

Toxicity Predictor
(LIVER/MIE-QSAR & LUNG/MIE-QSAR)

User Guide

July 2021 Ver.1.5

1. System overview.....	3
2. Recommended system requirements.....	3
3. Usage.....	4
3.1. Home page.....	4
3.1.1. Drawing structures.....	4
3.1.2. Input single SMILES.....	5
3.1.3. Upload SDF file.....	7
3.1.4. Upload SMILES list.....	7
3.2. Prediction Results.....	8
3.2.1. Wait completion of prediction.....	8
3.2.2. Check prediction results.....	8
3.2.3. Prediction results of DILI and DIPT.....	10
3.2.4. Prediction results of MIE.....	12
3.2.5. Similar medicines.....	13
3.2.6. Download prediction results.....	14
3.3. About prediction models.....	16
3.4. Sign up / Sign in.....	17
Acknowledgements.....	18

1. System overview

Toxicity Predictor is a QSAR system that predicts 59 types of molecular initiating events (MIE) in adverse outcome pathway (AOP), drug induced liver injuries (DILI) and drug induced pulmonary toxicities (DIPT) against input compounds. The prediction models are constructed by machine learning using activity data from PubChem and reported side effects data from JAPIC AERS. It also provides reported side effects data of similar medicines. Users can input compounds through a molecular editor on their web browsers, or by uploading SDF files or list of SMILES.

2. Recommended system requirements

The recommended web browsers and versions are as follows.

Web browser	Version
Google Chrome	78.0 and up
Firefox	70.0 and up
Safari	12.0.3 and up
Microsoft Edge	44 and up
Internet Explorer	11 and up

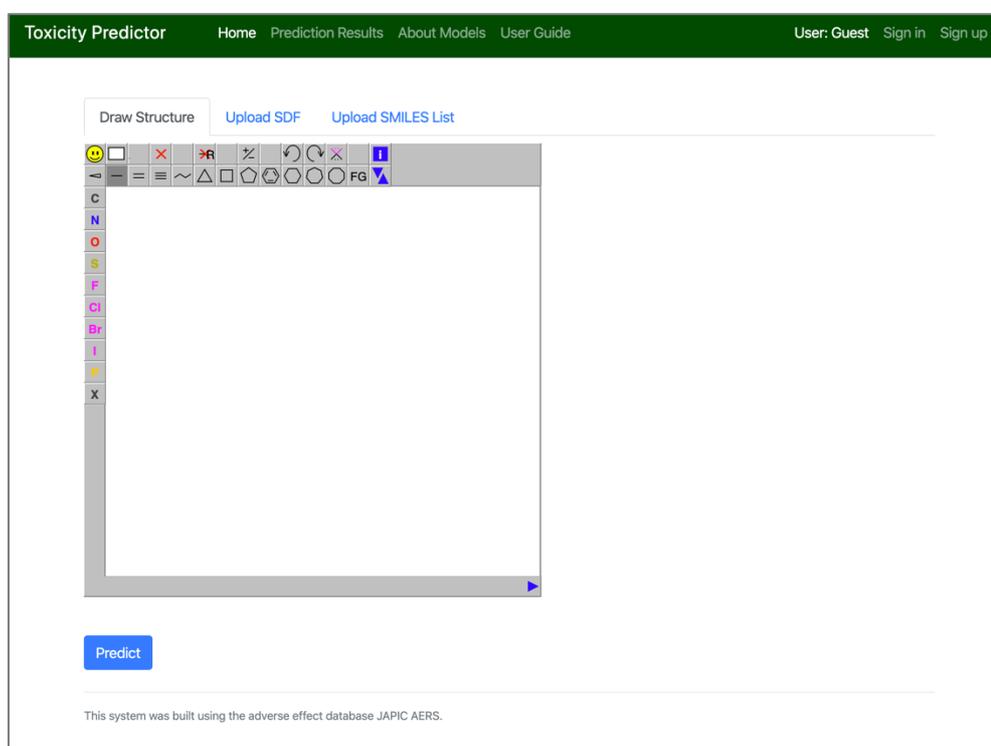
3. Usage

In this chapter, we explain usage of Toxicity Predictor.

3.1. Home page

When you access to the URL of Toxicity Predictor

(<http://mmi-03.my-pharm.ac.jp/tox1/>) , you can see the following home page. In this page, you can input compounds to predict. Compounds can be input by drawing structures on the molecular editor, paste single SMILES, uploading an SDF file or uploading a list of SMILES.



