Journal of Cheminformatics

DATABASE

Open Access

HepatoToxicity Portal (HTP): an integrated database of drug-induced hepatotoxicity knowledgebase and graph neural network-based prediction model

Jiyeon Han[1†], Wonho Zhung[2†], Insoo Jang[3†], Joongwon Lee[2], Min Ji Kang[4], Timothy Dain Lee[5], Seung Jun Kwack[6], Kyu-Bong Kim[7], Daehee Hwang[5], Byungwook Lee[3], Hyung Sik Kim[8*], Woo Youn Kim[2*] and Sanghyuk Lee[1,4*]

Abstract

Liver toxicity poses a critical challenge in drug development due to the liver's pivotal role in drug metabolism and detoxification. Accurately predicting liver toxicity is crucial but is hindered by scattered information sources, a lack of curation standards, and the heterogeneity of data perspectives. To address these challenges, we developed the HepatoToxicity Portal (HTP), which integrates an expert-curated knowledgebase (HTP-KB) and a state-of-the-art machine learning model for toxicity prediction (HTP-Pred). The HTP-KB consolidates hepatotoxicity data from nine major databases, carefully reviewed by hepatotoxicity experts and categorized into three levels: in vitro, in vivo, and clinical, using the Medical Dictionary for Regulatory Activities (MedDRA) terminology. The knowledgebase includes information on 8,306 chemicals. This curated dataset was used to build a hepatotoxicity prediction module by fine-tuning a GNN-based foundation model, which was pre-trained with approximately 10 million chemicals in the PubChem database. Our model demonstrated excellent performance, achieving an area under the ROC curve (AUROC) of 0.761, surpassing existing methods for hepatotoxicity prediction. The HTP is publicly accessible at https://kobic.re.kr/htp/, offering both curated data and prediction services through an intuitive interface, thus effectively supporting drug development efforts.

Scientific contributions

HTP-KB consolidates comprehensive curated information on liver toxicity gathered from nine sources. HTP-Pred utilizes advanced deep learning techniques, significantly enhancing predictive accuracy. Together, these tools provide valuable resources for researchers and practitioners in drug development, accessible through a user-friendly interface.

[†]Jiyeon Han, Wonho Zhung and Insoo Jang have contributed equally to the work.

*Correspondence:
Hyung Sik Kim
hkims@skku.edu
Woo Youn Kim
wooyoun@kaist.ac.kr
Sanghyuk Lee
sanghyuk@ewha.ac.kr
Full list of author information is available at the end of the article

Keywords Liver toxicity, Drug induced liver injury, Hepatotoxicity, Molecular graph, Fine-tuning, Foundation model, Deep neural network, Graph neural network

Introduction

Drug development is a complex and resource-intensive process with a low success rate of less than 10% in each developmental phase [1, 2]. A significant contributing factor to this high attrition rate is drug toxicity, often exacerbated by discrepancies between animal models and human responses [3–6]. Given the liver's pivotal role in chemical transformation and detoxification, it is particularly susceptible to drug-induced damage. Even after FDA market approval, drugs may have adverse effects such as drug-induced liver injury (DILI), a major cause of acute liver failure cases in U.S. tertiary care centers, accounting for over 50% of instances [7].

The need for comprehensive knowledge bases detailing drug effects on liver tissues has become apparent. The US FDA has made significant efforts to establish knowledge resources of DILI for FDA-approved drugs. The Liver Toxicity Knowledge Base (LTKB) is an umbrella project to develop content-rich resources on liver toxicity [8]. Notably, the DILIrank dataset [9] is the classification of 1,036 FDA-approved drugs into four classes according to their potential for causing DILI, determined by analyzing the hepatotoxic descriptions in the drug labeling documents and assessing causality evidence in literature. Similarly, LiverTox [10] provides clinical and research information on DILI for over 1,400 drugs. These databases are pivotal in hepatotoxicity research, yet their coverage is limited to drugs in the market only.

Experimental data remains crucial as it offers detailed insights into drug effects at cellular and organismal levels. Databases like InvitroDB [11] and CEBS [12] exemplify efforts to catalog chemical effects in biological systems based on drug experiments in cell lines, though translating these findings into clinical insights remains a challenge. Other approaches involve compiling drug experimental results from multiple publications to offer diverse perspectives on drug effects [13–15]. However, the usability of these databases is often hindered by the format of their reference data, typically stored as PDFs or CSVs, complicating data extraction for researchers.

To facilitate access to comprehensive drug data, various web servers have been developed to integrate disparate resources. Examples include CompTox [16], NITE-CHIRP [17], and eChemPortal [18], providing web-based access to toxicity reference databases in the U.S., Japan, and OECD, respectively. However, assessing overall compound toxicity or uncovering hidden biological connections remains challenging, as these platforms often lack additional curation and data visualization features.

Recent studies have focused on developing predictive models for hepatotoxicity using compiled datasets, reflecting diverse biological scenarios. Computational methods offer advantages over traditional in vitro and in vivo experiments in terms of time, coverage, and cost efficiency. Greene et al. introduced a model utilizing ECFP6 fingerprints to classify predefined hepatotoxicity labels [19], paving the way for subsequent algorithmic advancements. Bayesian models [20, 21], support vector machines (SVMs) [22–24], decision trees [25, 26], and random forests [24, 27, 28] have since been widely applied to predict hepatotoxicity, often integrating ensemble methods to enhance predictive performance [29–32].

With the emergence of deep learning methods, convolutional neural network (CNN)-based approaches have also been employed for toxicity predictions. Kang et al. applied deep neural networks to represent fingerprints of chemical compounds for hepatotoxicity prediction [33], while Xu et al. utilized undirected graph recursive neural networks for molecular structure encoding to identify DILI-positive molecules [34]. These approaches demonstrate the potential of deep learning in linking chemical structures and properties with hepatotoxicity outcomes, warranting further exploration of advanced algorithms and methodologies.

Beyond algorithmic research, efforts have been made to provide user-friendly web servers offering both prediction models and toxicity data. PASS Online supports diverse prediction modules trained on literature data with active maintenance [35]. Similarly, LAZAR [36], ProTox3 [37], admetSAR 2.0 [38], and eMolTox [39] provide prediction modules focusing on various aspects of ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity). However, these platforms, while comprehensive in terms of subject coverage, lack in-depth analysis specific to hepatotoxicity.

In response to these needs, we introduce the HepatoToxicity Portal (HTP), a specialized web application focused on liver toxicity. HTP integrates curated data from diverse toxicity databases and presents accurate prediction models trained on extensive datasets. Our knowledgebase systematically catalogs hepatotoxic compounds based on multiple reference sources, compiling the hepatotoxicity scores with manual curation, which could be valuable for both non-toxicology and toxicology

researchers. Moreover, to address the persistent issue of data scarcity in biology-based deep learning models, HTP leverages a generally pre-trained molecular graph-based model and fine-tuning techniques, resulting in improved performance compared to existing methods.

Construction and content

Database overview

The HTP comprises two modules, namely the HTP KnowledgeBase (HTP-KB) and HTP Prediction (HTP-Pred) (Fig. 1). HTP-KB serves as a knowledgebase, consolidating information from nine public resources. After annotating the compound ID from PubChem [40], the collected documents underwent manual curation classifying their information into three classes: clinical, in vivo, and in vitro evidence. Additionally, liver toxicity-related terms from Medical Dictionary for Regulatory Activities (MedDRA) were annotated based on the anticipated biological mechanisms of each compound. The overall hepatotoxicity score was computed considering the methodological importance of each reference of the contents. The subsequent module, HTP-Pred, is a hepatotoxicity prediction tool that leverages a pre-trained graph neural network on large unlabeled molecule data, which is fine-tuned using our curated dataset for hepatotoxicity prediction. HTP-KB and HTP-Pred are integrated into a web information portal with enhanced visualizations. Users can predict the toxicity score of new small molecules and identify substructural toxicophores.

Data collection and curation

Data collection and integration

The HTP-KB comprises a comprehensive collection of nine chemical-related databases, each established with diverse objectives and affiliations (Table 1). These databases are categorized based on their specific purposes, including organizing results from drug experiments (CEBS [12], InvitroDB [11]), aggregating information on commercially available drugs (DrugBank [41], DILIrank [9], SIDER [42], LiverTox [10]), and curating case studies on drug-environment effects along with relevant publications (T3DB [13], IRIS [14], ATSDR [15]).

