

GeoDILI: A Robust and Interpretable Model for Drug-Induced Liver Injury Prediction Using Graph Neural Network-Based Molecular Geometric Representation

Wenxuan Wu,[§] Jiayu Qian,[§] Changjie Liang, Jingya Yang, Guangbo Ge, Qingping Zhou,* and Xiaoqing Guan*



Cite This: *Chem. Res. Toxicol.* 2023, 36, 1717–1730



Read Online

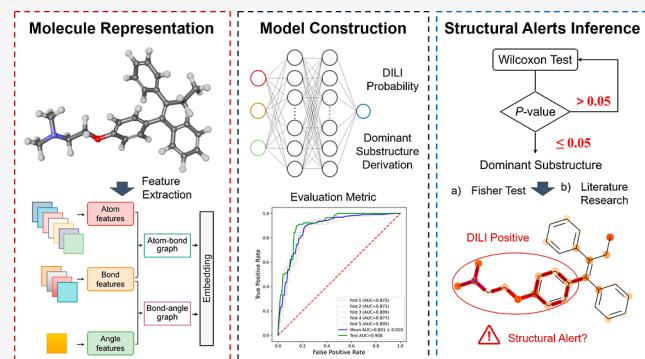
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Drug-induced liver injury (DILI) is a significant cause of drug failure and withdrawal due to liver damage. Accurate prediction of hepatotoxic compounds is crucial for safe drug development. Several DILI prediction models have been published, but they are built on different data sets, making it difficult to compare model performance. Moreover, most existing models are based on molecular fingerprints or descriptors, neglecting molecular geometric properties and lacking interpretability. To address these limitations, we developed GeoDILI, an interpretable graph neural network that uses a molecular geometric representation. First, we utilized a geometry-based pretrained molecular representation and optimized it on the DILI data set to improve predictive performance. Second, we leveraged gradient information to obtain high-precision atomic-level weights and deduce the dominant substructure. We benchmarked GeoDILI against recently published DILI prediction models, as well as popular GNN models and fingerprint-based machine learning models using the same data set, showing superior predictive performance of our proposed model. We applied the interpretable method in the DILI data set and derived seven precise and mechanistically elucidated structural alerts. Overall, GeoDILI provides a promising approach for accurate and interpretable DILI prediction with potential applications in drug discovery and safety assessment. The data and source code are available at GitHub repository (<https://github.com/CSU-QJY/GeoDILI>).



1. INTRODUCTION

The liver plays a vital role in metabolism and detoxification, which makes it vulnerable to damage from exogenous compounds such as drugs and environmental chemicals. Drug-induced liver injury (DILI) refers to the liver injury caused by the drug itself and/or its metabolites, which is one of the main reasons for drug failure and withdrawal or termination from the market in the later stage of clinical trials.^{1–4} DILI risk assessment has become one of the important issues in safe drug development, and therefore there is a high demand for developing predictive models to identify potential hepatotoxic compounds during the early stages of drug development.^{5,6}

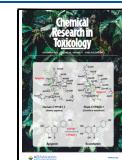
The underlying mechanisms of DILI are complex and varied.^{1,3} DILI can be classified as intrinsic or idiosyncratic based on the dose-dependent manner of the drug. Intrinsic DILI is dose-dependent and predictable in preclinical animal or *in vitro* studies, while idiosyncratic hepatotoxicity is dose-independent and usually unpredictable in regulatory-required animal/*cell* toxicity experiments.^{7–9} Many various *in vitro* and *in vivo* assays have been developed for DILI risk assessment. However, previous studies have shown that the results of

preclinical assessments of DILI (cellular models, animal models, and so on) and those of humans do not always coincide.^{10–13} A retrospective analysis showed that animal testing missed 40–45% of liver toxicity cases during clinical trials.^{14,15} The limited capability of the existing methods raises the need for more efficient testing approaches.

Over the past decade, high-throughput screening assays and combinatorial chemistry have generated a variety of biological data on millions of compounds, and the advancement of data-sharing programs has brought toxicology research into the era of big data. However, human hepatotoxicity data are extremely hard to collect and easily mislabeled. The main reasons are (1) the limited number of reports due to the voluntary character of information collection, (2) difficulty in the availability of

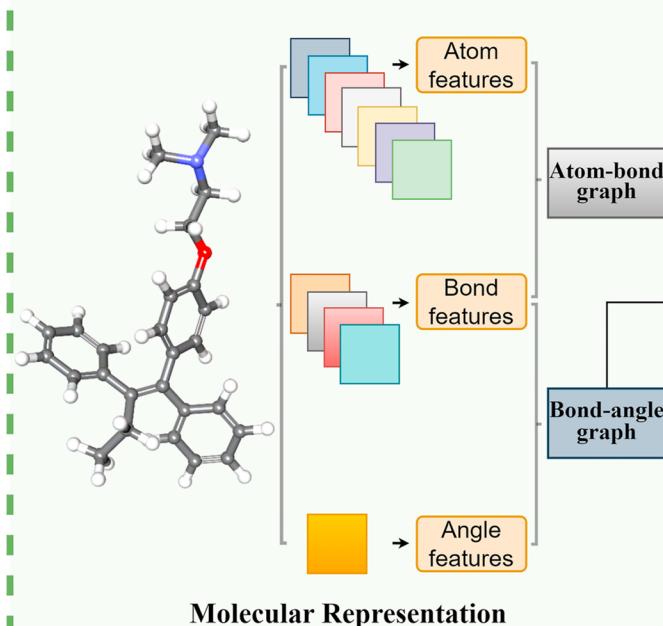
Received: July 10, 2023

Published: October 15, 2023

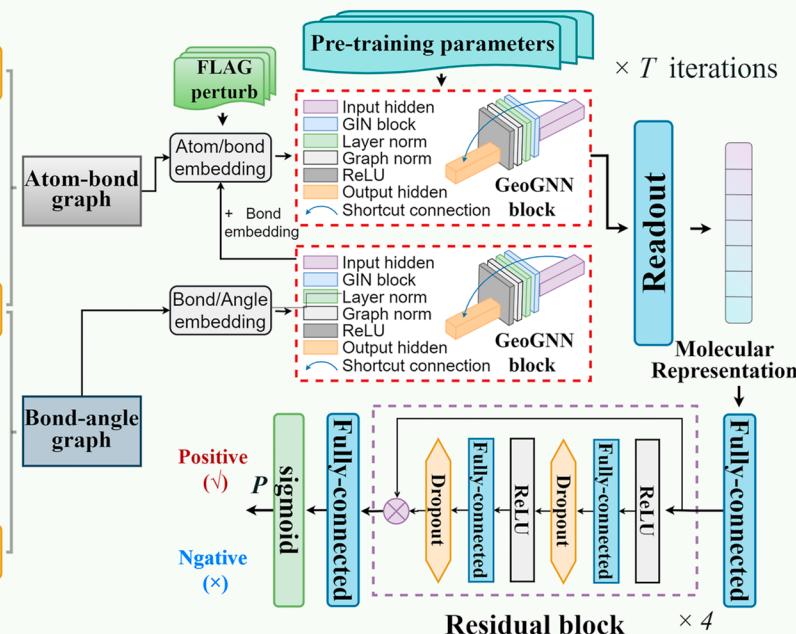


DILI Prediction Module

Feature Extraction Phase

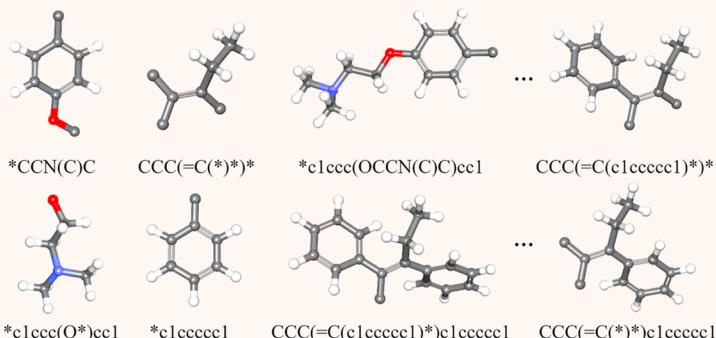


Fine-tuning Phase



Dominant Substructures Derivation

Fragmentation



Statistical Test

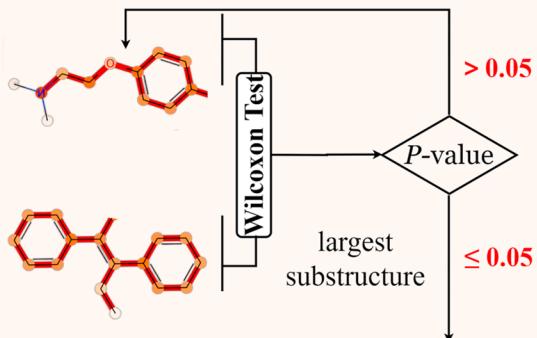


Figure 1. Overall framework of GeoDILI. The upper section shows the DILI prediction module, comprising two primary phases: the feature extraction phase and the fine-tuning phase. The lower part displays the results of dominant substructure derivation, which is accomplished by calculating atomic contribution scores using information from the last GIN layer within the GeoGNN block in conjunction with the fragmentation and Wilcoxon statistical test processes.

proprietary and postmarketing human toxicity data, and (3) the noncausal relationships in inferred hepatotoxicity since people take more than one drug or supplement at the same time.¹⁶ These problems are reflected in the conflicting classification labels of compounds between data sets from different sources. To address this issue, the U.S. Food and Drug Administration (FDA) developed an annotation scheme to label DILI risks for 1036 FDA-approved drugs based on the assessments by regulatory professionals and released the DILIRank public database in 2016.¹⁷ By far, DILIRank is the most widely used data set for developing DILI prediction models.¹⁸⁻²⁴ Recently, the FDA further augmented DILIRank

to DILIst, which contains 1279 drugs, by adding four additional literature data sets and applying consistency analysis.²⁵ Many other publications also annotated drugs of DILI risk based on different criteria, such as LiverTox,²⁶ Hepatox (<http://www.hepatox.org/>), LTKB,²⁷ SIDER,²⁸ and published works of literature.^{18,29–33} These databases provide a valuable resource for DILI prediction. Based on this, numerous in silico models have been generated in the past decade,^{34–42} and it has been shown that the quality of the data sets and the accuracy of the models are correlated.^{16,22}

Many of the existing DILI risk prediction models are based on molecular structures or properties of drugs, named

quantitative structure–activity relationship (QSAR) models.^{43–46} The advantage of QSAR models is that they do not require mechanistic information and could directly establish statistical relationships between structures or descriptors and biological activities. Several data-driven machine learning (ML)-based models have been proposed, such as Random Forest,^{16,22} and Support Vector Machine (SVM).^{34,47} Some works employed ensemble models to improve the accuracy and precision of complex DILI predictions, such as voting or average probability and neural-network-based meta-classifier strategy.^{36,37} These developed DILI models are mainly based on hand-coded or rule-based molecular descriptors/fingerprints or a combination of them to characterize the molecular properties of each compound, which may lead to poor performance if the molecular representations do not capture enough information.^{48,49}