3.1.1. Drawing structures

On the molecular editor, you can draw a two-dimensional structure of a compound to predict. After drawing, click 'Predict' button as the following screen shot. After clicking, the web browser will navigate to the Prediction Result page (see section 3.2) .

Toxicity Predictor Home Prediction Results About Models User Guide User: Guest Sign in Sign up

Draw Structure Upload SDF Upload SMILES List

The screenshot shows the Toxicity Predictor web interface. At the top, there is a navigation bar with 'Home', 'Prediction Results', 'About Models', and 'User Guide'. On the right, it says 'User: Guest' with 'Sign in' and 'Sign up' links. Below the navigation bar, there are three tabs: 'Draw Structure' (selected), 'Upload SDF', and 'Upload SMILES List'. The main area contains a molecular editor with a toolbar at the top and a vertical element list on the left (C, N, O, S, F, Cl, Br, I, P, X). A benzene ring is drawn in the editor, with a red callout box pointing to it labeled '① Draw molecule'. Below the editor is a blue 'Predict' button, with a red callout box pointing to it labeled '② Click 'Predict''. At the bottom, a small text note reads: 'This system was built using the adverse effect database JAPIC AERS.'

3.1.2. Input single SMILES

If you have a SMILES you want to predict, please ①right click on the molecular editor to show the menu, ②click "Paste MOL or SDF or SMILES", ③paste SMILES on the window shown and ④click "Accept" button. Please check that the pasted SMILES is shown on the molecular editor as a two-dimensional structure, then ⑤click "Predict" button. After clicking, the web browser will navigate to the Prediction Result page (see section 3.2) .

You can paste a MOL format and an SDF format as wells as SMILES in this way, however, if you paste a SDF format, only the first compound in SDF are input. If you want to upload multiple compounds at the same time, please upload your SDF as a file (see item 3.2) or upload smiles list as a file (see item 3.1.4) .

Draw Structure

[Upload SDF](#)[Upload SMILES List](#)

① Right click on the editor

↑ ② Click "Paste MOL or..."

Draw Structure

[Upload SDF](#)[Upload SMILES List](#)

← ③ Paste SMILES

↑ ④ Click "Accept"

↑ ⑤ Click "Predict"

3.1.3. Upload SDF file

By uploading an SDF file, you can start prediction of toxicity against multiple compounds at the same time. Note that the maximum number of predictions allowed for one user is limited up to 100. To upload SDF, please ①click "Upload SDF" to switch the tab, ②click "Select SDF" button to select a SDF file to upload and ③click "Predict". After clicking, the web browser will navigate to the Prediction Result page (see section 3.2) .

The screenshot shows the 'Toxicity Predictor' web application interface. The top navigation bar includes 'Home', 'Prediction Results', 'About Models', and 'User Guide', along with a user status 'User: Guest' and links for 'Sign in' and 'Sign up'. The main content area features three tabs: 'Draw Structure', 'Upload SDF', and 'Upload SMILES List'. The 'Upload SDF' tab is active, indicated by a red callout box with a left-pointing arrow and the text '①Click "Upload SDF"'. Below the tabs is a file input field containing the text '4_compounds.sdf'. To the right of this field is a blue button labeled 'Select SDF', with a red callout box above it containing an upward-pointing arrow and the text '②Click "Select SDF" and select a file to upload'. Below the file input field is a blue button labeled 'Predict', with a red callout box below it containing an upward-pointing arrow and the text '③Click "Predict"'. The overall interface has a green header and a white background.

3.1.4. Upload SMILES list

Instead of uploading an SDF file, you can also upload SMILES list to start multiple predictions. A file to be uploaded a SMILES list should exact one SMILES on each line. Note that the maximum number of predictions allowed for one user is limited up to 100. To upload a SMILES list, please ①click "Upload SMILES List" to switch the tab, ②click "Select SMILES List" to select a file to upload, then ③click "Predict" button. After clicking, the web browser will navigate to the Prediction Result page (see section 3.2) .