Depending on the database, liver-specific content was either readily accessible or required additional filtering from the complete dataset. The downloaded dataset underwent manual filtering to ensure its relevance to liver toxicity. Throughout the annotation process, PubChem Compound Identifiers (CIDs), widely used across most databases, were employed. In cases where assigning

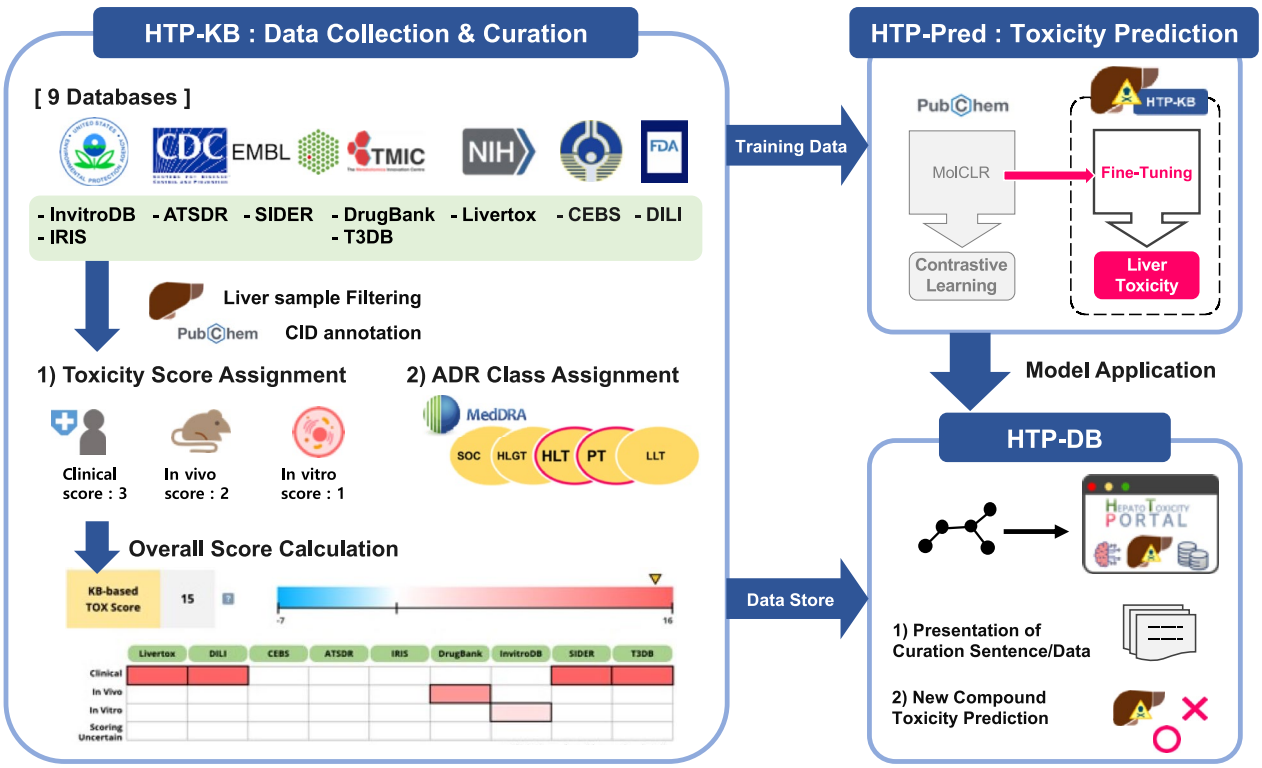

Fig. 1 An overview of the HTP database, illustrating the development of the knowledgebase, prediction module, and web portal

Table 1 Collection and characteristics of databases for HTP-KB

Database (Download date)	Affiliations	Hepatotoxicity information accessibility	Toxicity classification	Data format	No. of compound entries	No. of final curated CIDs	Data class
InvitroDB v3.4 (2022.02.04)	EPA (U.S.)	Low	Δ	CSV	9329	6466	In vitro
T3DB (2022.07.20)	TMIC (Canada)	Low	O	CSV	3673	955	In vitro In vivo clinical
LiverTox (2022.01.27)	NIH (U.S.)	High	O	PDF	1133	1005	In vitro In vivo clinical
DILIrank (2022.01.27)	FDA (U.S.)	High	O	PDF	1036	669	Clinical
SIDER (2022.02.04)	EMBL (Europe)	Middle	O	PDF	1430	749	Clinical
IRIS (2022.01.28)	EPA (U.S.)	Middle	Δ	PDF	666	138	In vivo clinical
CEBS (2022.02.04)	NTP (U.S.)	Low	X	CSV	9209 CASID	230	In vivo
ATSDR (2022.02.03)	CDC (U.S.)	Low	X	PDF	216	145	In vivo clinical
DrugBank v5.1.7 (2020.09.08)	TMIC (Canada)	Low	X	XML	13,580	154	In vitro In vivo clinical

* Hepatotoxicity accessibility levels (High, Easily accessible to liver-related toxicity results; Middle, Liver-related toxicity data available, yet needs additional filtration; Low, Difficult to extract liver-related results), Toxicity classification levels (O, Classification system is easily interpreted; Δ, Classification exists, yet not systematic; X, There is no classification levels)

a unique PubChem CID was unclear, PubChemPy (ver.1.0.4), a tool for retrieving related compound data using various substance identifiers, was utilized. The detailed curation processes varied due to disparities in available data across databases (Supplementary Fig. S1). The specific quantities of data before and after filtering is outlined in the Supplementary Materials.

MedDRA annotation

To describe biological activities with standardized vocabularies, we utilized the Medical Dictionary for Regulatory Activities (MedDRA) terms [43] to annotate documents and references aggregated from nine databases. MedDRA is an international medical ontology that supports a wide range of pharmaceutical and medical subject structured into four hierarchical levels under the System Organ Class (SOC) (Supplementary Fig. S2). The MedDRA ontology was accessed via BioPortal (BioPortal MedDRA 2019AB, accessed 2019.11.18). The four levels of the MedDRA structure consist of High-Level Group Terms (HLGT), High-Level Terms (HLT), Preferred Terms (PT), and Lowest Level Terms (LLT). For annotating references related to hepatotoxicity, we focused on SOC-level terms 'Hepatobiliary disorders' and 'Investigations', extracting their sub-hierarchical data from BioPortal. Under 'Hepatobiliary disorders', we selected five HLGT terms: 'Hepatic and hepatobiliary disorders', 'Hepatobiliary neoplasm', 'Bile duct disorders', 'Gallbladder disorders', and 'Hepatobiliary investigations'. Additionally, to include laboratory blood tests for liver function, we chose the HLGT term 'Hepatobiliary investigations' under 'Investigations'. We then utilized HLT and PT level terms within these selected HLGT terms to classify each reference in detail. Each HLT-PT set was paired to ensure precise clustering and annotation of data. To maintain focus on liver toxicity, we limited the inclusion of terms related to bile duct or gallbladder to one HLT-PT set per organ (*i.e.* 'Bile duct disorders'- 'Bile duct disorders' and 'Gallbladder disorders'- 'Gallbladder disorders'). Furthermore, recognizing the clinical complexity, we selected the HLGT term 'Hepatobiliary neoplasms' to cover terms related to liver cancer at the PT level.

Calculation of the hepatotoxicity score

Due to the heterogeneous nature of information resources, estimating the reliability and relevance of records to hepatotoxicity poses challenges. To consolidate multiple records into a single metric, we developed a scoring system that assigns higher weights to clinical references over in vitro and in vitro data. In our classification of references, we assigned arbitrary weights of 3 for clinical evidence, 2 for in vivo evidence, and 1 for in vitro evidence. The overall hepatotoxicity score for a compound (c) is calculated as the weighted sum of contributions from all records across nine source databases, taking into account whether each record has a positive or negative impact on hepatotoxicity:

$$S_c = \sum_{i=1}^{n_c} sign_c(i) \times weight_c(i) \quad (1)$$

where: n_c, number of records for compound c; $sign_c(i)$, $+1$ or -1 according to whether the record describes positive or negative evidence of hepatotoxicity; $weight_c(i)$, 3, 2, or 1 for clinical, in vivo, or in vitro evidence, respectively.

Conflicting records within a database are excluded from the sum (i.e., given a weight of zero). This scoring system allows us to assess the overall hepatotoxicity potential of a compound based on aggregated evidence from diverse sources while considering the varying quality and type of data provided by each database.