Recent progress in deep learning algorithms with novel neural architectures has greatly facilitated drug discovery and development. A promising approach is to represent molecules as graphs, with atoms as nodes and bonds as edges. Graph neural networks (GNN) can then process these graphs to extract features, with message-passing neural networks (MPNN) being the most popular architecture.^{50–53} It has been shown to outperform models built on human-designed molecular descriptors for certain biological properties but is still rarely applied in the DILI prediction area.^{54,55} In addition, the geometric information on a molecule plays an important role in its physicochemical properties and biological activities. For example, (*R*)-(+)thalidomide and (*S*)-(−)-thalidomide have the same topology structure but different geometries leading to different biological activities. Both (*R*)-(+)thalidomide and (*S*)-(−)-thalidomide have sedative and antiemetic effects, while only (*S*)-(−)-thalidomide is teratogenic and can cause embryonic malformations.⁵⁶ However, since MPNN only considers topological information such as nodes and edges and node adjacencies, it cannot distinguish molecules with different geometric structures. One popular research topic is geometric molecular representation, which encodes angular or three-dimensional (3D) coordinate features that allows GNNs to capture geometric shapes.⁵⁷ Recently, Fang et al. proposed a novel geometry-enhanced molecular (GEM) representation learning method, which is pretrained by large-scale unlabeled molecules with coarse 3D spatial structures using self-supervised learning, and achieved several state-of-the-art (SOTA) results on molecular property prediction benchmarks.⁵⁸

In addition to predictive performance, we also focus on the inferential ability of the model to derive key chemical substructures as structural alerts (SAs) for human liver toxicity. Using SAs, researchers can recognize potential hazardous compounds and modify them in very early stages. Explainable artificial intelligence (XAI) has become a very attractive research subject in theoretical ML, computer vision, natural language processing, and, more recently, cheminformatics.^{64–67} Compared with previously proposed SAs inference methods, such as expert knowledge-based^{59,68} or frequency-based,⁶¹ XAI-based methods infer substructures based on predictive accuracy and therefore have better predictive performance.^{63,67} Therefore, the construction of XAI-based SA inferring methods may be more effective for guiding lead optimization to reduce the risk of DILI. To date, these methods have not yet been applied to this issue yet.

In this study, we developed a highly accurate and interpretable human DILI prediction model named GeoDILI. An overview of the proposed model is shown in Figure 1. The GeoDILI model used a pretrained 3D spatial structure-based GNN to extract molecular representations, followed by a residual neural network to make an accurate DILI prediction. The gradient information from the final graph convolutional layer of GNN was utilized to obtain atom-based weights, which enabled the identification of dominant substructures that significantly contributed to the DILI prediction. We evaluated the performance of GeoDILI by comparing it with the SOTA DILI prediction tools, popular GNN models, as well as conventional deep neural networks (DNN) and ML models, confirming its effectiveness in predicting DILI. In addition, we applied our model to three different human DILI data sets from various sources, namely DILrank,¹⁷ DIList,²⁵ and a data set recently collected by Yan et al.³⁷ Results showed performance differences across data sets and suggested that a smaller, high-quality data set DILrank may lead to better results. Finally, we applied the dominant substructure inference method to analyze the entire DILrank data set and identified seven significant SAs with both high precision and potential mechanisms.

2. MATERIALS AND METHODS

2.1. Data Collection and Preparation. In this work, we used three human DILI data sets, named DILrank, DIList, and Yan et al. data set.^{17,25,37} A summary of these data sets is shown in Table 1.

Table 1. Summary of the DILI Datasets

data set	no. of compound (positive/negative)	classification rule
DILrank	720 (452/268)	^a Most- and ^b Less-DILI-Concern as positive; ^c No-DILI-Concern as negative
DIList	1002 (604/398)	1 as positive; 0 as negative
Yan et al.	2931 (1498/1433)	authors definition

The FDA-curated DILrank data set (2016) categorized 1036 FDA-approved drugs into four groups: “^aMost-,” “^bLess-,” and “^cNo-DILI-concern” with clear causal evidence of liver injury, and “Ambiguous-DILI-concern” without a clear causal relationship.¹⁷ In this study, drugs categorized as “^aMost-DILI concern” or “^bLess-DILI concern” were defined as “DILI positive,” while those categorized as “^cNo-DILI concern” were defined as “DILI negative.” The drugs categorized as “Ambiguous DILI Concern” were excluded. After checking for structural validity referenced to PubChem database and discarding unidentified compounds, the resulting data set contained 720 drugs, including 452 “DILI positive” and 268 “DILI negative”.

The DIList data set was an expanded set of the DILrank data set consisted of 1279 drugs with DILI binary classification, created by the FDA in 2020.²⁵ DIList was established by sequentially merging the DILrank data set (without the terminology of “Ambiguous DILI-concern” drugs) with four other data sets containing more than 350 drugs with the human DILI classification. The four data sets were established based on different approaches, including a clinical evidence-based approach (LiverTox),²⁶ a literature-based approach (Greene),²⁹ a case registry-based approach (Suzuki),³⁰ and an approach based on curating data from the FDA Adverse Event Reporting System (Zhu).³¹ After removing biologics, mixtures, and inorganics, the final DIList data set included 1002 drugs, of which 604 are “DILI positive” and 398 are “DILI negative.”

The Yan et al. data set was a large, comprehensive data set curated in 2022.³⁷ It was collected from five literature data sets (Greene,²⁹ Xu,³² Mulliner,³³ Shuaibing,¹⁸ and LiverTox²⁶) and two public databases (LTKB²⁷ and DILrank¹⁷). The SMILES of all compounds

was first converted to canonical SMILES format, followed by the removal of duplicate drugs, mixtures, and inorganic compounds. The DILI label for the two public databases was retained, and the DILI label for the remaining data sets was determined by the voting rules. The rules were described as follows: if the label of a drug is consistent in 80% or all data sets, the label of the drug will be retained; otherwise, the drug will be deleted. After the above procedures, the Yan et al. data set contained 2931 drugs, with 1498 classified as "DILI positive" and 1433 as "DILI negative."

2.2. Framework of GeoDILI. **2.2.1. Overview.** GeoDILI consists of two main components: a GNN-based DILI prediction model and a dominant substructure identification model. GeoDILI takes a molecular canonical SMILES as input and encodes it into a 32-dim vector using a fine-tuned geometry-based GNN (GeoGNN) model.⁵⁸ The encoded vector is then fed into a residual network (ResNet) for binary DILI classification. To identify DILI-related dominant substructures, a gradient-based algorithm is used to calculate the contribution score of each atom, and the larger substructures composed of atoms with significantly high atomic contribution scores are identified as dominant substructures.⁶⁶ The overall aim of GeoDILI is to accurately predict DILI and identify the key substructures that contribute to DILI, which can help in the design of safer drugs.

2.2.2. DILI Prediction. A molecule is made up of atoms and bonds connecting them, intuitively depicted as an atom-bond graph represented as $G = (\mathcal{V}, \mathcal{E})$ by taking the atoms $i \in \mathcal{V}$ as the nodes of G and the bonds $(i,j) \in \mathcal{E}$ connecting atom i and atom j as the edges of G . Considering the vital role of molecular geometric information in toxicity prediction tasks, we used a pretrained molecular geometric representation model named GeoGNN, that incorporates angular information.⁵⁸ A bond-angle graph H is introduced in the GeoGNN model, which is similar to the definition of G . The bond-angle graph $H = (\mathcal{E}, \mathcal{A})$ is defined by regarding the bonds $(j,i) \in \mathcal{E}$ as the nodes of H and the angles $(j, i, k) \in \mathcal{A}$ connecting bond (j,i) and bond (i,k) as the edges of H . The initial features of atoms, bonds, and angles were calculated according to the items listed in **Table S1**.

In GeoGNN, two blocks are used to learn and update features based on the angle-bond-atom order of the bond-angle graph H and atom-bond graph G , respectively. For the bond-angle graph H , GeoGNN aggregates messages from all of the bonds adjacent to bond (j,i) and the corresponding bond angles to learn and update the representation vector of bond (j,i) . At the t th iteration, the aggregated messages $\mathbf{a}_{ji}^{(t)}$ and hidden representation vector $\mathbf{h}_{ji}^{(t)}$ on a given bond (j,i) are expressed by

$$\begin{aligned} \mathbf{a}_{ji}^{(t)} &= A_{\text{bond-angle}}^{(t)}(\{\mathbf{h}_{ji}^{(t-1)}, \mathbf{h}_{jk}^{(t-1)}, \mathbf{x}_{kji}\}, \forall k: k \in \mathcal{N}(j)\} \\ &\cup \{\mathbf{h}_{ji}^{(t-1)}, \mathbf{h}_{ik}^{(t-1)}, \mathbf{x}_{jik}\}, \forall k: k \in \mathcal{N}(i)\} \end{aligned} \quad (1)$$

$$\mathbf{h}_{ji}^{(t)} = U_{\text{bond-angle}}^{(t)}(\mathbf{h}_{ji}^{(t-1)}, \mathbf{a}_{ji}^{(t)}) \quad (2)$$

where $\mathcal{N}(j)$ and $\mathcal{N}(i)$ are the sets of atoms adjacent to atom j and atom i , respectively, $\{(j, k), \forall k: k \in \mathcal{N}(j)\} \cup \{(i, k), \forall k: k \in \mathcal{N}(i)\}$ is the set of bonds adjacent to bond (j,i) . $A_{\text{bond-angle}}$ is the function for aggregating messages, while $U_{\text{bond-angle}}$ is the function for updating the representation vectors of the bonds.

For the atom-bond graph G , GeoGNN takes the updated representation vectors $\mathbf{h}_{ji}^{(t)}$ of bond (j,i) from H as the features for bonds in G . The aggregated messages $\mathbf{a}_j^{(t)}$ and hidden representation vector $\mathbf{h}_j^{(t)}$ of atom j for the t th iteration can be represented using the following formula

$$\mathbf{a}_j^{(t)} = A_{\text{atom-bond}}^{(t)}(\mathbf{h}_j^{(t-1)}, \mathbf{h}_i^{(t-1)}, \mathbf{h}_{ji}^{(t-1)}), \forall j: j \in \mathcal{N}(i) \quad (3)$$

$$\mathbf{h}_j^{(t)} = U_{\text{atom-bond}}^{(t)}(\mathbf{h}_j^{(t-1)}, \mathbf{a}_j^{(t)}) \quad (4)$$

where $\mathcal{N}(i)$ is the set of atoms adjacent to atom i , $A_{\text{atom-bond}}$ is the aggregate function in the atom-bond graph G , while $U_{\text{atom-bond}}$ is the update function.