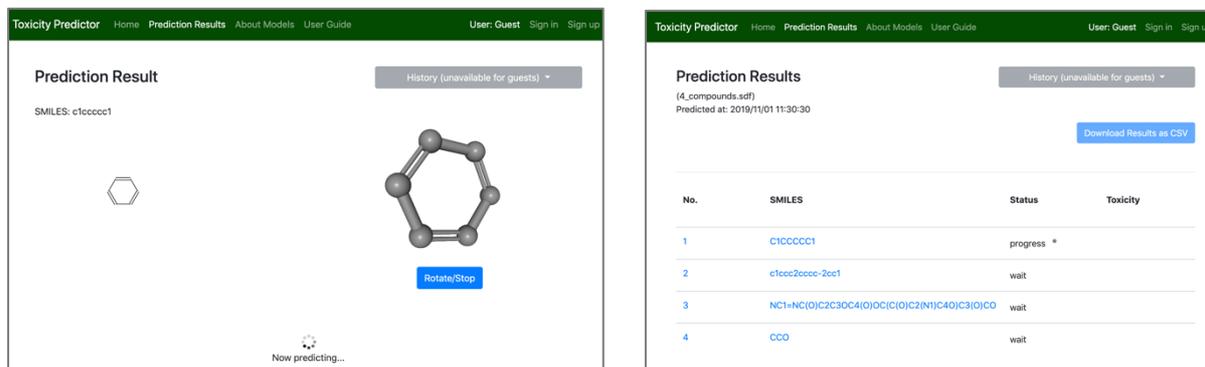
The screenshot shows the 'Toxicity Predictor' web application interface, similar to the previous one. The top navigation bar is the same. The 'Upload SMILES List' tab is active, indicated by a red callout box with a left-pointing arrow and the text '①Click "Upload SMILES List"'. Below the tabs is a file input field containing the text '2_compounds.smi'. To the right of this field is a blue button labeled 'Select SMILES List', with a red callout box above it containing an upward-pointing arrow and the text '②Click "Select SMILES List" and select a file'. Below the file input field is a blue button labeled 'Predict', with a red callout box below it containing an upward-pointing arrow and the text '③Click "Predict"'. A note below the file input field reads 'Uploaded file should contain one SMILES for each line'. The overall interface has a green header and a white background.

3.2. Prediction Results

On the "Prediction Results" page, you can see the results of toxicity prediction you performed. The web browser automatically navigates to this page after you start prediction. In addition, you can manually move to this page by clicking "Prediction Results" on the header menu on the top of each page.

3.2.1. Wait completion of prediction

After you started prediction, first you will see a waiting page as the following screen shots. If you input single compound, you will see the page like the left screen shot, otherwise you will see the right one. It takes about 1 minute to predict toxicity of one compound. Note that total time depends on the traffic condition of the server and the size of input compounds.



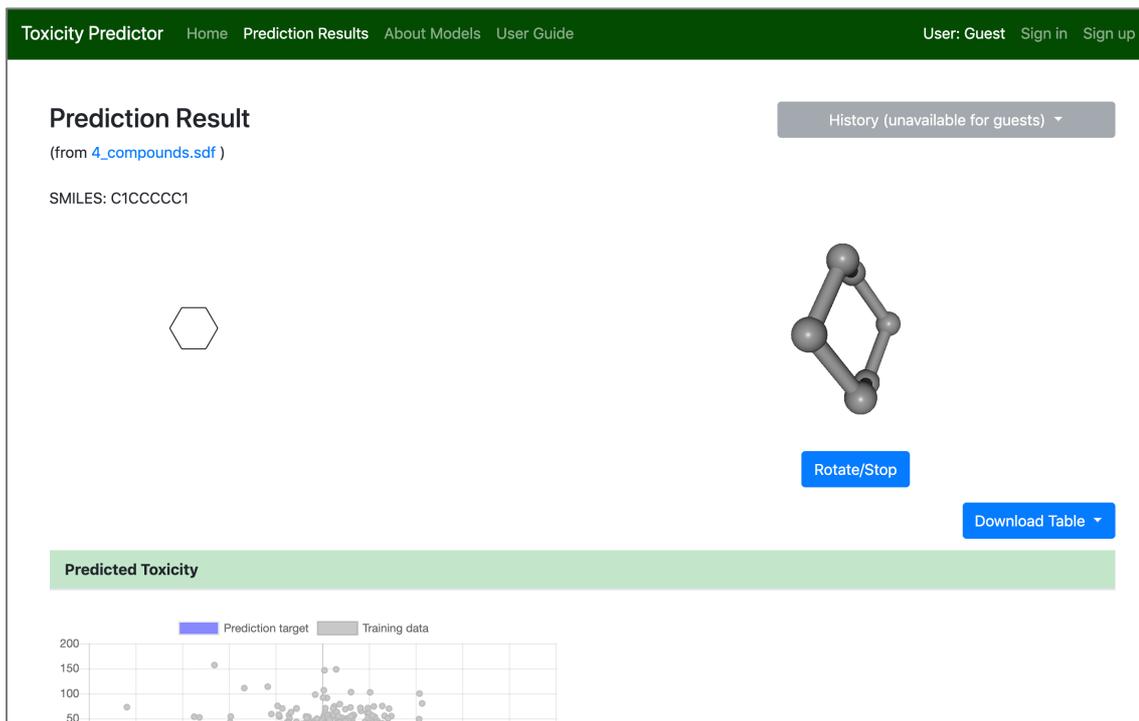
The left screenshot shows the "Prediction Result" page for a single compound. The SMILES string is c1ccccc1. A ball-and-stick model of the benzene ring is displayed, along with a "Rotate/Stop" button. The status at the bottom indicates "Now predicting...".