HTP-KB contents and statistics

The integration of nine databases followed by manual curation and scoring has resulted in the creation of the most comprehensive knowledgebase on hepatotoxicity. We provide a brief overview of the statistics for the HTP-KB contents, including evidence classes, source databases, annotation levels, and overall hepatotoxicity scores in Fig. 2. Additionally, the detailed contributions and compound overlaps from each database are presented in Supplementary Fig. S3. All statistics are based on the PubChem CIDs.

HTP-KB includes a total of 8306 compounds curated manually into three classes by toxicology experts. There are 2260 (27.2%) entries supported by clinical evidence, significantly surpassing entries found in LiverTox or DILIrank (Fig. 2a and b). Entries supported by in vitro evidence constitute the largest portion, with 6472 (77.9%) compounds, indicating that HTP-KB has substantially broadened the scope of hepatotoxic compounds by incorporating in vitro evidence.

Analyzing the source databases of the records, 2260 entries in the clinical class are distributed across databases such as LiverTox (1005), T3DB (890), SIDER (748), and DILI (669) (Fig. 2b). CEBS contributes the largest collection of in vivo evidence, albeit representing a smaller portion of the knowledgebase. Almost all in vitro evidence is sourced from InvitroDB.

Next, we examine the distribution of the hepatotoxicity scores within our database, ranging from -7 to $+16$ (Fig. 2c). The histogram plot showed a skewed distribution towards the positive side, likely because it is generally easier to determine positive hepatotoxicity compared to negative hepatotoxicity based on

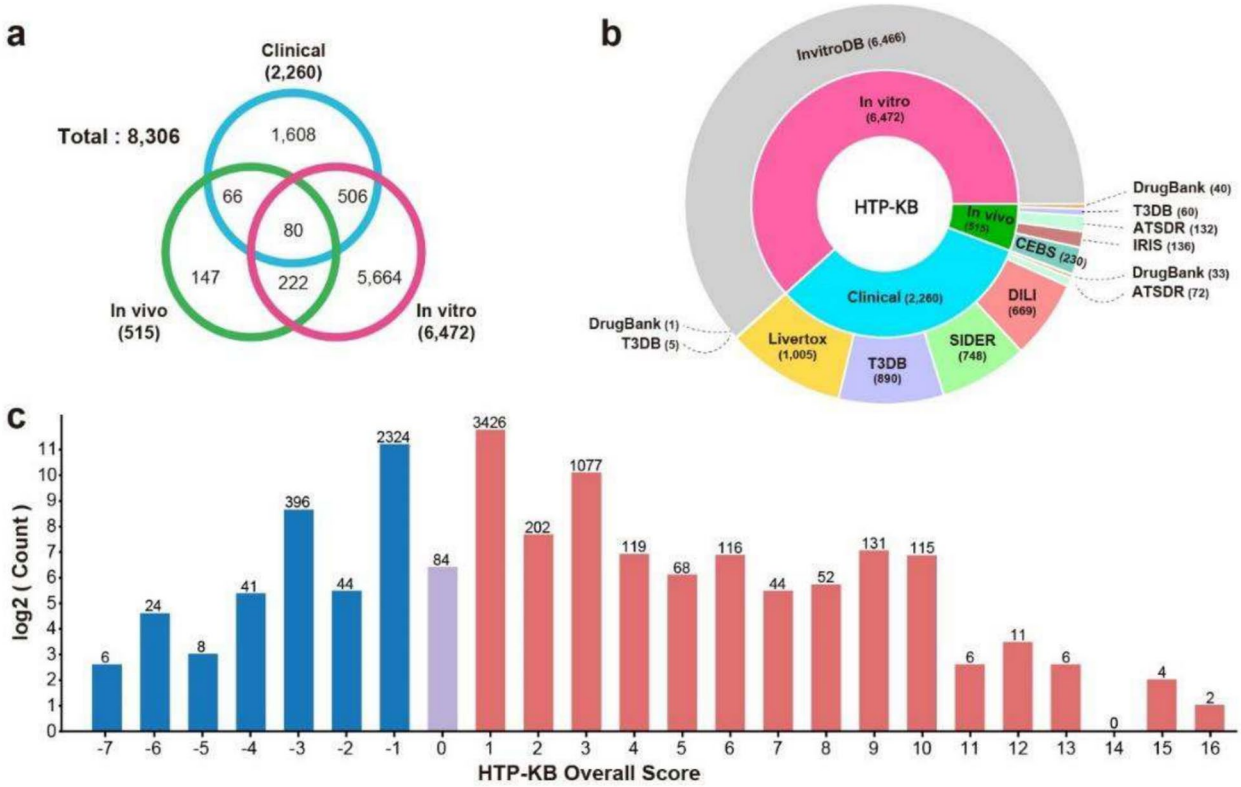

Fig. 2 Statistics of HTP-KB data. **a** Venn diagram of compounds with clinical, in vivo, and in vitro evidences. **b** Distribution of different classes of evidence records across source databases. Note that each compound may be annotated in multiple databases. **c** Histogram of overall hepatotoxicity score for all compounds in HTP-KB. Note that the frequency values are on a log2 scale to visualize the distribution effectively

experimental or literature evidence. Overall, HTP-KB includes 5379 compounds with positive scores and 2843 compounds with negative scores in terms of overall hepatotoxicity.

Annotation using MedDRA terms provides valuable insights into biological functions. Our annotation of hepatotoxicity utilizes a combination of High-Level Term (HLT) and Preferred Term (PT) terms from MedDRA terminology. The largest portion of the HLT terms is attributed to 'Hepatocellular damage and hepatitis NEC' (30%), encompassing various PT terms such as 'Hepatotoxicity', 'Hepatitis', 'Liver injury', and 'Hepatic necrosis' for sub-level categorizations (Supplementary Fig. S4). Other significant HLT terms include 'Cholestasis and jaundice' (14%), 'Hepatic enzymes and function abnormalities' (14%), and 'Hepatic and hepatobiliary disorders NEC' (13%). Cancer-related terms such as 'adenoma' and 'carcinoma' contributed to a relatively small portion (5% and 5%, respectively).

Development of HTP-Pred model

Pre-processing the HTP-KB dataset

To prepare the training and test data for the HTP-Pred model, we further curated the original HTP-KB dataset through additional pre-processing steps. Specifically, the data were re-labeled into binary classes as either hepatotoxic or non-hepatotoxic compounds after excluding molecules with fewer than three or more than 60 heavy atoms. Merging diverse hepatotoxicity datasets often results in data entries with conflicting labels. Excluding all such entries affects the model performance adversely due to insufficient amount of training data or overfitting limited amount of data. To address this, we resolved label conflicts by prioritizing the source database in the following order of reliability: clinical, in vivo, and in vitro. Additionally, we excluded ambiguous cases when the evidence for a compound contradicts each other at the same level of reliability. This approach ensures the model is trained on higher-confidence data while maintaining a sufficient number of data points. For evaluating robustness of the model upon imbalanced dataset, we employed stratified tenfold cross-validation to calculate the average performance score and standard deviation. Additionally, for

comparison with other hepatotoxicity prediction tools, we conducted hold-out validation. The dataset was split into training, validation, and test sets in an 8:1:1 ratio, maintaining an equivalent positive-to-negative class distribution. This split resulted in 5592 compounds in the training set, 699 in the validation set, and 700 in the test set.

Fine-tuning MolCLR with the HTP-KB dataset

Next, we developed a hepatotoxicity classification model by fine-tuning a pre-trained graph neural network (GNN) model (Fig. 3). Pre-trained deep learning models on large amount of data are widely employed as foundational frameworks for various downstream tasks, particularly in cases with limited labeled data [44, 45]. Hepatotoxicity prediction is one such case; despite rigorous data curation from diverse databases, training a model solely on the HTP-KB dataset is insufficient to capture a broad chemical space. To address this limitation, we employed MolCLR [46], a pre-trained GNN utilizing self-supervised learning techniques. MolCLR leverages approximately 10 million unique molecules from PubChem for contrastive learning task, enabling it to learn generalizable molecular representations. This approach allows the model to adapt to downstream tasks of molecular property prediction, demonstrating superior performance on both regression and classification benchmarks. Accordingly, we fine-tuned the base GNN model of MolCLR on the HTP-KB dataset, compensating for data scarcity and enhancing hepatotoxicity prediction.

