethal Concentrations and Points of Departure. *Environ. Sci. Technol.* **2019**, *53*, 12793–12802.

28. Li, F.; Fan, D.; Wang, H.; et al. *In Silico* Prediction of Pesticide Aquatic Toxicity with Chemical Category Approaches. *Toxicol. Res.* **2017**, *6*, 831–842.
29. Lunghini, F.; Marcou, G.; Azam, P.; et al. Consensus QSAR Models Estimating Acute Toxicity to Aquatic Organisms from Different Trophic Levels: Algae, Daphnia and Fish. *SAR QSAR Environ. Res.* **2020**, *31*, 655–675.
30. Gunnarsson, L.; Snape, J.R.; Verbruggen, B.; et al. Pharmacology Beyond the Patient—The Environmental Risks of Human Drugs. *Environ. Int.* **2019**, *129*, 320–332.
31. Chen, C.; Yang, B.; Li, M.; et al. Quantitative Structure–Activity Relationship Predicting Toxicity of Pesticides Towards Daphnia Magna. *Ecotoxicology* **2024**, *33*, 560–568.
32. Ding, F.; Wang, Z.; Yang, X.; et al. Development of Classification Models for Predicting Chronic Toxicity of Chemicals to Daphnia Magna and Pseudokirchneriella Subcapitata. *SAR QSAR Environ. Res.* **2019**, *30*, 39–50.
33. Jin, X.; Jin, M.; Sheng, L. Three Dimensional Quantitative Structure–Toxicity Relationship Modeling and Prediction of Acute Toxicity for Organic Contaminants to Algae. *Comput. Biol. Med.* **2014**, *51*, 205–213.
34. Yang, Y.; Yang, Y.; Pan, W.; et al. Multimodal Integration of Chemical and Biological Descriptors for Cross-Species Prediction of Fish Acute Toxicity. *Environ. Sci. Technol.* **2026**, *60*, 2556–2565.
35. Liang, W.; Li, J.; Wang, X.; et al. Machine Learning-Driven Cross-Species Toxicity Prediction for Advancing Ecologically Relevant PFAS Water Quality Criteria. *Environ. Sci. Technol.* **2025**, *59*, 25688–25702.
36. Pizzo, F.; Lombardo, A.; Manganaro, A.; et al. Integrated *in Silico* Strategy for PBT Assessment and Prioritization under REACH. *Environ. Res.* **2016**, *151*, 478–492.
37. Merkwirth, C.; Lengauer, T. Automatic Generation of Complementary Descriptors with Molecular Graph Networks. *J. Chem. Inf. Model.* **2005**, *45*, 1159–1168.
38. Yang, K.; Swanson, K.; Jin, W.; et al. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59*, 3370–3388.
39. Deng, J.; Yang, Z.; Wang, H.; et al. A Systematic Study Of key Elements Underlying Molecular Property Prediction. *Nat. Commun.* **2023**, *14*, 6395.
40. Li, J.; Zhao, T.; Yang, Q.; et al. A Review of Quantitative Structure-Activity Relationship: The Development and Current Status of Data Sets, Molecular Descriptors and Mathematical Models. *Chemom. Intell. Lab. Syst.* **2025**, *256*, 105278.
41. Toma, C.; Cappelli, C.I.; Manganaro, A.; et al. New Models to Predict the Acute and Chronic Toxicities of Representative Species of the Main Trophic Levels of Aquatic Environments. *Molecules* **2021**, *26*, 6983.
42. Kramer, L.; Schulze, T.; Klüver, N.; et al. Curated Mode-of-Action Data and Effect Concentrations for Chemicals Relevant for the Aquatic Environment. *Sci. Data* **2024**, *11*, 60.
43. Zhang, K.; Schwaller, P.; Fenner, K. Predicting Toxicity toward Nitrifiers by Attention-Enhanced Graph Neural Networks and Transfer Learning from Baseline Toxicity. *Environ. Sci. Technol.* **2025**, *59*, 4518–4529.
44. Tan, X.; Liu, Q.; Zhou, M.; et al. Pepadmet: A Novel Computational Platform for Systematic Admet Evaluation of Peptides. *J. Chem. Inf. Model.* **2026**, *66*, 936–946.