At the last iteration, GeoGNN uses an average pooling function to aggregate the atoms' representation vectors to obtain the molecular representation vector \mathbf{h}_G , which is expressed by

$$\mathbf{h}_G = R(\mathbf{h}_j^{(T)}), \forall j: j \in \mathcal{V} \quad (5)$$

where T is the total number of iterations. R is the average pooling function, which can be a complex nonlinear function. The resulting 32-dim representation vector \mathbf{h}_G captures the molecular information and is used as an input for the downstream ResNet.

A fully connected neural network (FCNN) is susceptible to the well-known issue of vanishing/exploding gradients as its depth increases, which can hinder model convergence and lead to degradation problems even when convergence is achieved. To solve the problem, ResNet introduces a shortcut connection, enabling smoother information flow within the network and speeding up convergence. Our ResNet consists of a fully connected layer, four residual blocks, and a final fully connected layer. Each residual block contains two fully connected layers with a specified dropout rate and an identity mapping shortcut connection between the input and output layers. The output of ResNet is used to predict the probability of a compound causing DILI. The overall framework of the DILI prediction model is shown at the top of **Figure 1**.

2.2.3. Dominant Substructures Derivation. As the graph isomorphism network (GIN) convolution layer naturally preserves geometries that are lost in fully connected layers, the last GIN convolution layer of the GeoGNN block can provide an optimal balance between high-dimensional semantics and molecular geometry information.⁶⁶ Specifically, the output of the final GIN convolution layer is represented by $L \in \mathbb{R}^{m \times n}$, where m is the number of atoms in the compound and n is the number of channels. For a given compound A, the importance of each channel can be calculated as

$$\alpha_k = \frac{1}{m} \sum_{i=1}^m \frac{\partial P}{\partial L_A^{i,k}} \quad (6)$$

where $L_A^{i,k}$ denotes a neuron at the k th channel and atom i of compound A, and P represents the probability of compound A causing DILI.

Then, we determined the atomic contribution scores vector $W_A \in \mathbb{R}^m$ of compound A by the weighted combination, which can be expressed as

$$W_A = \sum_{k=1}^n \alpha_k L_A^k \quad (7)$$

finally, we used min–max normalization to map W_A ranging from 0 to 1.

After the contribution scores of each atom in the molecule were obtained, the dominant substructures are extracted. More concretely, during the fragmentation step, each bond in the molecule (excluding ring bonds) is broken to generate two fragments. After iterations using a recursive algorithm, all possible substructures with 4–18 atoms are collected.^{67,69} Then, the one-sided Wilcoxon test was conducted to determine if a given substructure exhibits significantly higher atom attention weights than the remaining part of the molecule, with a p -value threshold set at 0.05.⁷⁰ Finally, only the largest substructures in each molecule were kept to eliminate redundancy.

The framework of dominant substructure derivation is depicted at the bottom of **Figure 1**.

2.3. Model Construction and Optimization. The training and test process of GeoDILI are shown in Algorithm 1.

Algorithm 1 GeoDILI

Require: Pre-training parameter w_{pre} ; embedding feature matrix X ; graph $\mathcal{G} = (G, H)$; ascent step M ; ascent step size α ; training epoch N ; GeoGNN model $u_w(\cdot)$; ResNet model $d_\theta(\cdot)$; forward function $f_{w+\theta}(\cdot) = d_\theta(u_w(\cdot))$; binary cross entropy loss function $L(\cdot)$.

1. Loading the pre-trained parameter $w = w_{pre}$. \triangleright pre-training
2. Initialize θ
3. **for** $epoch = 1 \dots N$ **do**
4. **for** $(\mathcal{G}, y) \in \mathbf{D}^{train}$ **do** \triangleright fine-tuning
5. $\delta_0 \leftarrow U(-\alpha, \alpha)$ \triangleright initialize from uniform distribution
6. $g_0 \leftarrow 0$
7. **for** $t = 1 \dots M$ **do**
8. $g_t \leftarrow g_{t-1} + \frac{1}{M} \cdot \nabla_{w+\theta} L(f_{w+\theta}(X + \delta_{t-1}; \mathcal{G}), y)$ $\triangleright \theta$ gradient accumulation
9. $g_\delta \leftarrow -\nabla_\delta L(f_{w+\theta}(X + \delta_{t-1}; \mathcal{G}), y)$ \triangleright perturbation δ gradient ascent
10. $\delta_t \leftarrow \delta_{t-1} + \alpha \cdot g_\delta / \|g_\delta\|_F$
11. **end for**
12. $w + \theta \leftarrow (w + \theta) - \tau \cdot g_M$ \triangleright model parameter θ gradient descent
13. **end for**
14. **end for**
15. **for** $(\mathcal{G}, y) \in \mathbf{D}^{test}$ **do** \triangleright test
16. Evaluate($f_{w+\theta}(X; \mathcal{G})$, y)
17. **end for**

2.3.1. Input. As shown in Table 2, the data set was split into a training set $\mathbf{D}^{train} = \{(G, H, y)\}$ and a test set $\mathbf{D}^{test} = \{(G, H, y)\}$ in a

Table 2. Distribution of “DILI-Positive” and “DILI-Negative” Samples in the DILIRank/DILIST/Yan et al. dataset

data set	DILI-class	training	test	total
DILIRank	positive	362	90	452
	negative	214	54	268
	total	576	144	720
DILIST	positive	483	121	604
	negative	318	80	398
	total	801	201	1002
Yan et al.	positive	1198	300	1498
	negative	1146	287	1433
	total	2344	587	2931

ratio of 8:2, where y denotes the label of the corresponding molecule represented by (G, H) . To address the problem of limited sample size in the available data sets, the model parameters were fine-tuned using fivefold cross-validation to maximize the utilization of the data sets for effective model training. Subsequently, we evaluated the model's performance in the test set (Table 2).

2.3.2. Model. The GeoGNN block is denoted as $u_w(\cdot)$, where w represents the set of model parameters. We initialized the parameters of the pretrained self-supervised learning model and then fine-tuned them by the downstream ResNet $d_\theta(\cdot)$.

2.3.3. Optimization. To prevent overfitting, we applied an early stop strategy based on the evaluation results of the training set and validation set. We also employed the Free Large-scale Adversarial Augmentation on Graphs (FLAG) approach, which introduces gradient-based adversarial perturbations to the input node features while keeping the graph structure unchanged.⁷¹ FLAG helps to generalize our model to out-of-distribution samples, reducing overfitting and improving the performance on the test set. Hyperparameters were optimized using grid search, and the final settings are bolded in Table 3.

2.4. Statistics for the Model Evaluation Criteria. Several evaluation metrics were used to evaluate the performance of our model, including the receiver-operating characteristic-area under the curve (AUC), accuracy (ACC), precision, sensitivity, specificity, F1-score, and Matthews correlation coefficient (MCC).^{72–76} AUC quantifies the model's ability to distinguish between “DILI positive”

Table 3. Hyperparameters Settings^a

hyperparameters	values
batch size	128 , 256
GeoGNN learning rate	1×10^{-3} , 1×10^{-4} , 5×10^{-5}
ResNet learning rate	1×10^{-3} , 1×10^{-4} , 5×10^{-5}
dropout	0.2 , 0.35, 0.4
optimizer	Adam, AdamW

^aBest hyperparameters are marked as bold.

and “DILI negative” labels.⁷² ACC represents the overall percentage of correct DILI label predictions.⁷³ Precision is the fraction of correctly predicted positive samples among all predicted positives.⁷⁴ Sensitivity measures the proportion of actual “DILI positive” drugs that were correctly predicted as such.^{74,75} Specificity (true-negative rate) indicates the percentage of drugs with the “DILI negative” label that were correctly predicted as such.⁷⁵ F1-score, which is the weighted harmonic mean of precision and recall, balances both metrics.⁷⁴ MCC, which takes into account all four categories (true positives, false negatives, true negatives, and false positives) of the binary confusion matrix, provides a comprehensive measure of the model's performance.⁷⁶ The mathematical formulas for these evaluation metrics are specified below

$$\text{accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (8)$$

$$\text{precision} = \frac{TP}{(TP + FP)} \quad (9)$$

$$\text{sensitivity} = \text{recall} = \frac{TP}{(TP + FN)} \quad (10)$$

$$\text{specificity} = \frac{TN}{(TN + FP)} \quad (11)$$

$$\text{F1-score} = \frac{2 \times \text{precision} \times \text{recall}}{(\text{precision} + \text{recall})} \quad (12)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (13)$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively.

2.5. Identification of DILI Structural Alerts. We derived dominant substructures for all true positives and -negatives in the DILIRank data set and identified DILI-related significant substructures using Fisher's exact test. The test statistic is the number of compounds with a specific dominant substructure in the set of “DILI positive” compounds, which follows a hypergeometric distribution (Table 4).

Table 4. Contingency Table for a Certain Dominant Substructure T

	DILI positive	DILI negative
with dominant substructure T	M	M'
without dominant substructure T	N	N'

The formula for the p-value can be represented as follows⁷⁷

$$P = \frac{\binom{M+M'}{M} \binom{N+N'}{N}}{\binom{M+N+M'+N'}{M+N}} \quad (14)$$

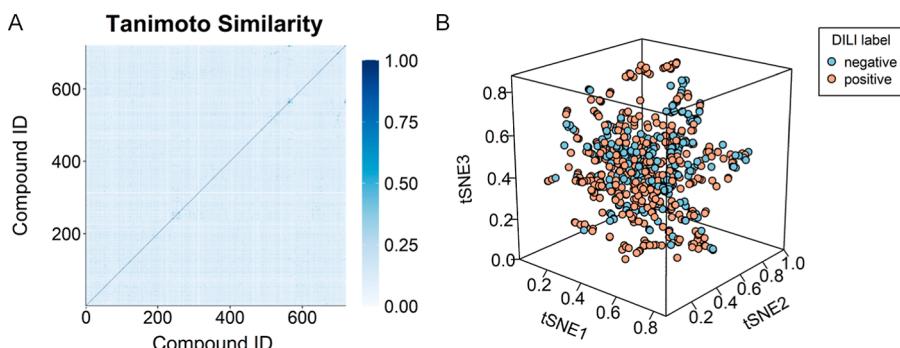


Figure 2. (A) Tanimoto similarity heatmap of the compounds in the DILIRank data set using Morgan fingerprint. (B) The t-SNE distribution of the compounds labeled “DILI positive” and “DILI negative” in the DILIRank data set.