We utilized either a graph convolutional network (GCN) [47] or graph isomorphism network (GIN) [48] as the GNN backbone for the pre-trained model, with pre-trained parameters provided by the original MolCLR implementation. For the binary classification task, we appended a randomly initialized multi-layer perceptron (MLP) prediction head to the pre-trained GNN feature extractor module. Following MolCLR's training protocol, we fine-tuned the model for 100 epochs, using an initial learning rate of 1×10^{-4} for the base model and 5×10^{-4} for the prediction head. The resulting fine-tuned model was named HTP-Pred.

The performance of HTP-Pred is summarized in Table 2. As baselines, we used molecular descriptors from InterDILI [49] to build input features and applied machine learning (ML) methods, including support

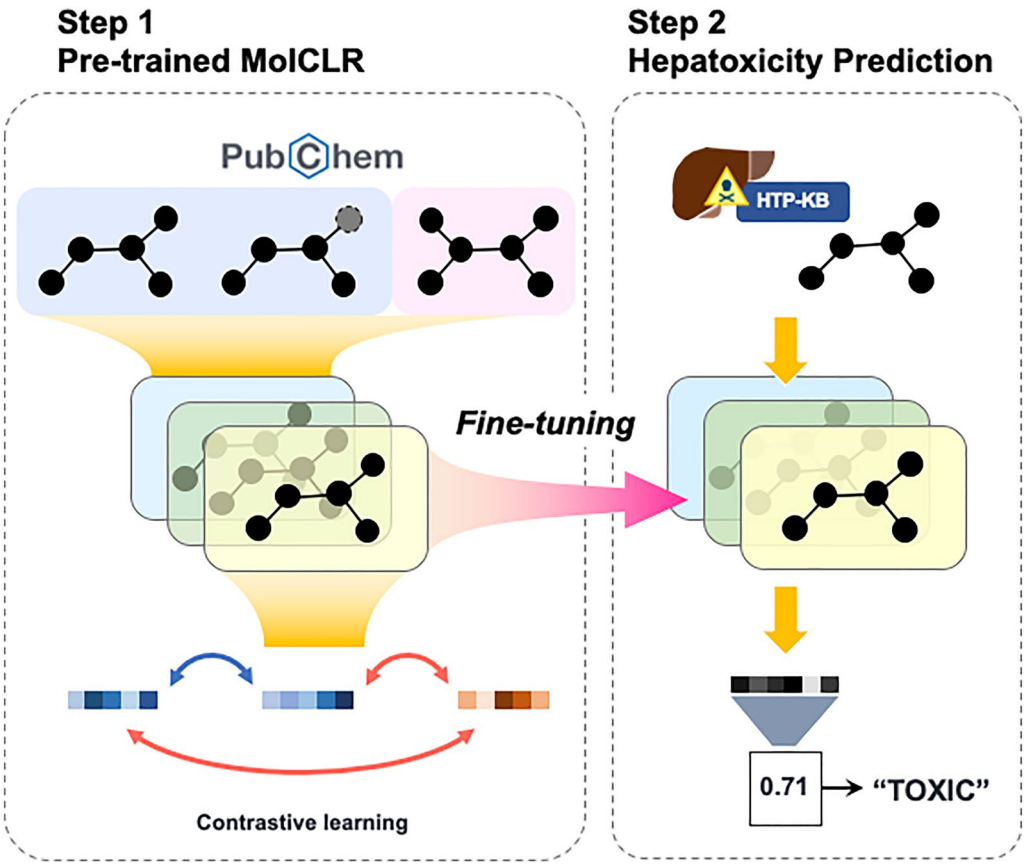

Fig. 3 Structure of HTP-Pred model

Table 2 Hepatotoxicity prediction performance of ML-based baseline models and HTP-Pred models with different pre-training method, with stratified tenfold cross-validation

Models	Pre-training method	AUROC
SVM (linear)	None	0.638 ± 0.015
SVM (polynomial)	None	0.538 ± 0.009
SVM (RBF)	None	0.658 ± 0.017
RF	None	0.677 ± 0.011
Logistic regression	None	0.633 ± 0.022
HTP-Pred-GCN	None	0.755 ± 0.020
	MolCLR	0.740 ± 0.023
HTP-Pred-GIN	None	0.763 ± 0.021
	MolCLR	**0.772 ± 0.019**

Best AUROC score shown in bold

vector machine (SVM), random forest (RF), and logistic regression, for classification. For SVM, we tested three kernel types: linear, polynomial, and radial basis function (RBF). Additionally, we conducted an ablation study on pre-training by training the backbone model from scratch. We also compared the performance of GCN- and GIN-based pre-trained models. AUROC scores were used as an evaluation metric, which captures the binary classification performance across different thresholds. Among the ML-based methods, RF achieved the best performance, consistent with the results from InterDILI. However, even without pre-training, the GNN-based classifiers outperformed the baseline ML models in terms of AUROC scores. Between the two backbones, GIN consistently outperformed GCN. Fine-tuning MolCLR further improved GIN-based performance, achieving the best AUROC score of 0.772. These results demonstrate that the pre-trained GIN-based MolCLR effectively captures informative molecular representations, leading to superior hepatotoxicity prediction.

Next, we evaluated the concordance between HTP-Pred predictions and the hepatotoxicity curation scores from HTP-KB, using the model trained on the hold-out validation set. Compounds in HTP-KB were categorized into three groups based on their hepatotoxicity scores: hepatotoxicity negative (KB score: − 7 to 0), moderately positive (KB score: 0 to 7), and highly positive (KB score: > 7). Compounds in the negative group exhibited significantly lower HTP-Pred scores compared to those in the positive groups, indicating that HTP-Pred effectively distinguishes hepatotoxicity-negative compounds from hepatotoxicity-positive ones (Supplementary Fig. S5). However, the moderately positive and highly positive groups showed similar score distributions, likely because the model was trained to predict the binary presence or absence of hepatotoxicity rather than specific score values.

Additionally, we compared HTP-Pred's performance against previous liver toxicity prediction tools for compounds (Table 3). Although we aimed to use the full test set of 700 compounds, some tools were limited by input constraints, restricting the comparison to 644 overlapping compounds. The list of these compounds is available in the model repository, alongside the model scripts (https://github.com/WonhoZhung/HTP_Pred). The results demonstrate that HTP-Pred outperforms existing toxicity prediction tools, likely due to the combination of a robustly curated dataset and advanced deep learning techniques, including the GIN-based molecular representation and fine-tuning of a GNN model pre-trained on large, unlabeled datasets. Further details on this comparative analysis can be found in the Supplementary materials.

Table 3 Performance comparison with existing prediction tools using 644 overlapping compounds

Models (Prediction tool)	Liver toxicity data	Algorithm	Accuracy	AUROC	References
ProTox [37] 3 (Webserver)	LiverTox	RF	0.457	0.549	[37]
ToxSTAR (Webserver)	DILIrank, PharmaPendium	SVM, RF	0.534	0.495	[31]
eMolTox (Webserver)	38/174 journal survey dataset	RF	0.598	0.553	[39]
admetSAR (Webserver)	–[a]	–[a]	0.624	0.498	[38]
Hepatopred-EL (Webserver)	DILIrank, FDA orange book	Ensemble (RF, SVM, XGB)	0.618	0.558	[29]
DeepDILI (Code-based)	DILIst, DrugBank, Wikipedia, FDALabelDB	Ensemble (LR, KNN, XGB, RF, SVM)	0.620	0.586	[32]
HTP-Pred	9 Datasets	GIN	0.717	0.761	

[a] Unable to specify the liver-specific training dataset or model algorithm

Hepato-toxicophore calculation

To enhance our comprehension and explainability of hepatotoxicity predictions for small molecules, it is essential to identify the contributions of individual atoms or substructures within a molecule. Gradient-based methods [50, 51], originally developed to assess pixel contributions in image-based predictions, were adapted for use with the HTP-Pred model. For a given molecular graph $\mathcal{G}$, each atom a is represented as a node feature $X_a \in \mathbb{R}^F$, where F denotes the feature dimension. To determine the contribution of each atom to the prediction, we first compute the absolute gradient of the prediction output $y_{\mathcal{G}}$, with respect to the input node features:

$$\widetilde{c}(\mathcal{G}, a) = \sum_{i=1}^{F} \left| \frac{\partial y_{\mathcal{G}}}{\partial X_{a,i}} \right|,$$

where $\widetilde{c}(\mathcal{G}, a)$ represents the unnormalized contribution score for atom a. Next, these scores are normalized across all atoms in the molecule to obtain the atom contribution score $c(\mathcal{G}, a)$:

$$c(\mathcal{G}, a) = \frac{\widetilde{c}(\mathcal{G}, a)}{\sum_b \widetilde{c}(\mathcal{G}, b)}.$$

This approach quantifies the contribution of each atom or substructure to the model's prediction outcome, enabling the identification of hepato-toxicophores (toxic substructures) within the input molecule. Note that the unnormalized contribution score is positive, so the normalized atom contribution score ranges between 0 and 1.