45. Xiao, Z.; Zhu, M.; Chen, J.; et al. Integrated Transfer Learning and Multitask Learning Strategies to Construct Graph Neural Network Models for Predicting Bioaccumulation Parameters of Chemicals. *Environ. Sci. Technol.* **2024**, *58*, 15650–15660.
46. Xiao, Z.; Zhu, M.; Chen, J. Improving Prediction Performance on Solute Parameters Using Multitask Relational Graph Convolutional Networks with Explicit Hydrogens. *ACS ES&T Water* **2024**, *4*, 5969–5979.
47. Mayr, A.; Klambauer, G.; Unterthiner, T.; et al. Large-Scale Comparison of Machine Learning Methods for Drug Target Prediction on Chembl. *Chem. Sci.* **2018**, *9*, 5441–5451.
48. Schlender, T.; Viljanen, M.; van Rijn, J.N.; et al. The Bigger Fish: A Comparison of Meta-Learning QSAR Models on Low-Resourced Aquatic Toxicity Regression Tasks. *Environ. Sci. Technol.* **2023**, *57*, 17818–17830.
49. Muratov, E.N.; Bajorath, J.; Sheridan, R.P.; et al. QSAR without Borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564.
50. Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug Discovery with Explainable Artificial Intelligence. *Nat. Mach. Intell.* **2020**, *2*, 573–584.
51. Proietti, M.; Ragno, A.; Rosa, B.L.; et al. Explainable Ai in Drug Discovery: Self-Interpretable Graph Neural Network for Molecular Property Prediction Using Concept Whitening. *Mach. Learn.* **2024**, *113*, 2013–2044.
52. Wong, F.; Zheng, E.J.; Valeri, J.A.; et al. Discovery of a Structural Class of Antibiotics with Explainable Deep Learning. *Nature* **2024**, *626*, 177–185.
53. Novakovsky, G.; Dexter, N.; Libbrecht, M.W.; et al. Obtaining Genetics Insights from Deep Learning Via Explainable Artificial Intelligence. *Nat. Rev. Genet.* **2023**, *24*, 125–137.
54. Polishchuk, P. Interpretation of Quantitative Structure–Activity Relationship Models: Past, Present, and Future. *J. Chem. Inf. Model.* **2017**, *57*, 2618–2639.
55. Wang, H.; Wang, Z.; Chen, J.; et al. Graph Attention Network Model with Defined Applicability Domains for Screening PBT Chemicals. *Environ. Sci. Technol.* **2022**, *56*, 6774–6785.
56. Wang, H.; Liu, W.; Chen, J.; et al. Transfer Learning with a Graph Attention Network and Weighted Loss Function for Screening of Persistent, Bioaccumulative, Mobile, and Toxic Chemicals. *Environ. Sci. Technol.* **2025**, *59*, 578–590.
57. Xiong, Z.; Wang, D.; Liu, X.; et al. Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. *J. Med. Chem.* **2020**, *63*, 8749–8760.
58. Wu, J.; D'Ambrosi, S.; Ammann, L.; et al. Predicting Chemical Hazard across Taxa through Machine Learning. *Environ. Int.* **2022**, *163*, 107184.
59. Organisation for Economic Co-operation and Development (OECD). *Guidance Document on the Validation of (Quantitative) Structure Activity Relationships (Q)SAR Models*; OECD: Paris, France, 2014. https://doi.org/10.1787/9789264085442-en.

60. Zhang, Z.; Sangion, A.; Wang, S.; et al. Chemical Space Covered by Applicability Domains of Quantitative Structure–Property Relationships and Semiempirical Relationships in Chemical Assessments. *Environ. Sci. Technol.* **2024**, *58*, 3386–3398.
61. Cumming, J.G.; Davis, A.M.; Muresan, S.; et al. Chemical Predictive Modelling to Improve Compound Quality. *Nat. Rev. Drug Discov.* **2013**, *12*, 948–962.
62. Weaver, S.; Gleeson, M.P. The Importance of the Domain of Applicability in QSAR Modeling. *J. Mol. Graph. Model.* **2008**, *26*, 1315–1326.