The right screenshot shows the "Prediction Results" page for a batch of 4 compounds. The prediction was completed on 2019/11/01 at 11:30:30. A "Download Results as CSV" button is available. The results are shown in the following table:

No.	SMILES	Status	Toxicity
1	C1CCCCC1	progress *	
2	c1ccc2ccccc2cc1	wait	
3	NC1=NC(O)C2C3OC4(O)OC(C)(O)C2N1C4O)C3(O)CO	wait	
4	CCO	wait	

3.2.2. Check prediction results

After a prediction completed, the result of the prediction will be shown. If you input single compound, the detail of the prediction result will be shown right away as the following screen shot. Please see item 3.2.3 and up for the explanation of this view.



On the other hand, if you upload multiple compounds, the status of prediction for each compound will be shown as the following screenshot. After a prediction completed, the "Status" column will change to "success". After it changes to "success", you can click the "No." column or the "SMILES" column to move to the detail page of the prediction. After all prediction are completed, you can download a list of prediction result in a CSV file (see section 3.2.6 for the contents of CSVs). If you signed in this system (see section 3.4), you can also view your history of predictions by clicking the upper right "History" button.

Toxicity Predictor Home Prediction Results About Models User Guide User: Guest Sign in Sign up

Prediction Results

(4_compounds.sdf)
Predicted at: 2019/11/01 11:30:30

Download Results as CSV

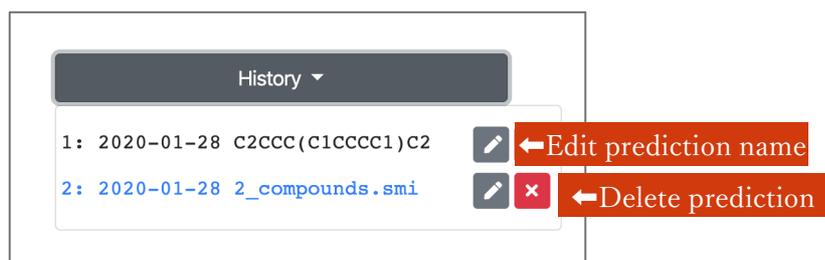
No.	SMILES	Status	Toxicity
1	C1CCCCC1	success	
2	c1ccc2cccc-2cc1	success	positive
3	NC1=NC(O)C2C3OC4(O)OC(C(O)C2(N1)C4O)C3(O)CO	success	positive
4	CCO	success	positive

View history (needs sign in)