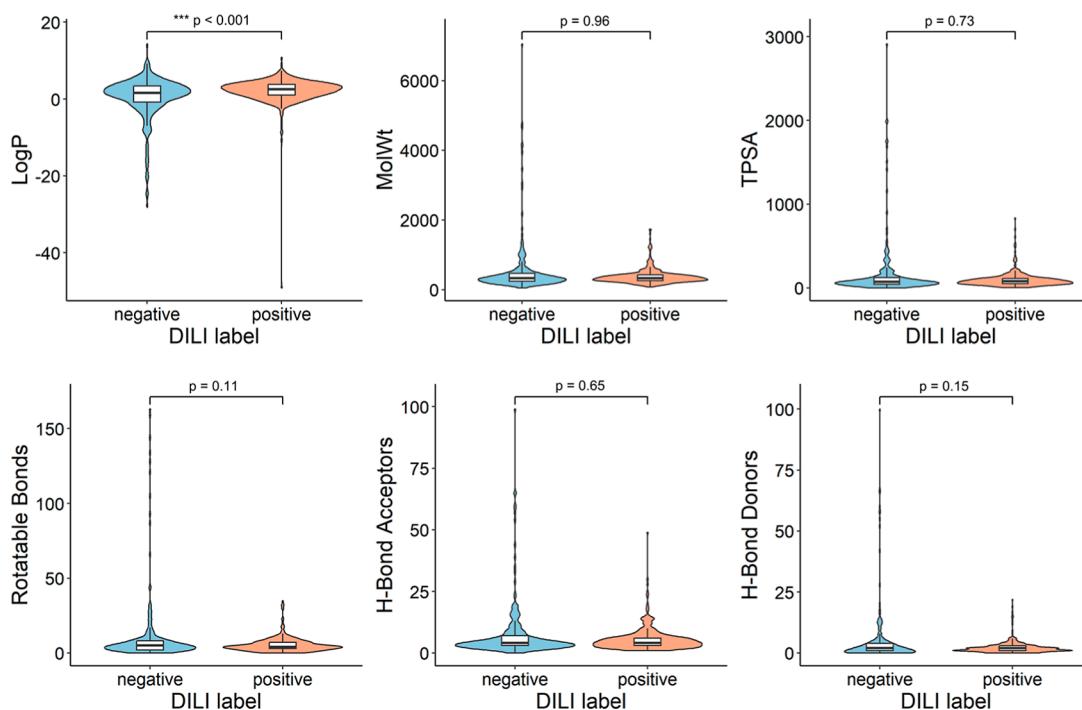


Figure 3. Physicochemical property distributions of compounds with “DILI positive” and “DILI negative” in the DILIRank data set. LogP, the Wildman–Crippen log *P* value;^{83,84} MolWt, the average molecular weight of the molecule;⁸³ TPSA, topological polar surface area;⁸⁵ rotatable bonds, the number of rotatable bonds;⁸³ H-bond acceptors, the number of hydrogen bond acceptors;⁸³ H-bond donors, the number of hydrogen bond donors.⁸³

In addition, enrichment factor (EF) was used to identify true positives compared to a random selection.⁷⁸ Higher EF indicates better enrichment of true positives in the top ranking. The formula of EF is

$$\text{precision} = P(X = 1|T) = \frac{M}{(M + M')} \quad (15)$$

$$\text{EF} = \frac{P(X = 1|T)}{P(X = 1)} = \frac{M/(M + M')}{(M + N)/(M + N + M' + N')} \quad (16)$$

where $P(X = 1|T)$ is the probability of drug X containing substructure T as “DILI positive.”

To obtain substructures with both a high precision and high coverage rate, we employed information gain (IG).⁷⁹ IG measures the difference in information entropy before $H(X)$ and after separation by substructure $H(X|T)$.^{62,80}

$$\text{IG}(T) = H(X) - H(X|T) \quad (17)$$

3. RESULTS

3.1. Data Analysis. To visualize the chemical diversity of compounds in the DILIRank data set, we employed the Tanimoto similarity analysis and the t-distributed stochastic neighbor embedding (t-SNE) algorithm using molecular Morgan fingerprints.^{81,82} The Tanimoto coefficient is a widely utilized metric to measure the chemical structure similarity between two molecules. The Tanimoto similarity heatmap in Figure 2A reveals high compound diversity with an average similarity of 0.105. The t-SNE algorithm is an unsupervised ML method that transforms high-dimensional data into a low-dimensional representation while preserving the relationships between the data points.⁸¹ Figure 2B shows that the distributions of “DILI positive” and “DILI negative” compounds largely overlap, indicating limited differences in their chemical structures.

We further analyzed the distribution of physicochemical properties between “DILI positive” and “DILI negative”

compounds in the DILIRank data set. Results shown in Figure 3 reveal statistically significant differences in LogP values between the two categories, with “DILI positive” compounds being more lipophilic (median log P 2.55 for “DILI positive” vs 1.62 for “DILI negative”, $p < 0.001$). However, no significant differences were observed for the other five molecular descriptors ($p > 0.05$). Similar analyses on the DILIST and Yan et al. data sets are presented in Figures S1 and S2, and Table S2. These findings highlight the challenges in developing accurate and robust DILI prediction models due to the complexity of the DILI toxic end point and the difficulty in distinguishing between “DILI positive” and “DILI negative” compounds solely based on chemical structures and physicochemical properties.

3.2. Model Performance. **3.2.1. Model Construction and Evaluation.** We trained the model for a total of 100 epochs with a batch size of 128. The optimizer used for training was AdamW, which is an extension of the Adam optimizer that incorporates weight decay to avoid overfitting with a learning rate of 1×10^{-4} and a dropout of 0.2. During the training process, the network parameters were updated by minimizing the binary cross-entropy loss function. In the fine-tuning stage, we transferred the parameters from the pretrained model to the current prediction task. In addition, we utilized the FLAG strategy to improve the model’s robustness and reduce overfitting by generating and incorporating adversarial nodes into the graph data.⁷¹ We performed fivefold cross-validation runs with different random seeds to evaluate the reliability of the classification models. The performance of GeoDILI in the DILIRank data set is summarized in Figure 4 and Table 5. For

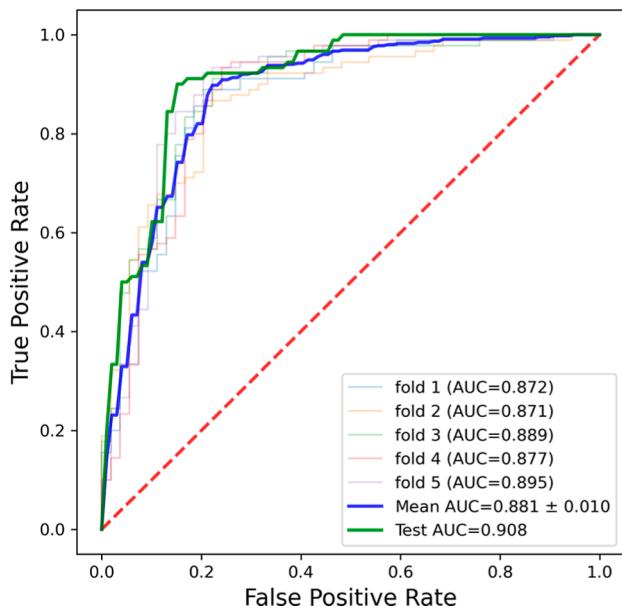


Figure 4. Receiver operating characteristic curves of the fivefold cross-validation and the test set in the DILIRank data set.

the training set given by the 5-fold cross-validation method, the mean and standard deviation of the AUC of the validation set were 0.881 and 0.010, respectively. The final model was retrained throughout the whole training set and further validated with the test set, obtaining a good performance with an AUC value of 0.908.

3.2.2. Comparison with Other Models. Although several DILI prediction models have been published, it is difficult to compare their performance directly as the data sets and source codes are not available.^{38,54} To make a fair comparison of our model with previous DILI models and other popular models, we performed a systematic comparison using the same training and test data.

First, we compared our model GeoDILI with the open-source SOTA DILI models from the past 3 years, of which only three are available: CNN-MFE,³⁵ DeepDILI,³⁶ and R-E-GA (Table 5).³⁷ In comparison with the current DILI prediction models, our model GeoDILI achieved the best result in most evaluation metrics, with accuracy, precision, sensitivity, F1 score, specificity, and MCC of 0.875, 0.860, 0.956, 0.905, 0.875, and 0.732, respectively. Sensitivity, a crucial indicator of the ability to correctly identify hepatotoxicity, is particularly high in our model (0.956). This is important for drug development, as it enables drug developers to conduct further verification to determine the potential for adverse effects from candidate compounds. The well-balanced sensitivity/specification ratio of 0.956:0.875 further demonstrates the efficacy of the GeoDILI model. Additionally, the F1-score and MCC, important metrics for an imbalanced data set, indicate a substantial performance improvement compared to the other evaluated models for the DILI prediction task.

Then, we compared GeoDILI to seven pretrained GNN models, including GEM,⁵⁸ GIN_AttrMasking, GIN_ContextPred, GIN_InfoMax, GIN_EdgePred,⁸⁶ 3D InfoMax and GraphMVP,^{87,88} and four popular GNN models, which are AttentiveFP,⁸⁹ directed message passing neural network (D-MPNN),⁹⁰ graph convolutional network (GCN), graph attention network (GAT). As shown in Table 5, GeoDILI outperforms most metrics among all GNN models (pretrained and popular GNNs) on the DILIRank data set. Among the pretrained and popular GNNs, the pretrained GNNs exhibit superior performance compared to the popular GNNs. Specifically, GeoDILI, GEM,⁵⁸ and GIN_ContextPred⁸⁶ achieve the highest AUCs of 0.908, 0.874, and 0.866, respectively. Correspondingly, their ACCs are 0.875, 0.861, and 0.778, respectively. These results highlight the effectiveness of the pretrained strategy in predicting DILI. Furthermore, most pretrained GNNs exhibited a notable balance between sensitivity and specificity, with a sensitivity/specification ratio close to 1. This finding implies that these pretrained models effectively identified both positive and negative samples in DILI prediction. Moreover, both GeoDILI and GEM outperformed the remaining four pretrained GNNs with significantly higher AUC values (0.908 and 0.874), indicating that they were more effective in predicting DILI by leveraging geometric information. Additionally, GeoDILI exhibits significantly better performance than GEM, which demonstrates the suitability and efficacy of our model.