63. Wang, H.; Liu, W.; Chen, J.; et al. Applicability Domains Based on Molecular Graph Contrastive Learning Enable Graph Attention Network Models to Accurately Predict 15 Environmental End Points. *Environ. Sci. Technol.* **2023**, *57*, 16906–16917.
64. Liu, W.; Wang, Z.; Chen, J.; et al. Machine Learning Model for Screening Thyroid Stimulating Hormone Receptor Agonists Based on Updated Datasets and Improved Applicability Domain Metrics. *Chem. Res. Toxicol.* **2023**, *36*, 947–958.
65. Wu, C.; Chen, J.; Zhang, Y.; et al. Machine Learning Models Based on Enlarged Chemical Spaces for Screening Carcinogenic Chemicals. *Chem. Res. Toxicol.* **2025**, *38*, 1192–1202.
66. Organisation for Economic Co-Operation and Development (OECD). *Test No. 203: Fish, Acute Toxicity Test*; OECD: Paris, France, 2025. https://doi.org/10.1787/9789264069961-en.
67. Organisation for Economic Co-Operation and Development (OECD). *Test No. 210: Fish, Early-Life Stage Toxicity Test*; OECD: Paris, France, 2025. https://doi.org/10.1787/9789264203785-en.
68. Organisation for Economic Co-Operation and Development (OECD). *Test No. 202: Daphnia Sp. Acute Immobilisation Test*; OECD: Paris, France, 2004. https://doi.org/10.1787/9789264069947-en.
69. Organisation for Economic Co-Operation and Development (OECD). *Test No. 211: Daphnia Magna Reproduction Test*; OECD: Paris, France, 2012. https://doi.org/10.1787/9789264185203-en.
70. Organisation for Economic Co-Operation and Development (OECD). *Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test*; OECD: Paris, France, 2011. https://doi.org/10.1787/9789264069923-en.
71. U.S. Environmental Protection Agency. ECOTOX Knowledgebase; Washington, DC, USA. Available online: https://cfpub.epa.gov/ecotox/ (accessed on 31 July 2025).
72. Organisation for Economic Co-Operation and Development (OECD). eChemPortal: Global Portal to Information on Chemical Substances; Paris. Available online: https://echemportal.org/echemportal/ (accessed on 31 July 2025).
73. Connors, K.A.; Beasley, A.; Barron, M.G.; et al. Creation of a Curated Aquatic Toxicology Database: EnviroTox. *Environ. Toxicol. Chem.* **2019**, *38*, 1062–1073.
74. Wu, L.; Yan, B.; Han, J.; et al. TOXRIC: A Comprehensive Database of Toxicological Data and Benchmarks. *Nucleic Acids Res.* **2022**, *51*, D1432–D1445.
75. Ministry of the Environment Government of Japan. Results of Aquatic Toxicity Tests of Chemicals Conducted by Ministry of the Environment in Japan; Tokyo, 2016. Available online: https://www.env.go.jp/en/chemi/sesaku/aquatic_Mar_2016.pdf (accessed on 31 July 2025).
76. Wang, Z.; Chen, J.; Hong, H. Developing QSAR Models with Defined Applicability Domains on PPARγ Binding Affinity Using Large Data Sets and Machine Learning Algorithms. *Environ. Sci. Technol.* **2021**, *55*, 6857–6866.
77. Liu, W.; Wang, H.; Fu, Z.; et al. Mcgpcr: A Multimodal Learning Model with Improved Applicability Domain Characterization for Predicting G Protein-Coupled Receptor Affinity of Plastic Chemicals. *Environ. Sci. Technol.* **2025**, *59*, 25938–25949.
78. Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *J. Chem. Inf. Model.* **2010**, *50*, 1189–1204.
79. Wang, M.; Zheng, D.; Ye, Z.; et al. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. *arXiv* **2020**, arXiv:1909.01315.
80. Wu, Z.; Jiang, D.; Wang, J.; et al. Mining Toxicity Information from Large Amounts of Toxicity Data. *J. Med. Chem.* **2021**, *64*, 6924–6936.