To define toxicophores, we used a set of SMiles ARbitrary Target Specification (SMARTS) patterns derived from Yang et al. [52], which employ a cheminformatics language for describing chemical patterns. RDKit functions were utilized to search for these substructure patterns within each compound. Atom contribution scores obtained earlier were summed for each pattern's corresponding atoms to derive an overall score $c(\mathcal{G}, \mathcal{S})$ for each substructure $\mathcal{S}$:

$$c(\mathcal{G}, \mathcal{S}) = \sum_{a \in V(\mathcal{S})} c(\mathcal{G}, a),$$

where $V(\mathcal{S})$ represents the set of atoms comprising substructure $\mathcal{S}$. The score of substructures, identified through toxicophore SMARTS matching, can also range from 0 to 1, indicating the contribution of the substructure to the model's decision. This methodology enabled the identification of key toxicophores by ranking substructures based on their overall scores. These ranked toxicophores provide insights into the molecular features most critical for hepatotoxicity prediction.

HTP Database and web server implementation

Database construction

PubChem CID was utilized as the primary identifier for each compound to efficiently link specific contents from individual databases with overall curation summary results. Additionally, sample IDs were created for references from their respective databases, formatted as numeric identifiers prefixed with the abbreviated database name. An SQL file was compiled to consolidate all database sample IDs with the main PubChem CID, integrating additional molecular properties and the corresponding HTP-Pred results. The web server operates by querying this comprehensive SQL file, ensuring seamless access to integrated data.

Web interface overview

The HTP web interface is designed to provide users with accessible and comprehensive information on chemical hepatotoxicity. The 'Search' section allows users to identify compounds through various methods, supporting multiple chemical ID formats and featuring visual representations of chemical structures for enhanced usability. An integrated statistics page presents a summary of the dataset, offering users a broad and detailed view of hepatotoxicity data coverage. The 'Downloads' section allows users to download the entire curated dataset or specific subsets from individual databases, enabling further analysis and research. To assist users in navigating and utilizing the HTP web server effectively, detailed instructions and usage guidelines are provided on the 'Help' page. This user-friendly interface ensures streamlined access to hepatotoxicity data for research and exploration.

Compound searching and browsing

In the 'Search' module, users can search for chemical compounds either by querying compound IDs or by drawing chemical structures (Fig. 4). Alongside PubChem CID, the primary identifier, we support diverse ID formats such as general compound names, IUPAC names, SMILES, CASIDs, and molecular formulas. Users have the option to choose between exact matching results or explore structurally similar or substructural compounds as per their needs. Additionally, users can input their original molecules using the JSME molecule editor. In cases where no matching compound is found in HTP-KB, only the HTP-Pred result is displayed, which is further detailed in the result interface section.

Alternatively, users can utilize the 'Statistics' module to explore overall data across each database and select preferred compounds. While this page provides comprehensive statistics for our data, clicking on each database name directs users to a detailed data browsing table. The result table includes user-friendly filtering options

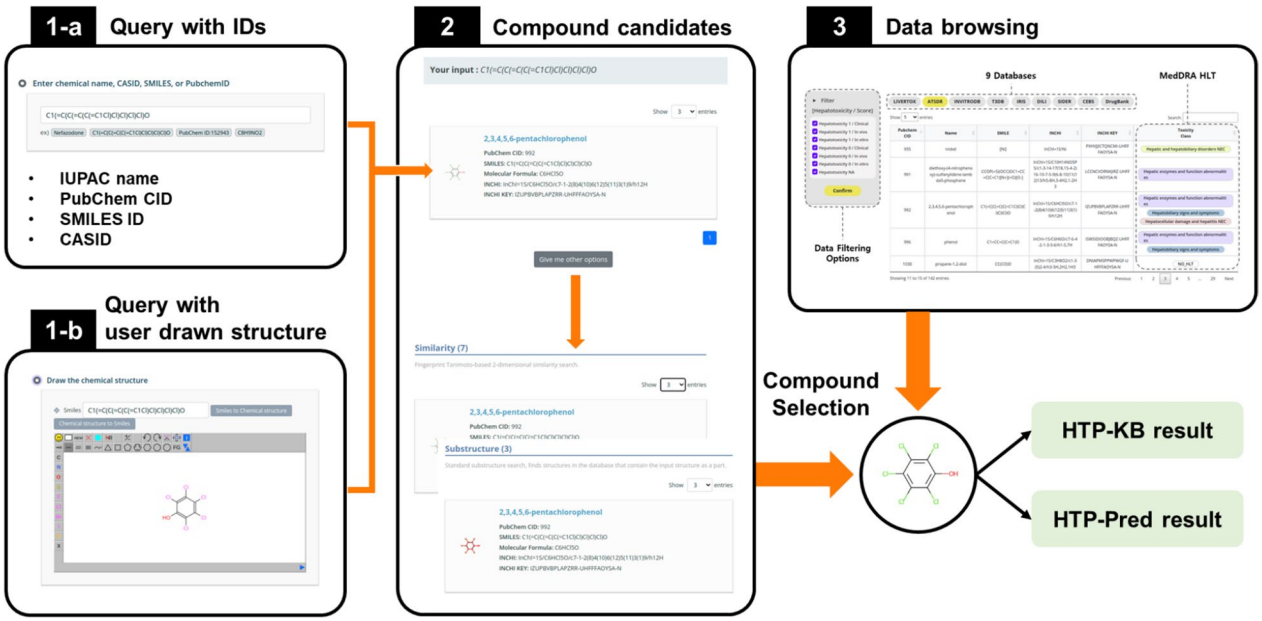

Fig. 4 User interface for compound searches. The figure illustrates example pages for searching and browsing compounds. Users can query compounds using several ID types or the JSME molecular editor. The search results screen allows users to select exact matching compounds, similar compounds, and substructural compounds through various options. The 'Statistics' menu provides access to individual database-wise browsing tables, allowing users to filter by toxicity class and score options. It also includes basic ID information and molecular properties for each compound. Upon final selection, users are presented with two main pages: HTP-KB search results and HTP-Pred results

via a selection bar adjacent to the table, allowing users to obtain a filtered list of compounds within each database. Each table entry features basic identifiers such as PubChem CID, SMILES, InChI, and InChI Key, alongside all unique lists of matched High-Level Terms (HLT). Clicking on any row navigates users to the specific compound result page.

HTP-KB result page

The HTP-KB result for the queried compound consists of several active subpages (Fig. 5). At the top of each HTP-KB subpage, a color bar indicates the overall hepatotoxicity score of the queried compound relative to the score distribution.

In the center of the page, a main table allows users to quickly assess the hepatotoxicity references from each database, along with their corresponding importance classes. Colored compartments within the table signify the characteristics of the data: red for hepatotoxic and blue for non-hepatotoxic. Clicking on each activated compartment reveals detailed results at the bottom of the screen.

Each database subpage varies in format due to distinct characteristics and evidence information for hepatotoxicity determination. However, all subpages include links to the original database web server and annotated MedDRA toxicity classification terms at the top. Even within a single database, multiple reference buttons may be provided to display results corresponding to various MedDRA terms. Clicking these buttons shows the main evidence sentence and overall data used for MedDRA term decisions. For databases such as ATSDR, DILI, LiverTox, and IRIS, which offer PDF-formatted files as resources, pages containing relevant sentences are prioritized, with additional pages accessible by scrolling through the embedded PDF file. If a database's primary data file is in CSV format (e.g., CEBS and InvitroDB), a subpage presents a table with selectable columns. Initially, pre-selected columns are displayed, but users can customize the view by selecting columns of interest. Some databases follow different formats not covered above. For instance, T3DB highlights crucial sentences related to data decisions among multiple sections on its subpages, while SIDER provides all MedDRA-related reference files. DrugBank presents only the critical sentence used in toxicity determination directly.