Finally, we compared our proposed model GeoDILI to eight fingerprint-based ML models,⁹¹ including DNN, k-nearest neighbors (k-NN), SVM, Naïve Bayes, Decision Tree, Random Forest, Gradient Boosting, and AdaBoost. Among the ML models, the Naïve Bayes and Random Forest have the best prediction performance, with AUC and ACC of 0.860, 0.819 for the Naïve Bayes algorithm and 0.849, 0.813 for Random Forest, respectively (Table 5). These models have also been widely used in previous DILI prediction tasks.^{22,39,42} However, most ML models display low MCC due to the imbalance between sensitivity and specificity with higher sensitivity and

Table 5. Comparison of Different Models on the DILrank Dataset^a

model	AUC	ACC	precision	sensitivity	F1-score	specificity	MCC
GeoDILI	0.908	0.875	0.860	0.956	0.905	0.875	0.732
CNN-MFE	0.866	0.854	0.879	0.889	0.884	0.796	0.688
DeepDILI	0.756	0.692	0.766	0.702	0.733	0.675	0.372
R-E-GA	0.882	0.847	0.840	0.933	0.830	0.704	0.670
GEM	0.874	0.861	0.850	0.944	0.895	0.861	0.701
GIN_AttrMasking	0.822	0.819	0.905	0.809	0.854	0.840	0.626
GIN_ContextPred	0.866	0.778	0.780	0.886	0.830	0.607	0.522
GIN_EdgePred	0.798	0.736	0.825	0.733	0.776	0.741	0.462
GIN_InfoMax	0.829	0.764	0.872	0.739	0.800	0.808	0.527
3D InfoMax	0.778	0.723	0.737	0.837	0.784	0.550	0.408
GraphMVP	0.860	0.792	0.816	0.870	0.842	0.654	0.539
AttentiveFP	0.843	0.806	0.833	0.900	0.865	0.591	0.522
D-MPNN	0.821	0.757	0.798	0.820	0.808	0.652	0.480
GCN	0.835	0.778	0.755	0.930	0.833	0.552	0.536
GAT	0.830	0.736	0.731	0.884	0.800	0.517	0.439
DNN	0.764	0.690	0.735	0.727	0.731	0.507	0.334
k-NN	0.795	0.750	0.807	0.789	0.798	0.685	0.471
SVM	0.829	0.799	0.790	0.922	0.851	0.593	0.561
Naïve Bayes	0.860	0.819	0.833	0.889	0.860	0.704	0.609
Decision Tree	0.694	0.688	0.800	0.667	0.727	0.722	0.377
Random Forest	0.849	0.813	0.812	0.911	0.859	0.648	0.592
Gradient Boosting	0.822	0.778	0.774	0.911	0.837	0.556	0.513
AdaBoost	0.785	0.764	0.792	0.844	0.817	0.630	0.487

^aHigher is better. Best results are marked as bold. The models listed in the table have the same training and test sets, as indicated by Table 2.

Table 6. Summary of the Optimal Performance of the Published Models Using DILrank/DIList/Yan et al. dataset^a

author	data set	AUC	ACC	precision	sensitivity	F1-score	specificity	MCC
This work	DILrank	0.908	0.875	0.860	0.956	0.905	0.875	0.732
He et al. ¹⁸		0.859	0.783	/	0.818	/	0.748	/
Wang et al. ¹⁹		0.804	0.817	/	0.646	/	0.962	/
Mora et al. ²⁰		/	0.810	/	0.817	/	0.793	0.566
Ancuceanu et al. ²¹		/	/	/	0.890	/	0.565	/
Liu et al. ²²		0.824	0.763	0.736	0.724	0.730	0.794	0.523
Jaganathan et al. ²³		/	0.811	/	0.840	/	0.783	0.623
Kang, M.-G. and Kang, N. S. ²⁴		/	0.731	/	0.714	/	0.750	/
This work	DILst	0.851	0.786	0.820	0.826	0.823	0.786	0.553
Li et al. ³⁶		0.659	0.687	/	0.805	0.755	0.51	0.331
Lim et al. ³⁵		0.691	0.687	/	/	0.784	/	0.338
This work	Yan et al.	0.843	0.773	0.781	0.846	0.788	0.773	0.549
Yan et al. ³⁷		0.842	0.770	/	/	0.769	/	/

^aHigher is better. Best results are marked as bold. “/” denotes no result for this metric.

lower specificity. For example, SVM has a sensitivity of 0.922 and specificity of 0.593, which suggests that the model is biased toward the positive class. This may be due to the ratio of positive to negative samples in the DILrank data set being nearly 6:4. In summary, GeoDILI not only has good prediction performance but also predicts positive and negative compounds in a more balanced way than the other models.

3.2.3. Comparison of the Three DILI Data Sets. The way data are collected greatly affects the performance of a model, particularly because acquiring human hepatotoxicity data is a challenging task. In this study, we evaluated two FDA-approved DILI data sets (DILrank and DILst) and a recent and larger data set (Yan et al.) by comparing the reported results of the same data sets and systematically comparing the performance of different models on the three data sets.

DILrank is the first DILI benchmark data set released by the FDA in 2016, which is the most widely employed data set

and has been applied by several models.¹⁷ We compared GeoDILI's performance with reported models and found that GeoDILI outperforms other models in most metrics using the same data set. One exception is that Wang's model had a high specificity of 0.962, but its sensitivity/specification ratio was 0.646:0.962, indicating that the model is biased toward the negative class (Table 6).¹⁹ DILst is an expanded data set of DILrank by the FDA.²⁵ We searched for relevant work that employed the DILst data set and compared our results with theirs. As Table 6 demonstrates, GeoDILI outperforms these models in all metrics. The Yan et al. data set is a newly published large data set curated in 2022.³⁷ The results show that our metrics are superior to those of their model, demonstrating the power of GeoDILI.

We compared the performance of GeoDILI with other popular models using the same training and test sets in three data sets (Tables 5, 6, S3 and S4). It is worth noting that the

Table 7. Summary of the Significant SAs^a

No.	SMILES	2D Structure	No. of positive	No. of negative	Precision	EF	IG	<i>p</i> _value
1	C(=O)NCc1ccccc1		18	0	1.000	1.593	0.017	0.00028
2	c1cccc1Cl		71	18	0.798	1.271	0.014	0.00039
3	c2ccc(S(=O)(=O)N)cc2		25	2	0.926	1.475	0.013	0.00078
4	c1ccc(N)c1		100	32	0.758	1.207	0.012	0.00067
5	NC1C(=O)N2C(=CCSC12)		12	0	1.000	1.593	0.011	0.00497
6	CC(NN)		12	0	1.000	1.593	0.011	0.00497
7	CC1=C(C(=O)O)N2C(=O)CC2S1		11	0	1.000	1.593	0.010	0.00890

^aNo. of positive, the number of “DILI-positive” compounds containing this SA; no. of negative, the number of “DILI-negative” compounds containing this SA; EF, enrichment factor; IG, information gain.

performance of GeoDILI optimal models varied across the three data sets, with the best performance on DILIrak (AUC 0.908 and ACC 0.875), followed by DIL1st (AUC 0.851 and ACC 0.786), and the worst on the Yan et al. data set (AUC 0.843 and ACC 0.773).

Moreover, to explore the performance differences between the three data sets, we conducted a comparative analysis based on data set overlap, their distribution in chemical space, and the similarity between training and test sets (Figure S4). Unfortunately, these aspects failed to yield a reasonable explanation. We hypothesized that the performance disparities may be attributed to inherent data characteristics, potentially related to the data source and quality. Consequently, it is crucial to use a uniform benchmark data set in the DILI prediction tasks.

3.3. Structural Alerts Analysis. We performed a substructure inference on all true-positive and true-negative samples in the DILIrak data set and used statistical metrics such as EF, *p*-value, and IG to assess the quality of SAs and identify DILI-related significant substructures. We set thresholds of EF > 1, *p*-value < 0.05, and IG > 0.001 to filter out insignificant substructures and ranked the remaining SAs based on their IG values. Seven significant SAs were identified and are listed in Table 7.

N-benzylformamide moiety (no. 1) is present in 18 compounds, all exhibiting severe or mild hepatotoxicity (Table S4). The containing amide moiety is present as a lactam structure in eight structures, including the antiepileptic drug phenytoin.⁹² These 18 drugs have diverse structures with molecular weights ranging from 252 to 1449 and often coexist with other SAs, such as the chlorobenzene and sulfanilamide moiety in the withdrawn drug chlormezanone and the hydrazine moiety in nialamide. Chlorobenzene moiety (no. 2) is known to be metabolized in the liver by cytochrome P450 enzymes to form epoxides, which can bind to proteins, DNA, and RNA, contributing to its toxicity.⁹³ Induction of the cytochrome P450 system can increase the rate of formation of these epoxides and thus the toxic effects, especially on the liver.⁹³ Exposure to high levels of chlorobenzene has been shown to cause liver damage as well as other adverse effects such as kidney damage and neurological symptoms. Sulfonamide moiety (no. 3) is known to cause idiosyncratic liver injury that exhibits features of drug allergy or hypersensitivity. They have been associated with cases of acute liver failure and remain among the top 5–10 causes of drug-induced, idiosyncratic fulminant hepatic failure.⁹⁴ The aniline moiety (no. 4) is commonly associated with severe toxicities.⁹⁵ It is present in 100 hepatotoxic compounds and 32

nonhepatotoxic compounds according to DILIrank, giving it a relatively higher coverage (18.3%) but lower precision (0.758). 2-Azetidinone moiety is a part of the derived substructure of no. 5 and no. 7, which is a common structural feature of many broad-spectrum β -lactam antibiotics. These antibiotics have been associated with minor liver injuries.⁹⁶ Hydrazine compound (no. 6) is a known human carcinogen that can also cause hepatic necrosis leading to acute liver failure.^{97,98}

4. DISCUSSION

The development of novel artificial intelligence approaches based on publicly available large-scale toxicity data is urgently needed to generate accurate predictive models for chemical toxicity evaluation in the early stages of drug development. However, the lack of internal consistency in publicly available data presents a significant challenge to building accurate predictive models. In this study, we systematically compared two FDA-published DILI-standardized data sets with a recently published large data set and found that the optimal performance of the model differs across data sets. Smaller but higher quality data set DILIrank showed the best performance.^{16,22} Due to the lack of a large-scale DILI benchmark data set, reported DILI prediction models were built on different data sets, making it difficult to compare their performance.^{34–38} Therefore, establishing a DILI benchmark data set and collecting more high-quality DILI data are imperative.

Besides the data set, the way molecules are characterized plays a decisive role in the prediction results. Most DILI prediction models are built based on molecular descriptors, molecular fingerprints, or a combination of both.^{19,20,36,37} In recent years, many pretrained molecular representation learning methods have been proposed, such as MolMap,⁹⁹ GEM,⁵⁸ ImageMol,¹⁰⁰ and so on. These models are trained on large-scale unlabeled molecules in a self-supervised methodology to extract low-dimensional features from the input data. As shown in this study, due to the high structural diversity and small size of the DILI data set, using predefined fingerprints and simple graph representations may not fully capture molecular features and lead to poor prediction performance. By employing a pretraining scheme and fine-tuning on a small number of labeled molecules, significant improvements can be achieved in DILI prediction models.