Download CSV

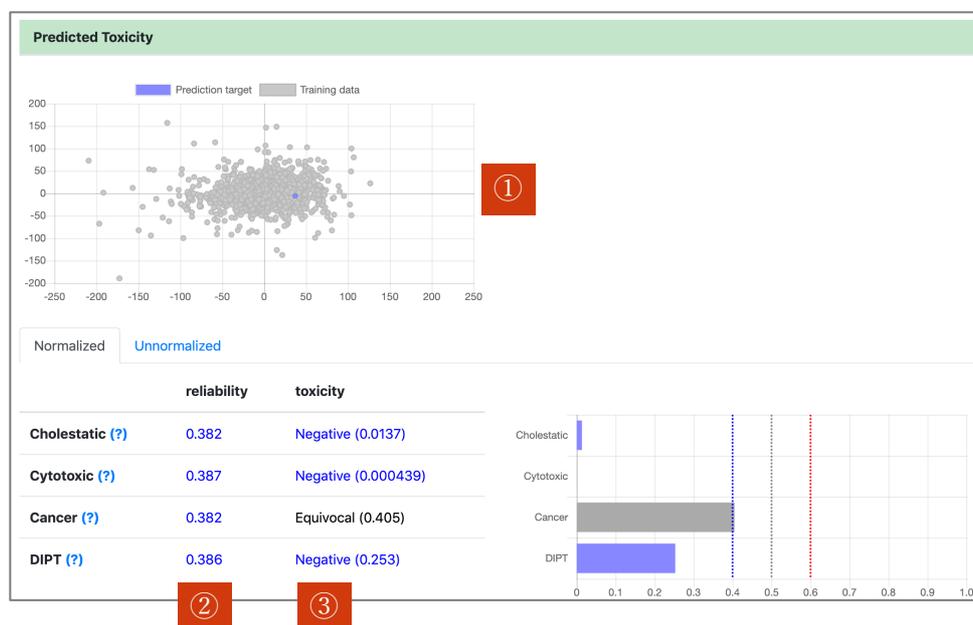
↑ Link to prediction details

If you click the "History" button, you can see list of file names and SMILES you input so far. If you click a name of a history, you can move to the prediction result page of the past prediction. In this window, you can also edit a name of a prediction by clicking the  button. You can also delete your prediction by clicking the  button.



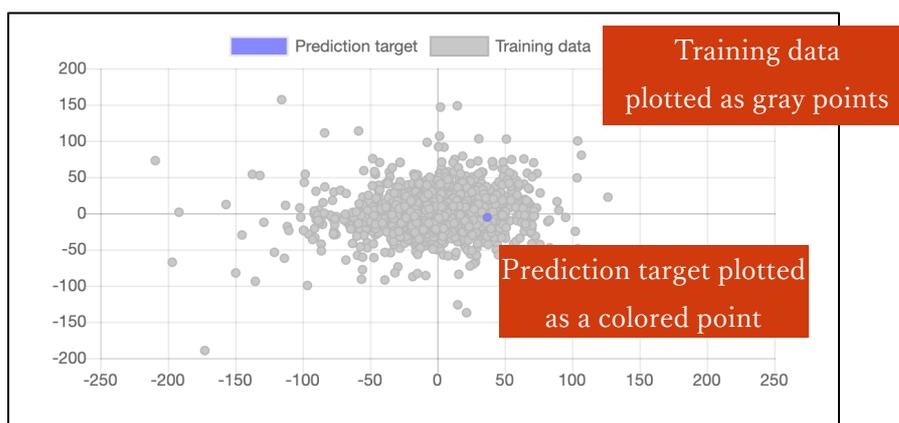
3.2.3. Prediction results of DILI and DIPT

On the top of the prediction detail page, you can see the prediction results of liver injuries and pulmonary toxicities. The contents of this part are a ① scatter plot of trained datasets and prediction target, ② reliability of prediction and ③ predicted toxicity.



①The scatter plot of trained datasets and prediction target shows the overview of position of the prediction target among the trained dataset. In this scatter plot, the trained data are plotted as gray points and the prediction target is plotted as a colored point. The coordinate of each point is computed by dimensional reduction of the high-dimensional

Euclidean space that includes corresponding molecular descriptors as data points. The point of prediction target is colored **Green (for known compounds)** , **blue (reliability: high)** , **red (reliability: low)** depending on the value of ②reliability. The reliability is computed based on the geometric mean of Euclidean distance between the prediction target and the k-nearest training data points.



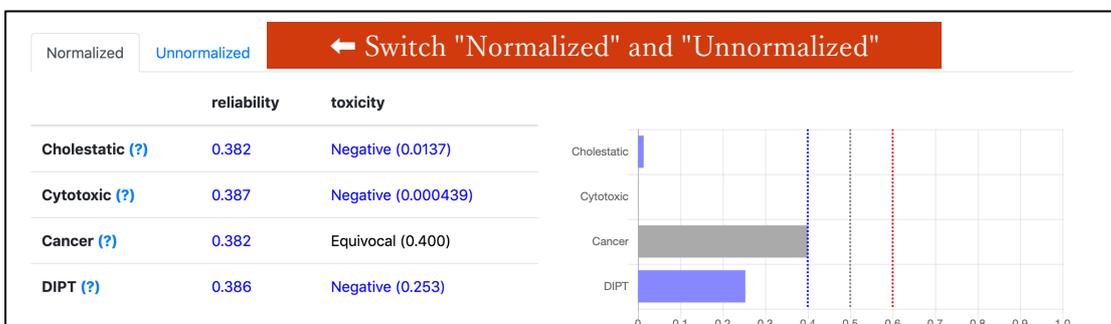
③The predicted toxicity is shown in a numeric value and a bar chart. The types of predicted toxicities are as the following table. You can check the performance of prediction models (see section 3.3) by clicking "(?)" links on the right of the name of the toxicities.