HTP-Pred result page

Another significant output of HTP is the prediction result generated by the HTP-Pred module (Fig. 6). The primary toxicity prediction score, displayed at the upper right part of the figure, indicates the likelihood of hepatotoxicity. This score is represented as a green dotted line on a plot showing the distribution of

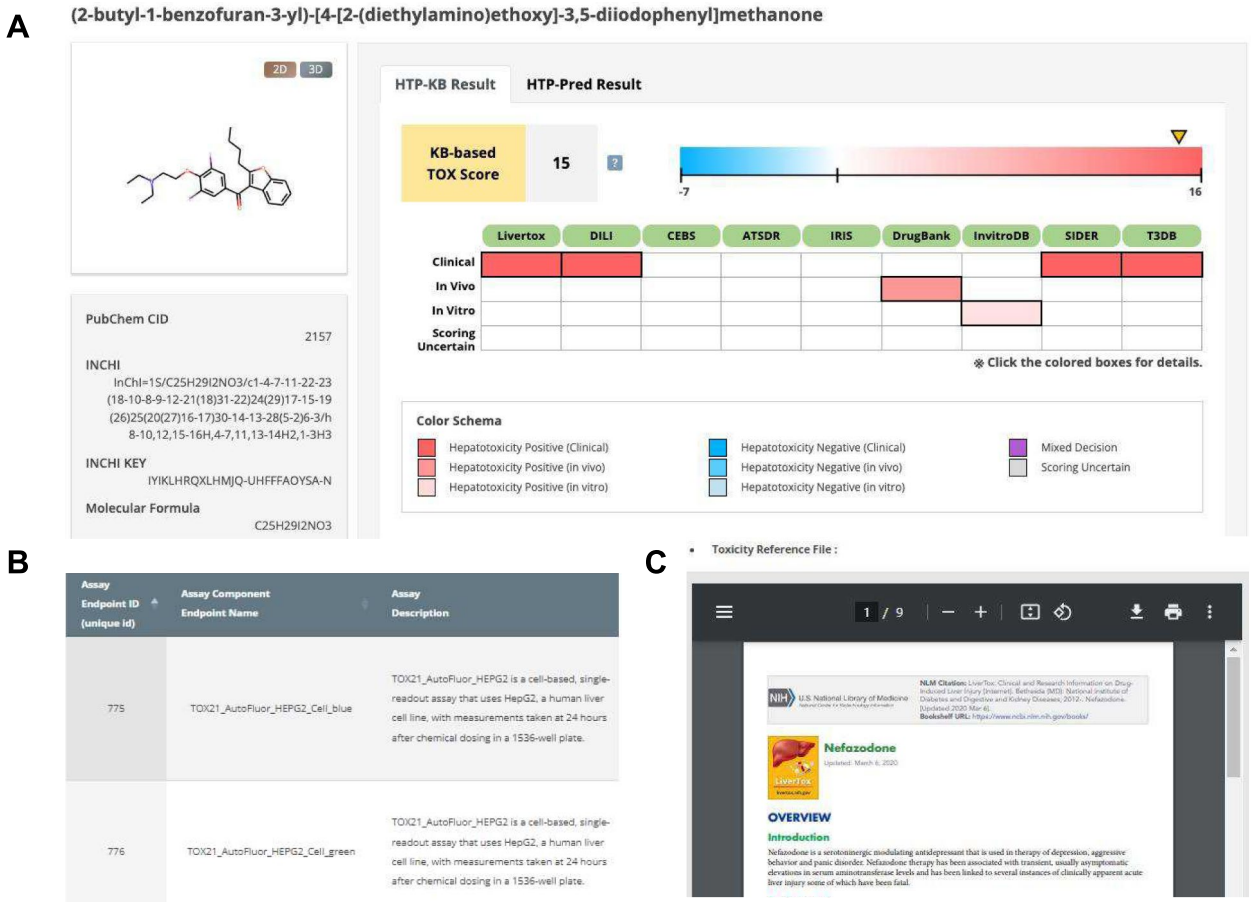

Fig. 5 HTP-KB result of queried compound. **a** Users can expand the compound's property information on the left side. A question mark icon next to the overall score explains score calculation, with a yellow triangle indicating its relative position on the color bar. Colored blocks in the main database table denote available curation sources. Selection of a database highlights it in yellow, revealing detailed information in distinct formats on subpages. **b** InvitroDB in CSV format, and **c** LiverTox in PDF format

prediction scores for HTP-KB compounds. To aid in assessing the confidence of the prediction result, HTP-KB compounds are categorized into three hepatotoxicity classes based on overall curation scores: negative (− 7 to 0), moderately positive (0 to 7), and highly positive (7 to 16). This categorization assists users in interpreting the prediction score relative to established thresholds for hepatotoxicity classification.

On the left side of the page, the compound structure is displayed, with each atom's importance score depicted in contours. A detailed table at the bottom of the figure specifies the importance score for each atom, highlighting the primary atom responsible for predicting the hepatotoxicity score. The lower part of the subpage presents the toxicophores result, accompanied by a detailed table on the right side. This table outlines the identified patterns of the toxicophore in SMARTS format, including the origin of SMARTS patterns, numerical identifiers, and a summation score derived from atom importance scores. Multiple toxicophores may correspond to the same SMARTS pattern, each identified with a distinct numerical identifier. Users can conveniently verify the location of each pattern highlighted on the compound by clicking the respective rows in the table.

Discussion and conclusions

The HepatoToxicity Portal (HTP) represents a pioneering effort in consolidating comprehensive hepatotoxicity data and advancing predictive modeling using state-of-the-art techniques. Both the knowledgebase (HTP-KB) and prediction modules (HTP-Pred) are designed to address critical gaps in understanding and predicting drug-induced liver injury. HTP-KB stands out for its extensive content and expert curation, classifying evidence into clinical, in vivo, and in vitro categories. A unique hepatotoxicity scoring system aggregates data from multiple sources into a unified metric, providing