81. Wang, Z.; Chen, J.; Hong, H. Applicability Domains Enhance Application of PPARγ Agonist Classifiers Trained by Drug-Like Compounds to Environmental Chemicals. *Chem. Res. Toxicol.* **2020**, *33*, 1382–1388.
82. Gajewicz-Skretna, A.; Gromelski, M.; Wyrzykowska, E.; et al. Aquatic Toxicity (Pre)Screening Strategy for Structurally Diverse Chemicals: Global or Local Classification Tree Models? *Ecotoxicol. Environ. Saf.* **2021**, *208*, 111738.
83. Lambert, F.N.; Vivian, D.N.; Raimondo, S.; et al. Relationships between Aquatic Toxicity, Chemical Hydrophobicity, and Mode of Action: Log Kow Revisited. *Arch. Environ. Contam. Toxicol.* **2022**, *83*, 326–338.
84. Yang, L.; Sang, C.; Wang, Y.; et al. Development of QSAR Models for Evaluating Pesticide Toxicity against Skeletonema Costatum. *Chemosphere* **2021**, *285*, 131456.
85. Tinkov, O.V.; Grigorev, V.Y.; Razdolsky, A.N.; et al. Effect of the Structural Factors of Organic Compounds on the Acute Toxicity toward Daphnia Magna. *SAR QSAR Environ. Res.* **2020**, *31*, 615–641.
86. Mackay, D.; Arnot, J.A.; Petkova, E.P.; et al. The Physicochemical Basis of QSARs for Baseline Toxicity. *SAR QSAR Environ. Res.* **2009**, *20*, 393–414.
87. Netzeva, T.I.; Pavan, M.; Worth, A.P. Review of (Quantitative) Structure–Activity Relationships for Acute Aquatic Toxicity. *QSAR Comb. Sci.* **2008**, *27*, 77–90.
88. Enoch, S.J.; Ellison, C.M.; Schultz, T.W.; et al. A Review of the Electrophilic Reaction Chemistry Involved in Covalent Protein Binding Relevant to Toxicity. *Crit. Rev. Toxicol.* **2011**, *41*, 783–802.

89. Escher, B.I.; Hermens, J.L.M. Modes of Action in Ecotoxicology: Their Role in Body Burdens, Species Sensitivity, QSARs, and Mixture Effects. *Environ. Sci. Technol.* **2002**, *36*, 4201–4217.
90. Park, W.H. A Comprehensive Review of Pyrogallol: From Fundamental Chemistry to Advanced Applications and Toxicological Insights. *Biotechnol. Bioeng.* **2026**, *123*, 527–543.
91. Yan, Z.-G.; Zhang, Z.-S.; Wang, H.; et al. Development of Aquatic Life Criteria for Nitrobenzene in China. *Environ. Pollut.* **2012**, *162*, 86–90.
92. Li, Y.; Wen, W.; Zhang, J.; et al. Halogenated Organic Pollutants in Aquatic Invertebrate—Fish Food Webs: Global Distributions and Trophic Magnification. *J. Environ. Sci.* **2025**, *157*, 123–133.
93. Li, J.J.; Yue, Y.X.; Shi, S.J.; et al. Investigation on Toxicity Mechanism of Halogenated Aromatic Disinfection by-Products to Zebrafish Based on Molecular Docking and QSAR Model. *Chemosphere* **2023**, *341*, 139916.
94. Sharma, G.; Chadha, P. Toxic Effects of Aniline in Liver, Gills and Kidney of Freshwater Fish Channa Punctatus after Acute Exposure. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **2024**, *281*, 109916.
95. Dewulf, F.; Semmouri, I.; Van Nieuwerburgh, F.; et al. Exploring Mode of Action for Acute Toxicity of Primary Aromatic Amines with Daphnia Magna Using Differential Gene Expression Analysis. *J. Hazard. Mater.* **2025**, *488*, 137377.
96. Dobaradaran, S.; Schmidt, T.C.; Mutke, X.A.M.; et al. Aromatic Amines Leachate from Cigarette Butts into Aquatic Environments: Is There Risk for Water Organisms? *Environ. Res.* **2023**, *216*, 114717.
97. Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way to Cluster Small and Large Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 747–750.