SAs are commonly used in toxicology to quickly identify potentially toxic molecules, such as those with genotoxicity and mutagenicity, endocrine disruptors, skin sensitivity, hepatotoxicity, and more.¹⁰¹ A recent review discussed that the methods for investigating SAs can be divided into three categories: expert systems, frequency analysis, and interpretable ML models, each with its own advantages and disadvantages.⁶² Expert systems focus on the mechanics of SAs and often result in false positives.^{59,68} In contrast, frequency-based and ML methods have been shown to outperform expert inference-based methods in terms of prediction accuracy.^{61,63} The frequency-based approaches are commonly used methods that involve statistical methods, such as *p*-value, precision, and enrichment factor to assess substructures occurring more frequently in toxic compounds than in nontoxic ones.⁶⁹ Interpretable neural networks represent a promising new direction due to their ability to identify SAs by optimizing complex neural network parameters to achieve high predictive performance while providing structural information about how it is predicted.^{66,67} In this study, we utilized an attention-free

GNN interpretation method to obtain atom-level weights for DILI prediction and then extracted dominant substructures.⁶⁶ We applied this method to DILIrank to obtain seven high-precision SAs. Their potential toxic mechanisms were discussed in the results section, demonstrating that the obtained SAs with both high precision and mechanistic interpretability.

QSAR models can rapidly predict large amounts of new compounds and prioritize toxic ones, but they have limitations in practicality and discrimination.^{43–46,48,49} Classic QSAR models for hepatotoxicity only use structure information and oversimplify the problem into a binary classification, missing out relevant toxicity assays data such as dose–response relationships and assays that evaluate relevant mechanisms such as oxidative stress and mitochondrial reductive activity.¹⁰³ Consequently, these models are unable to fully utilize the information for *in vitro* to *in vivo* extrapolation and accurately differentiate between similar compounds with different toxicities. Future models should incorporate more information, such as *in vitro* activity data and target information, to establish more robust and reliable toxicity prediction models.¹⁰²

5. CONCLUSIONS

In this work, we proposed a robust, interpretable end-to-end DILI predictor GeoDILI for safe drug development. We systematically compared our model with other models using the same data set and demonstrated its superior performance. We utilized a molecular geometric representation strategy that includes angle information to provide more accurate predictions for stereoisomeric molecules. Additionally, our model was capable of deducing dominant substructures, which can provide suggestions for molecular structure optimization.

■ ASSOCIATED CONTENT

Data Availability Statement

All data involved in this study and the source code of GeoDILI are available at <https://github.com/CSU-QJY/GeoDILI>.

■ Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.chemrestox.3c00199>.

Similarity heat maps and t-SNE plots of data sets, physicochemical property analysis, receiver operating characteristic curves, comparison of the three DILI data sets, and tables with initial features, statistics for physicochemical properties, additional model metrics, and drugs containing no. 1 SA ([PDF](#))

Model training and test sets ([ZIP](#))

■ AUTHOR INFORMATION

Corresponding Authors

Xiaoqing Guan — Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China;  orcid.org/0000-0002-4755-820X; Email: guanxq@shutcm.edu.cn

Qingping Zhou — School of Mathematics and Statistics, Central South University, Changsha, Hunan 410083, China;  orcid.org/0000-0003-1530-3525; Email: qpzhou@csu.edu.cn

Authors

Wenxuan Wu — Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional

Chinese Medicine, Shanghai 201203, China;  orcid.org/0000-0002-5371-1186

Jiayu Qian – School of Mathematics and Statistics, Central South University, Changsha, Hunan 410083, China;  orcid.org/0009-0004-9929-3444

Changjie Liang – Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

Jingya Yang – School of Mathematics and Statistics, Central South University, Changsha, Hunan 410083, China

Guangbo Ge – Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China;  orcid.org/0000-0002-9670-4349

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.chemrestox.3c00199>

Author Contributions

[§]W.W. and J.Q. contributed equally to this paper. CRediT: **Wenxuan Wu** data curation, formal analysis, investigation, methodology, project administration, software, validation, writing-original draft; **Jiayu Qian** formal analysis, investigation, methodology, software, visualization, writing-original draft; **Changjie Liang** formal analysis, investigation, software, visualization, writing-original draft; **Jingya Yang** formal analysis, investigation, methodology, software; **Guangbo Ge** conceptualization, funding acquisition, resources; **Qingping Zhou** conceptualization, funding acquisition, methodology, supervision, writing-review & editing; **Xiaoqing Guan** conceptualization, funding acquisition, methodology, project administration, resources, supervision, writing-review & editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (grant nos 82003847, 81922070, 82273897 and 12101614), the Natural Science Foundation of Hunan Province, China (grant no. 2021JJ40715).

REFERENCES

- (1) Shehu, A. I.; Ma, X.; Venkataraman, R. Mechanisms of Drug-Induced Hepatotoxicity. *Clin. Liver Dis.* **2017**, *21* (1), 35–54.
- (2) Navarro, V. J.; Khan, I.; Björnsson, E.; Seeff, L. B.; Serrano, J.; Hoofnagle, J. H. Liver Injury from Herbal and Dietary Supplements. *Hepatology* **2017**, *65* (1), 363–373.
- (3) Tujios, S.; Fontana, R. J. Mechanisms of Drug-Induced Liver Injury: From Bedside to Bench. *Nat. Rev. Gastroenterol. Hepatol.* **2011**, *8* (4), 202–211.
- (4) Weaver, R. J.; Blomme, E. A.; Chadwick, A. E.; Copple, I. M.; Gerets, H. H. J.; Goldring, C. E.; Guillouzo, A.; Hewitt, P. G.; Ingelman-Sundberg, M.; Jensen, K. G.; Juhila, S.; Klingmüller, U.; Labbe, G.; Liguori, M. J.; Lovatt, C. A.; Morgan, P.; Naisbitt, D. J.; Pieters, R. H. H.; Snoeys, J.; van de Water, B.; Williams, D. P.; Park, B. K. Managing the Challenge of Drug-Induced Liver Injury: A Roadmap for the Development and Deployment of Preclinical Predictive Models. *Nat. Rev. Drug Discovery* **2020**, *19* (2), 131–148.
- (5) Dowden, H.; Munro, J. Trends in Clinical Success Rates and Therapeutic Focus. *Nat. Rev. Drug Discovery* **2019**, *18* (7), 495–496.
- (6) Isella, C. J.; Johnson, H. J.; Dunn, M. A. Adverse Drug Reactions: Type A (Intrinsic) or Type B (Idiosyncratic). *Clin. Liver Dis.* **2017**, *21* (1), 73–87.
- (7) Hoofnagle, J. H.; Björnsson, E. S. Drug-Induced Liver Injury — Types and Phenotypes. *N. Engl. J. Med.* **2019**, *381* (3), 264–273.
- (8) Mosedale, M.; Watkins, P. Drug-Induced Liver Injury: Advances in Mechanistic Understanding That Will Inform Risk Management. *Clin. Pharmacol. Ther.* **2017**, *101* (4), 469–480.
- (9) Dara, L.; Liu, Z.-X.; Kaplowitz, N. Mechanisms of Adaptation and Progression in Idiosyncratic Drug Induced Liver Injury, Clinical Implications. *Liver Int.* **2016**, *36* (2), 158–165.
- (10) Blomme, E. A. G.; Yang, Y.; Waring, J. F. Use of Toxicogenomics to Understand Mechanisms of Drug-Induced Hepatotoxicity during Drug Discovery and Development. *Toxicol. Lett.* **2009**, *186* (1), 22–31.
- (11) Xu, J. J.; Henstock, P. V.; Dunn, M. C.; Smith, A. R.; Chabot, J. R.; de Graaf, D. Cellular Imaging Predictions of Clinical Drug-Induced Liver Injury. *Toxicol. Sci.* **2008**, *105* (1), 97–105.
- (12) Elferink, M.; Olinga, P.; Draisma, A.; Merema, M.; Bauerschmidt, S.; Polman, J.; Schoonen, W.; Groothuis, G. Microarray Analysis in Rat Liver Slices Correctly Predicts in Vivo Hepatotoxicity. *Toxicol. Appl. Pharmacol.* **2008**, *229* (3), 300–309.
- (13) McGill, M. R.; Jaeschke, H. Animal Models of Drug-Induced Liver Injury. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2019**, *1865* (5), 1031–1039.
- (14) O'Brien, P. J.; Irwin, W.; Diaz, D.; Howard-Cofield, E.; Krejsa, C. M.; Slaughter, M. R.; Gao, B.; Kaludercic, N.; Angeline, A.; Bernardi, P.; Brain, P.; Hougham, C. High Concordance of Drug-Induced Human Hepatotoxicity with in Vitro Cytotoxicity Measured in a Novel Cell-Based Model Using High Content Screening. *Arch. Toxicol.* **2006**, *80* (9), 580–604.
- (15) Olson, H.; Betton, G.; Robinson, D.; Thomas, K.; Monro, A.; Kolaja, G.; Lilly, P.; Sanders, J.; Sipes, G.; Bracken, W.; Dorato, M.; Van Deun, K.; Smith, P.; Berger, B.; Heller, A. Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals. *Regul. Toxicol. Pharmacol.* **2000**, *32* (1), 56–67.
- (16) Kotsampasakou, E.; Montanari, F.; Ecker, G. F. Predicting Drug-Induced Liver Injury: The Importance of Data Curation. *Toxicology* **2017**, *389*, 139–145.
- (17) Chen, M.; Suzuki, A.; Thakkar, S.; Yu, K.; Hu, C.; Tong, W. DILIrank: The Largest Reference Drug List Ranked by the Risk for Developing Drug-Induced Liver Injury in Humans. *Drug Discovery Today* **2016**, *21* (4), 648–653.
- (18) He, S.; Ye, T.; Wang, R.; Zhang, C.; Zhang, X.; Sun, G.; Sun, X. An In Silico Model for Predicting Drug-Induced Hepatotoxicity. *Int. J. Mol. Sci.* **2019**, *20* (8), 1897.
- (19) Wang, Y.; Xiao, Q.; Chen, P.; Wang, B. In Silico Prediction of Drug-Induced Liver Injury Based on Ensemble Classifier Method. *Int. J. Mol. Sci.* **2019**, *20* (17), 4106.
- (20) Mora, J. R.; Marrero-Ponce, Y.; García-Jacas, C. R.; Suárez Causado, A. Ensemble Models Based on QuBiLS-MAS Features and Shallow Learning for the Prediction of Drug-Induced Liver Toxicity: Improving Deep Learning and Traditional Approaches. *Chem. Res. Toxicol.* **2020**, *33* (7), 1855–1873.
- (21) Ancuceanu, R.; Hovanet, M. V.; Anghel, A. I.; Furtunescu, F.; Neagu, M.; Constantin, C.; Dinu, M. Computational Models Using Multiple Machine Learning Algorithms for Predicting Drug Hepatotoxicity with the DILIrank Dataset. *Int. J. Mol. Sci.* **2020**, *21* (6), 2114.
- (22) Liu, A.; Walter, M.; Wright, P.; Bartosik, A.; Dolciami, D.; Elbasir, A.; Yang, H.; Bender, A. Prediction and Mechanistic Analysis of Drug-Induced Liver Injury (DILI) Based on Chemical Structure. *Biol. Direct* **2021**, *16* (1), 6.
- (23) Jaganathan, K.; Tayara, H.; Chong, K. T. Prediction of Drug-Induced Liver Toxicity Using SVM and Optimal Descriptor Sets. *Int. J. Mol. Sci.* **2021**, *22* (15), 8073.
- (24) Kang, M.-G.; Kang, N. S. Predictive Model for Drug-Induced Liver Injury Using Deep Neural Networks Based on Substructure Space. *Molecules* **2021**, *26* (24), 7548.
- (25) Thakkar, S.; Li, T.; Liu, Z.; Wu, L.; Roberts, R.; Tong, W. Drug-Induced Liver Injury Severity and Toxicity (DILIST): Binary