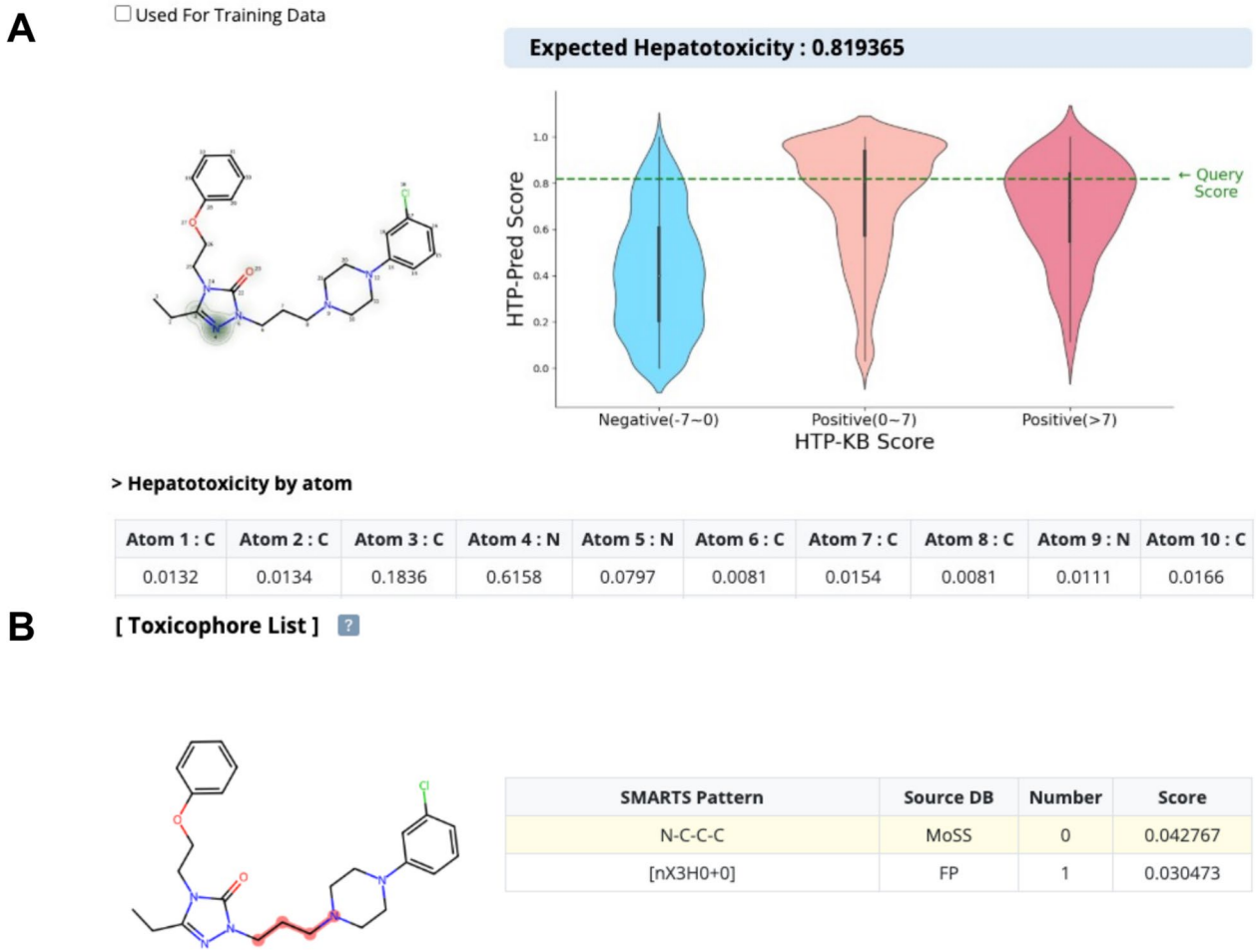

Fig. 6 Result page of HTP-Pred. The HTP-Pred result pages illustrate the predicted toxicity score and the contribution of each atom to toxicity assessment. **a** The HTP-Pred score is displayed with the distribution plot in the upper section. Additionally, the atomic importance scores are visually represented on the compound structure plot. **b** The toxicophore list is accessible through the table with visual representation on the compound structure. Columns include the SMARTS pattern, its source database, and the overall importance score of the substructure. The 'Number' column enumerates instances where multiple substructures correspond to a single SMARTS pattern. Users can interactively highlight specific substructures on the compound plot by selecting corresponding rows in the table

researchers across disciplines with a comprehensive overview of hepatotoxic compounds.

HTP-Pred leverages the pre-trained GIN model, MolCLR, which is trained on approximately 10 million unlabeled molecular data from PubChem and fine-tuned on the curated HTP-KB dataset. Comparative evaluation demonstrates superior performance compared to traditional ML-based baselines and other web servers for hepatotoxicity prediction. Additionally, HTP-Pred supports the identification of toxicophores, enabling researchers to pinpoint specific molecular features contributing to hepatotoxicity predictions, thereby aiding informed decision-making in drug design and optimization. However, the model may face intrinsic biases arising from the merged databases and the model itself. Quantifying and distinguishing aleatoric and epistemic uncertainties would provide deeper insights into the hepatotoxicity prediction results.

The HTP web interface provides intuitive access to curated data and predictive models, facilitating seamless navigation for users seeking detailed compound information on hepatotoxicity. It includes robust search functionalities and offers comprehensive curated information from HTP-KB along with prediction results from HTP-Pred.

Looking forward, ongoing updates and enhancements to HTP promise to refine predictive capabilities and expand database coverage, meeting evolving research needs in toxicology and pharmacology. HTP is poised to make a lasting impact on pharmaceutical research by

providing critical insights into liver toxicity mechanisms and facilitating the development of safer and more effective therapeutic agents.

In conclusion, HTP represents a significant advancement in toxicology and drug development. By integrating curated data from multiple databases and employing cutting-edge predictive models, HTP offers a comprehensive resource for assessing hepatotoxicity risks associated with chemical compounds. Its ability to merge sophisticated data curation with advanced deep learning methodologies underscores its potential to enhance drug safety evaluation and accelerate therapeutic innovation. In summary, HTP exemplifies the transformative potential of integrating curated data and advanced computational techniques, paving the way for enhanced drug safety assessment and biomedical research.

Abbreviations

ADME	Absorption, distribution, metabolism, and excretion
AUROC	Area under the receiver operating characteristic
CASID	Chemical abstracts service identifier
CID	Compound ID
DB	Database
DILI	Drug-induced liver toxicity
EPA	U.S. Environmental Protection Agency
GCN	Graph Convolutional Network
GIN	Graph Isomorphism Network
GNN	Graph Neural Network
HTP	HepatoToxicity Portal
INCHI	International Chemical Identifier
IUPAC	International Union of Pure and Applied Chemistry
KB	Knowledgebase
MedDRA	Medical Dictionary for Regulatory Activities
ML	Machine learning
Pred	Prediction
RBF	Radial basis function
RF	Random forest
SMARTS	SMiles ARbitrary Target Specification
SMILES	Simplified molecular–input line–entry system
SSL	Self-supervised learning
SVM	Support vector machine

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-025-00992-8.

Additional file 1.

Acknowledgements
The authors sincerely thank the curators of this project for their valuable contributions to the development of HTP-KB. We also extend our gratitude to the anonymous reviewers for their insightful suggestions to utilize a foundation model and fine-tuning method to enhance hepatotoxicity prediction.

Author contributions
JH – Data curation and validation, Writing—Original draft, Visualization; WZ and JL –Software development, Writing—Original draft; IJ – Webserver development; MJK and TDL – Data curation; SJK and KBK – Data curation, Project administration; DH – Project administration, Funding acquisition; BL – Supervision, Web development; HSK – Supervision, Project management, Funding acquisition; WYK – Supervision, Project administration, Writing—Review & editing; SL – Conceptualization, Supervision, Project administration, Writing—Review & editing; All authors corrected and approved the final manuscript.

Funding
This work was supported by the Ministry of Food and Drug Safety of Korea (Grant no. 20183MFDS410) and the National Research Foundation (NRF) of Korea (Grant no. 2020M3A916A0036057 for the KBDS program). This work was also supported by the Korea Bio Data Station (K-BDS) program in Korea Institute of Science and Technology Information (KISTI) with computing resources and technical supports.

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Bio-Information Science, Ewha Womans University, Seoul 03760, Republic of Korea. [2]Department of Chemistry, KAIST, Daejeon 34141, Republic of Korea. [3]Korea Bioinformation Center (KOBIC), KRIBB, 125 Gwahangno, Yuseong-Gu, Daejeon 34141, Republic of Korea. [4]Department of Life Sciences, Ewha Womans University, Seoul 03760, Republic of Korea. [5]School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea. [6]Department Bio Health Science, College of Natural Sciences, Changwon National University, Changwon 51140, Republic of Korea. [7]Center for Human Risk Assessment and College of Pharmacy, Dankook University, 119 Dandae-Ro, Cheonan 31116, Republic of Korea. [8]School of Pharmacy, Sungkyunkwan University, Suwon 16419, Republic of Korea.

Received: 3 August 2024 Accepted: 20 March 2025
Published online: 08 April 2025

References

1. David T (2021) Clinical development success rates and contributing factors 2011–2020
2. Harrison RK (2016) Phase II and phase III failures: 2013–2015. Nat Rev Drug Discov 15(12):817–818
3. Denayer T, Stöhr T, Van Roy M (2014) Animal models in translational medicine: validation and prediction. New Horizons Transl Med 2(1):5–11. https://doi.org/10.1016/j.nhtm.2014.08.001
4. McGonigle P, Ruggeri B (2014) Animal models of human disease: challenges in enabling translation. Biochem Pharmacol 87(1):162–171. https://doi.org/10.1016/j.bcp.2013.08.006
5. Ruggeri BA, Camp F, Miknyoczki S (2014) Animal models of disease: preclinical animal models of cancer and their applications and utility in drug discovery. Biochem Pharmacol 87(1):150–161. https://doi.org/10.1016/j.bcp.2013.06.020
6. Olson H, Betton G, Robinson D et al (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul Toxicol Pharmacol 32(1):56–67. https://doi.org/10.1006/rtph.2000.1399
7. Ostapowicz G, Fontana RJ, Schiødt FV et al (2002) Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States. Ann Intern Med 137(12):947–954. https://doi.org/10.7326/0003-4819-137-12-200212170-00007
8. Chen M, Vijay V, Shi Q et al (2011) FDA-approved drug labeling for the study of drug-induced liver injury. Drug Discov Today 16(15–16):697–703. https://doi.org/10.1016/j.drudis.2011.05.007
9. Chen M, Suzuki A, Thakkar S et al (2016) DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. Drug Discov Today 21(4):648–653. https://doi.org/10.1016/j.drudis.2016.02.015

10. LiverTox: Clinical and Research Information on Drug-Induced Liver Injury. 2022. https://www.ncbi.nlm.nih.gov/books/NBK547852/. Accessed 17 Feb 2022
11. Feshuk M, Brown J, Davidson-Fritz S et al (2022) Invitrodb version 3.5 release. U.S. Environmental Protection Agency, Washington DC. https://doi.org/10.23645/epacomptox.6062623.v8
12. Waters M, Stasiewicz S, Alex Merrick B et al (2007) CEBS—Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Res 36(1):D892–D900. https://doi.org/10.1093/nar/gkm755
13. Wishart D, Arndt D, Pon A et al (2015) T3DB: the toxic exposome database. Nucleic Acids Res 43(D1):D928–D934. https://doi.org/10.1093/nar/gku1004
14. Integrated risk information system, U.S. EPA. https://www.epa.gov/iris. Accessed 3 Feb 2022
15. Agency for toxic substances and disease registry (ATSDR). https://www.atsdr.cdc.gov/index.html. Accessed 3 Feb 2022
16. Williams AJ, Grulke CM, Edwards J et al (2017) The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminform 9:1–27. https://doi.org/10.1186/s13321-017-0247-6
17. NITE-CHRIP: NITE chemical risk information platform. https://www.nite.go.jp/en/chem/chrip/chrip_search/systemTop. Accessed 20 Aug 2022
18. eChemPortal. https://www.echemportal.org/echemportal. Accessed 20 Aug 2022
19. Greene N, Fisk L, Naven RT et al (2010) Developing structure—activity relationships for the prediction of hepatotoxicity. Chem Res Toxicol 23(7):1215–1222. https://doi.org/10.1021/tx1000865
20. Zhang H, Ding L, Zou Y et al (2016) Predicting drug-induced liver injury in human with Naïve Bayes classifier approach. J Comput Aided Mol Des 30:889–898. https://doi.org/10.1007/s10822-016-9972-6
21. Ekins S, Williams AJ, Xu JJ (2010) A predictive ligand-based Bayesian model for human drug-induced liver injury. Drug Metab Dispos 38(12):2302–2308. https://doi.org/10.1124/dmd.110.035113
22. Mulliner D, Schmidt F, Stolte M et al (2016) Computational models for human and animal hepatotoxicity with a global application scope. Chem Res Toxicol 29(5):757–767. https://doi.org/10.1021/acs.chemrestox.5b00465
23. Zhang C, Cheng F, Li W et al (2016) In silico prediction of drug induced liver toxicity using substructure pattern recognition method. Mol Inf 35(3–4):136–144. https://doi.org/10.1002/minf.201500055
24. Liu A, Walter M, Wright P et al (2021) Prediction and mechanistic analysis of drug-induced liver injury (DILI) based on chemical structure. Biol Direct 16:1–15. https://doi.org/10.1186/s13062-020-00285-0
25. Hong H, Thakkar S, Chen M et al (2017) Development of decision forest models for prediction of drug-induced liver injury in humans using a large set of FDA-approved drugs. Sci Rep 7(1):17311. https://doi.org/10.1038/s41598-017-17701-7
26. Chen M, Hong H, Fang H et al (2013) Quantitative structure-activity relationship models for predicting drug-induced liver injury based on FDA-approved drug labeling annotation and using a large collection of drugs. Toxicol Sci 136(1):242–249. https://doi.org/10.1093/toxsci/kft189
27. Kim E, Nam H (2017) Prediction models for drug-induced hepatotoxicity by using weighted molecular fingerprints. BMC Bioinform 18:25–34. https://doi.org/10.1186/s12859-017-1638-4
28. Zhu X-W, Li S-J (2017) In silico prediction of drug-induced liver injury based on adverse drug reaction reports. Toxicol Sci 158(2):391–400. https://doi.org/10.1093/toxsci/kfx099
29. Ai H, Chen W, Zhang L et al (2018) Predicting drug-induced liver injury using ensemble learning methods and molecular fingerprints. Toxicol Sci 165(1):100–107. https://doi.org/10.1093/toxsci/kfy121
30. He S, Ye T, Wang R et al (2019) An in silico model for predicting drug-induced hepatotoxicity. Int J Mol Sci 20(8):1897. https://doi.org/10.3390/ijms20081897
31. Shin HK, Chun H-S, Lee S et al (2022) ToxSTAR: drug-induced liver injury prediction tool for the web environment. Bioinformatics 38(18):4426–4427. https://doi.org/10.1093/bioinformatics/btac490
32. Li T, Tong W, Roberts R et al (2020) DeepDILI: deep learning-powered drug-induced liver injury prediction using model-level representation. Chem Res Toxicol 34(2):550–565. https://doi.org/10.1021/acs.chemrestox.0c00374
33. Kang M-G, Kang NS (2021) Predictive model for drug-induced liver injury using deep neural networks based on substructure space. Molecules 26(24):7548. https://doi.org/10.3390/molecules26247548
34. Xu Y, Dai Z, Chen F et al (2015) Deep learning for drug-induced liver injury. J Chem Inf Model 55(10):2085–2093. https://doi.org/10.1021/acs.jcim.5b00238
35. Lagunin A, Stepanchikova A, Filimonov D et al (2000) PASS: prediction of activity spectra for biologically active substances. Bioinformatics 16(8):747–748. https://doi.org/10.1093/bioinformatics/16.8.747
36. Maunz A, Gütlein M, Rautenberg M et al (2013) Lazar: a modular predictive toxicology framework. Front Pharmacol 4:38. https://doi.org/10.3389/fphar.2013.00038
37. Banerjee P, Kemmler E, Dunkel M et al (2024) ProTox 3.0: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Res. https://doi.org/10.1093/nar/gkae303
38. Yang H, Lou C, Sun L et al (2019) admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. Bioinformatics 35(6):1067–1069. https://doi.org/10.1093/bioinformatics/bty707
39. Ji C, Svensson F, Zoufir A et al (2018) eMolTox: prediction of molecular toxicity with confidence. Bioinformatics 34(14):2508–2509. https://doi.org/10.1093/bioinformatics/bty135
40. Kim S, Chen J, Cheng T et al (2019) PubChem 2019 update: improved access to chemical data. Nucleic Acids Res 47(D1):D1102–D1109. https://doi.org/10.1093/nar/gky1033
41. Wishart DS, Feunang YD, Guo AC et al (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46(D1):1074–1082. https://doi.org/10.1093/nar/gkx1037
42. Kuhn M, Letunic I, Jensen LJ et al (2016) The SIDER database of drugs and side effects. Nucleic Acids Res 44(D1):D1075–D1079. https://doi.org/10.1093/nar/gkv1075
43. Medical dictionary for regulatory activities (MedDRA). http://www.meddra.org/. Accessed 5 June 2022
44. Brown TB (2020) Language models are few-shot learners. arXiv preprint. arXiv:2005.14165, https://doi.org/10.48550/arXiv.2005.14165
45. Lin Z, Akin H, Rao R et al (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model. Science 379(6637):1123–1130. https://doi.org/10.1126/science.ade2574
46. Wang Y, Wang J, Cao Z et al (2022) Molecular contrastive learning of representations via graph neural networks. Nat Mach Intell 4(3):279–287. https://doi.org/10.1038/s42256-022-00447-x
47. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. https://doi.org/10.48550/arXiv.1609.02907
48. Xu K, Hu W, Leskovec J, et al. (2018) How powerful are graph neural networks? arXiv preprint arXiv:1810.00826. https://doi.org/10.48550/arXiv.1810.00826
49. Lee S, Yoo S (2024) InterDILI: interpretable prediction of drug-induced liver injury through permutation feature importance and attention mechanism. J Cheminform 16(1):1. https://doi.org/10.1186/s13321-023-00796-8
50. Sundararajan M, Taly A, Yan Q (2017) Axiomatic attribution for deep networks. Int Conf machine learning. https://doi.org/10.48550/arXiv.1703.01365
51. Simonyan K (2013) Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034. https://doi.org/10.48550/arXiv.1312.6034
52. Yang H, Li J, Wu Z et al (2017) Evaluation of different methods for identification of structural alerts using chemical ames mutagenicity data set as a benchmark. Chem Res Toxicol 30(6):1355–1364. https://doi.org/10.1021/acs.chemrestox.7b00083

Publisher's Note