- Classification of 1279 Drugs by Human Hepatotoxicity. *Drug Discovery Today* **2020**, *25* (1), 201–208.
- (26) Björnsson, E. S.; Hoofnagle, J. H. Categorization of Drugs Implicated in Causing Liver Injury: Critical Assessment Based on Published Case Reports. *Hepatology* **2016**, *63* (2), 590–603.
- (27) Chen, M.; Vijay, V.; Shi, Q.; Liu, Z.; Fang, H.; Tong, W. FDA-Approved Drug Labeling for the Study of Drug-Induced Liver Injury. *Drug Discovery Today* **2011**, *16* (15–16), 697–703.
- (28) Kuhn, M.; Letunic, I.; Jensen, L. J.; Bork, P. The SIDER Database of Drugs and Side Effects. *Nucleic Acids Res.* **2016**, *44* (D1), D1075–D1079.
- (29) Greene, N.; Fisk, L.; Naven, R. T.; Note, R. R.; Patel, M. L.; Pelletier, D. J. Developing Structure-Activity Relationships for the Prediction of Hepatotoxicity. *Chem. Res. Toxicol.* **2010**, *23* (7), 1215–1222.
- (30) Suzuki, A.; Andrade, R. J.; Björnsson, E.; Lucena, M. I.; Lee, W. M.; Yuen, N. A.; Hunt, C. M.; Freiston, J. W. Drugs Associated with Hepatotoxicity and Their Reporting Frequency of Liver Adverse Events in VigiBase: Unified List Based on International Collaborative Work. *Drug Saf.* **2010**, *33* (6), 503–522.
- (31) Zhu, X.; Kruhlak, N. L. Construction and Analysis of a Human Hepatotoxicity Database Suitable for QSAR Modeling Using Post-Market Safety Data. *Toxicology* **2014**, *321*, 62–72.
- (32) Xu, Y.; Dai, Z.; Chen, F.; Gao, S.; Pei, J.; Lai, L. Deep Learning for Drug-Induced Liver Injury. *J. Chem. Inf. Model.* **2015**, *55* (10), 2085–2093.
- (33) Mulliner, D.; Schmidt, F.; Stolte, M.; Spirk, H.-P.; Czich, A.; Amberg, A. Computational Models for Human and Animal Hepatotoxicity with a Global Application Scope. *Chem. Res. Toxicol.* **2016**, *29* (5), 757–767.
- (34) Tharwat, A.; Moemen, Y. S.; Hassanien, A. E. Classification of Toxicity Effects of Biotransformed Hepatic Drugs Using Whale Optimized Support Vector Machines. *J. Biomed. Inf.* **2017**, *68*, 132–149.
- (35) Nguyen-Vo, T.-H.; Nguyen, L.; Do, N.; Le, P. H.; Nguyen, T.-N.; Nguyen, B. P.; Le, L. Predicting Drug-Induced Liver Injury Using Convolutional Neural Network and Molecular Fingerprint-Embedded Features. *ACS Omega* **2020**, *5* (39), 25432–25439.
- (36) Li, T.; Tong, W.; Roberts, R.; Liu, Z.; Thakkar, S. DeepDILI: Deep Learning-Powered Drug-Induced Liver Injury Prediction Using Model-Level Representation. *Chem. Res. Toxicol.* **2021**, *34* (2), 550–565.
- (37) Yan, B.; Ye, X.; Wang, J.; Han, J.; Wu, L.; He, S.; Liu, K.; Bo, X. An Algorithm Framework for Drug-Induced Liver Injury Prediction Based on Genetic Algorithm and Ensemble Learning. *Molecules* **2022**, *27* (10), 3112.
- (38) Chen, Z.; Jiang, Y.; Zhang, X.; Zheng, R.; Qiu, R.; Sun, Y.; Zhao, C.; Shang, H. ResNet18DNN: Prediction Approach of Drug-Induced Liver Injury by Deep Neural Network with ResNet18. *Briefings Bioinf.* **2022**, *23* (1), bbab503.
- (39) Ai, H.; Chen, W.; Zhang, L.; Huang, L.; Yin, Z.; Hu, H.; Zhao, Q.; Zhao, J.; Liu, H. Predicting Drug-Induced Liver Injury Using Ensemble Learning Methods and Molecular Fingerprints. *Toxicol. Sci.* **2018**, *165* (1), 100–107.
- (40) Zhang, H.; Ding, L.; Zou, Y.; Hu, S.-Q.; Huang, H.-G.; Kong, W.-B.; Zhang, J. Predicting Drug-Induced Liver Injury in Human with Naive Bayes Classifier Approach. *J. Comput. Aided Mol. Des.* **2016**, *30* (10), 889–898.
- (41) Ekins, S.; Williams, A. J.; Xu, J. J. A Predictive Ligand-Based Bayesian Model for Human Drug-Induced Liver Injury. *Drug Metab. Dispos.* **2010**, *38* (12), 2302–2308.
- (42) Williams, D. P.; Lazic, S. E.; Foster, A. J.; Semenova, E.; Morgan, P. Predicting Drug-Induced Liver Injury with Bayesian Machine Learning. *Chem. Res. Toxicol.* **2020**, *33* (1), 239–248.
- (43) Gadaleta, D.; Manganelli, S.; Roncaglioni, A.; Toma, C.; Benfenati, E.; Mombelli, E. QSAR Modeling of ToxCast Assays Relevant to the Molecular Initiating Events of AOPs Leading to Hepatic Steatosis. *J. Chem. Inf. Model.* **2018**, *58* (8), 1501–1517.
- (44) Ambure, P.; Halder, A. K.; González Díaz, H.; Cordeiro, M. N. D. S. QSAR-Co: An Open Source Software for Developing Robust Multitasking or Multitarget Classification-Based QSAR Models. *J. Chem. Inf. Model.* **2019**, *59* (6), 2538–2544.
- (45) Khan, K.; Roy, K.; Benfenati, E. Ecotoxicological QSAR Modeling of Endocrine Disruptor Chemicals. *J. Hazard. Mater.* **2019**, *369*, 707–718.
- (46) Chen, S.; Xue, D.; Chuai, G.; Yang, Q.; Liu, Q. FL-QSAR: A Federated Learning-Based QSAR Prototype for Collaborative Drug Discovery. *Bioinformatics* **2021**, *36* (22–23), 5492–5498.
- (47) Zhang, C.; Cheng, F.; Li, W.; Liu, G.; Lee, P. W.; Tang, Y. In Silico Prediction of Drug Induced Liver Toxicity Using Substructure Pattern Recognition Method. *Mol. Inf.* **2016**, *35* (3–4), 136–144.
- (48) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without Borders. *Chem. Soc. Rev.* **2020**, *49* (11), 3525–3564.
- (49) Johnson, S. R. The Trouble with QSAR (or How I Learned To Stop Worrying and Embrace Fallacy). *J. Chem. Inf. Model.* **2008**, *48* (1), 25–26.
- (50) Jiang, D.; Wu, Z.; Hsieh, C.-Y.; Chen, G.; Liao, B.; Wang, Z.; Shen, C.; Cao, D.; Wu, J.; Hou, T. Could Graph Neural Networks Learn Better Molecular Representation for Drug Discovery? A Comparison Study of Descriptor-Based and Graph-Based Models. *J. Cheminf.* **2021**, *13* (1), 12.
- (51) Wang, Y.; Wang, J.; Cao, Z.; Barati Farimani, A. Molecular Contrastive Learning of Representations via Graph Neural Networks. *Nat. Mach. Intell.* **2022**, *4* (3), 279–287.
- (52) Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph Neural Networks: A Review of Methods and Applications. *AI Open* **2020**, *1*, 57–81.
- (53) Withnall, M.; Lindelöf, E.; Engkvist, O.; Chen, H. Building Attention and Edge Message Passing Neural Networks for Bioactivity and Physical-Chemical Property Prediction. *J. Cheminf.* **2020**, *12* (1), 1.
- (54) Ma, H.; An, W.; Wang, Y.; Sun, H.; Huang, R.; Huang, J. Deep Graph Learning with Property Augmentation for Predicting Drug-Induced Liver Injury. *Chem. Res. Toxicol.* **2021**, *34* (2), 495–506.
- (55) Lim, S.; Kim, Y.; Gu, J.; Lee, S.; Shin, W.; Kim, S. Supervised Chemical Graph Mining Improves Drug-Induced Liver Injury Prediction. *iScience* **2023**, *26* (1), 105677.
- (56) Ma, Z.; Wang, J.; Loskill, P.; Huebsch, N.; Koo, S.; Svedlund, F. L.; Marks, N. C.; Hua, E. W.; Grigoropoulos, C. P.; Conklin, B. R.; Healy, K. E. Self-Organizing Human Cardiac Microchambers Mediated by Geometric Confinement. *Nat. Commun.* **2015**, *6* (1), 7413.
- (57) Atz, K.; Grisoni, F.; Schneider, G. Geometric Deep Learning on Molecular Representations. *Nat. Mach. Intell.* **2021**, *3* (12), 1023–1032.
- (58) Fang, X.; Liu, L.; Lei, J.; He, D.; Zhang, S.; Zhou, J.; Wang, F.; Wu, H.; Wang, H. Geometry-Enhanced Molecular Representation Learning for Property Prediction. *Nat. Mach. Intell.* **2022**, *4* (2), 127–134.
- (59) Hewitt, M.; Enoch, S. J.; Madden, J. C.; Przybylak, K. R.; Cronin, M. T. D. Hepatotoxicity: A Scheme for Generating Chemical Categories for Read-across, Structural Alerts and Insights into Mechanism(s) of Action. *Crit. Rev. Toxicol.* **2013**, *43* (7), 537–558.
- (60) Ahlberg, E.; Carlsson, L.; Boyer, S. Computational Derivation of Structural Alerts from Large Toxicology Data Sets. *J. Chem. Inf. Model.* **2014**, *54* (10), 2945–2952.
- (61) Liu, R.; Yu, X.; Wallqvist, A. Data-Driven Identification of Structural Alerts for Mitigating the Risk of Drug-Induced Human Liver Injuries. *J. Cheminf.* **2015**, *7*, 4.
- (62) Yang, H.; Lou, C.; Li, W.; Liu, G.; Tang, Y. Computational Approaches to Identify Structural Alerts and Their Applications in Environmental Toxicology and Drug Discovery. *Chem. Res. Toxicol.* **2020**, *33* (6), 1312–1322.

- (63) Jia, X.; Wen, X.; Russo, D. P.; Aleksunes, L. M.; Zhu, H. Mechanism-Driven Modeling of Chemical Hepatotoxicity Using Structural Alerts and an in Vitro Screening Assay. *J. Hazard. Mater.* **2022**, *436*, 129193.
- (64) Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug Discovery with Explainable Artificial Intelligence. *Nat. Mach. Intell.* **2020**, *2* (10), 573–584.
- (65) Rodríguez-Pérez, R.; Bajorath, J. Explainable Machine Learning for Property Predictions in Compound Optimization. *J. Med. Chem.* **2021**, *64* (24), 17744–17752.
- (66) Yang, Z.; Zhong, W.; Zhao, L.; Yu-Chian Chen, C. MGraphDTA: Deep Multiscale Graph Neural Network for Explainable Drug-Target Binding Affinity Prediction. *Chem. Sci.* **2022**, *13* (3), 816–833.
- (67) Lou, C.; Yang, H.; Wang, J.; Huang, M.; Li, W.; Liu, G.; Lee, P. W.; Tang, Y. IDL-PPBopt: A Strategy for Prediction and Optimization of Human Plasma Protein Binding of Compounds via an Interpretable Deep Learning Method. *J. Chem. Inf. Model.* **2022**, *62* (11), 2788–2799.
- (68) Sushko, I.; Salmina, E.; Potemkin, V. A.; Poda, G.; Tetko, I. V. ToxAlerts: A Web Server of Structural Alerts for Toxic Chemicals and Compounds with Potential Adverse Reactions. *J. Chem. Inf. Model.* **2012**, *S2* (8), 2310–2316.
- (69) Ferrari, T.; Cattaneo, D.; Gini, G.; Golbamaki Bakhtyari, N.; Manganaro, A.; Benfenati, E. Automatic Knowledge Extraction from Chemical Structures: The Case of Mutagenicity Prediction. *SAR QSAR Environ. Res.* **2013**, *24* (5), 365–383.
- (70) Murakami, H. The Power of the Modified Wilcoxon Rank-Sum Test for the One-Sided Alternative. *Statistics* **2015**, *49* (4), 781–794.
- (71) Kong, K.; Li, G.; Ding, M.; Wu, Z.; Zhu, C.; Ghanem, B.; Taylor, G.; Goldstein, T. Robust Optimization as Data Augmentation for Large-Scale Graphs. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New Orleans, LA, USA, 2022; pp 60–69.
- (72) Lobo, J. M.; Jiménez-Valverde, A.; Real, R. AUC: A Misleading Measure of the Performance of Predictive Distribution Models. *Global Ecol. Biogeogr.* **2008**, *17* (2), 145–151.
- (73) Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A. F.; Nielsen, H. Assessing the Accuracy of Prediction Algorithms for Classification: An Overview. *Bioinformatics* **2000**, *16* (5), 412–424.
- (74) Yacoubi, R.; Axman, D. Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*; Association for Computational Linguistics: Online, 2020; pp 79–91.
- (75) Lalkhen, A. G.; McCluskey, A. Clinical Tests: Sensitivity and Specificity. *Cont. Educ. Anaesth. Crit. Care Pain* **2008**, *8* (6), 221–223.
- (76) Zhu, Q. On the Performance of Matthews Correlation Coefficient (MCC) for Imbalanced Dataset. *Pattern Recognit. Lett.* **2020**, *136*, 71–80.
- (77) Thiese, M. S.; Ronna, B.; Ott, U. P Value Interpretations and Considerations. *J. Thorac. Dis.* **2016**, *8* (9), E928–E931.
- (78) Du, H.; Cai, Y.; Yang, H.; Zhang, H.; Xue, Y.; Liu, G.; Tang, Y.; Li, W. In Silico Prediction of Chemicals Binding to Aromatase with Machine Learning Methods. *Chem. Res. Toxicol.* **2017**, *30* (5), 1209–1218.
- (79) Lee, C.; Lee, G. G. Information Gain and Divergence-Based Feature Selection for Machine Learning-Based Text Categorization. *Inf. Process. Manag.* **2006**, *42* (1), 155–165.
- (80) Yang, H.; Sun, L.; Li, W.; Liu, G.; Tang, Y. Identification of Nontoxic Substructures: A New Strategy to Avoid Potential Toxicity Risk. *Toxicol. Sci.* **2018**, *165* (2), 396–407.
- (81) Arora, S.; Hu, W.; Kothari, P. K. An Analysis of the T-SNE Algorithm for Data Visualization. **2018**, arXiv:1803.01768.
- (82) Capecchi, A.; Probst, D.; Reymond, J.-L. One Molecular Fingerprint to Rule Them All: Drugs, Biomolecules, and the Metabolome. *J. Cheminf.* **2020**, *12* (1), 43.
- (83) Lipinski, C. A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discov. Today Technol.* **2004**, *1* (4), 337–341.
- (84) Koutsoumpas, S.; Chronaki, M.; Tsionos, C.; Karakasidis, T.; Guazzelli, L.; Mezzetta, A.; Moutzouris, K. On the Application of the Wildman-Crippen Model to Ionic Liquids. *Results Mater.* **2022**, *16*, 100350.
- (85) Prasanna, S.; Doerksen, R. Topological Polar Surface Area: A Useful Descriptor in 2D-QSAR. *Curr. Med. Chem.* **2009**, *16* (1), 21–41.
- (86) Li, M.; Zhou, J.; Hu, J.; Fan, W.; Zhang, Y.; Gu, Y.; Karypis, G. DGL-LifeSci: An Open-Source Toolkit for Deep Learning on Graphs in Life Science. *ACS Omega* **2021**, *6* (41), 27233–27238.
- (87) Stärk, H.; Beaini, D.; Corso, G.; Tossou, P.; Dallago, C.; Günemann, S.; Liò, P. 3D Infomax Improves GNNs for Molecular Property Prediction. *Proceedings of the 39th International Conference on Machine Learning*, 2022.
- (88) Liu, S.; Wang, H.; Liu, W.; Lasenby, J.; Guo, H.; Tang, J. Pre-Training Molecular Graph Representation with 3D Geometry. **2022**, arXiv:2110.07728.
- (89) Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H.; Zheng, M. Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. *J. Med. Chem.* **2020**, *63* (16), 8749–8760.
- (90) Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; Palmer, A.; Settels, V.; Jaakkola, T.; Jensen, K.; Barzilay, R. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59* (8), 3370–3388.
- (91) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (92) Vidaurre, J.; Gedela, S.; Yarosz, S. Antiepileptic Drugs and Liver Disease. *Pediatr. Neurol.* **2017**, *77*, 23–36.
- (93) Younis, H. S. The Role of Hepatocellular Oxidative Stress in Kupffer Cell Activation during 1,2-Dichlorobenzene-Induced Hepatotoxicity. *Toxicol. Sci.* **2003**, *76* (1), 201–211.
- (94) Stepan, A. F.; Walker, D. P.; Bauman, J.; Price, D. A.; Baillie, T. A.; Kalgutkar, A. S.; Aleo, M. D. Structural Alert/Reactive Metabolite Concept as Applied in Medicinal Chemistry to Mitigate the Risk of Idiosyncratic Drug Toxicity: A Perspective Based on the Critical Examination of Trends in the Top 200 Drugs Marketed in the United States. *Chem. Res. Toxicol.* **2011**, *24* (9), 1345–1410.
- (95) Wang, Y.; Gao, H.; Na, X.-L.; Dong, S.-Y.; Dong, H.-W.; Yu, J.; Jia, L.; Wu, Y.-H. Aniline Induces Oxidative Stress and Apoptosis of Primary Cultured Hepatocytes. *Int. J. Environ. Res. Publ. Health* **2016**, *13* (12), 1188.
- (96) Paech, F.; Messner, S.; Spickermann, J.; Wind, M.; Schmitt-Hoffmann, A.-H.; Witschi, A. T.; Howell, B. A.; Church, R. J.; Woodhead, J.; Engelhardt, M.; Krähenbühl, S.; Maurer, M. Mechanisms of Hepatotoxicity Associated with the Monocyclic β -Lactam Antibiotic BAL30072. *Arch. Toxicol.* **2017**, *91* (11), 3647–3662.
- (97) Scales, M. D. G.; Timbrell, J. A. Studies on Hydrazine Hepatotoxicity. 1. Pathological Findings. *J. Toxicol. Environ. Health* **1982**, *10* (6), 941–953.
- (98) Timbrell, J. A.; Scales, M. D. C.; Streeter, A. J. Studies on Hydrazine Hepatotoxicity. 2. Biochemical Findings. *J. Toxicol. Environ. Health* **1982**, *10* (6), 955–968.
- (99) Shen, W. X.; Zeng, X.; Zhu, F.; Wang, Y. l.; Qin, C.; Tan, Y.; Jiang, Y. Y.; Chen, Y. Z. Out-of-the-Box Deep Learning Prediction of Pharmaceutical Properties by Broadly Learned Knowledge-Based Molecular Representations. *Nat. Mach. Intell.* **2021**, *3* (4), 334–343.
- (100) Zeng, X.; Xiang, H.; Yu, L.; Wang, J.; Li, K.; Nussinov, R.; Cheng, F. Accurate Prediction of Molecular Properties and Drug Targets Using a Self-Supervised Image Representation Learning Framework. *Nat. Mach. Intell.* **2022**, *4* (11), 1004–1016.
- (101) Erve, J. C. L. Chemical Toxicology: Reactive Intermediates and Their Role in Pharmacology and Toxicology. *Expert Opin. Drug Metab. Toxicol.* **2006**, *2* (6), 923–946.

- (102) Zhao, L.; Russo, D. P.; Wang, W.; Aleksunes, L. M.; Zhu, H. Mechanism-Driven Read-Across of Chemical Hepatotoxins Based on Chemical Structures and Biological Data. *Toxicol. Sci.* **2020**, *174* (2), 178–188.
- (103) Low, Y.; Sedykh, A.; Fourches, D.; Golbraikh, A.; Whelan, M.; Rusyn, I.; Tropsha, A. Integrative Chemical-Biological Read-Across Approach for Chemical Hazard Classification. *Chem. Res. Toxicol.* **2013**, *26* (8), 1199–1208.



CAS BIOFINDER DISCOVERY PLATFORM™

BRIDGE BIOLOGY AND CHEMISTRY FOR FASTER ANSWERS

Analyze target relationships,
compound effects, and disease
pathways

Explore the platform



A division of the
American Chemical Society