Short name	Description
Cholestatic	Cholestatic Liver Injury
Cytotoxic	Cytotoxic Liver Injury
Cancer	Liver Cancer
DIPT	Drug-Induced Pulmonary Toxicity

This system provides two tabs of prediction results, the "Normalized" tab and the "Unnormalized" tab. By default, the "Normalized" tab is selected. In the "Normalized" tab, the values x_n are shown which is obtained by normalizing the direct predicted value x_u using the following equation. Here, c is the cutoff value of each prediction model based on Youden Index of ROC curve) .

$$x_n := x_u^{-\log_c 2}$$

By clicking the "Unnormalized" tab, you can see the predicted value x_u without normalization.



In this table, the predicted value is colored depending on the value of x_n . If $x_n < 0.4$, the value is shown as **Negative (blue)**, **Equivocal (black)** if $0.4 \leq x_n \leq 0.6$, **Positive (赤)** if $0.6 \leq x_n$. If the input compound is included by training data, the experimented value is shown instead of the predicted value. When checking whether the input compound is included by the training dataset, this system distinguishes isomers if the input SMILES contains one of @, /, and ¥, otherwise it identifies isomers.

3.2.4. Prediction results of MIE

Below the results of DILI / DIPT prediction, the prediction results of 59 types of MIE follows. The basic meaning of the table and charts are the same as the part of DILI/DIPT, however, there are two models for each MIE so that two predicted values are shown. One model is constructed using criteria of the PubChem activity score $s \geq 1$ to be regarded as a positive sample. The other model is using criteria $s \geq 40$.

In the MIE part, there are a few models that could not obtain enough prediction performance. Such model is shown as "Criteria 1 or 40 is unavailable".



3.2.5. Similar medicines

Below the results of MIE prediction, medicines which are similar with the input compounds are shown. These medicines are retrieved from the report of JAPIC AERS.

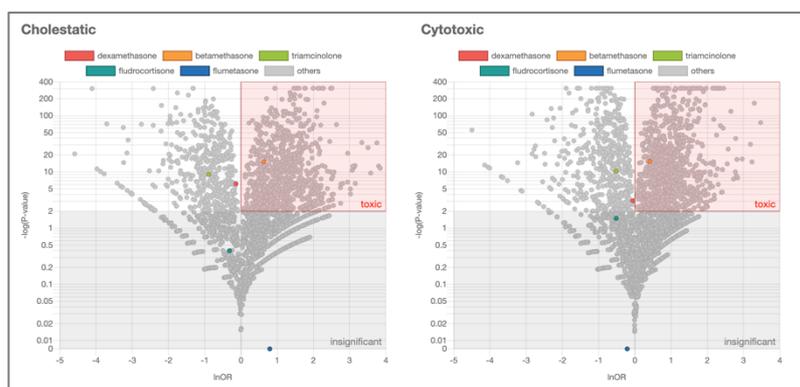
Similar Medicines												
name	similarity	n	Cholestatic			Cytotoxic			Cancer			DIPT
			toxicity	ROR	p-value	toxicity	ROR	p-value	toxicity	ROR	p-value	
calcium	0.333	566028	Negative	0.371	1.41×10^{-144}	Negative	0.457	1.63×10^{-276}	Negative	0.540	4.88×10^{-12}	Nega
hydrochloric acid	0.333	1698	Negative	0.663	0.600	Negative	1.02	1.00	Negative	2.64	0.434	Nega
oxaliplatin	0.182	138486	Positive	1.78	1.53×10^{-34}	Positive	2.56	1.11×10^{-302}	Positive	1.54	0.000682	Positi
sodium perchlorate	0.167	1028	Positive	3.30	0.00894	Positive	2.35	0.00677	Positive	7.28	0.0472	Nega
cyclamic acid	0.158	143	Negative	1.57	1.00	Negative	2.90	0.217	Negative	10.4	1.00	Nega

For each similar medicine, the following information is shown.

Name	Description
similarity	Similarity to the input compound (0 ~ 1)
n	Reported count

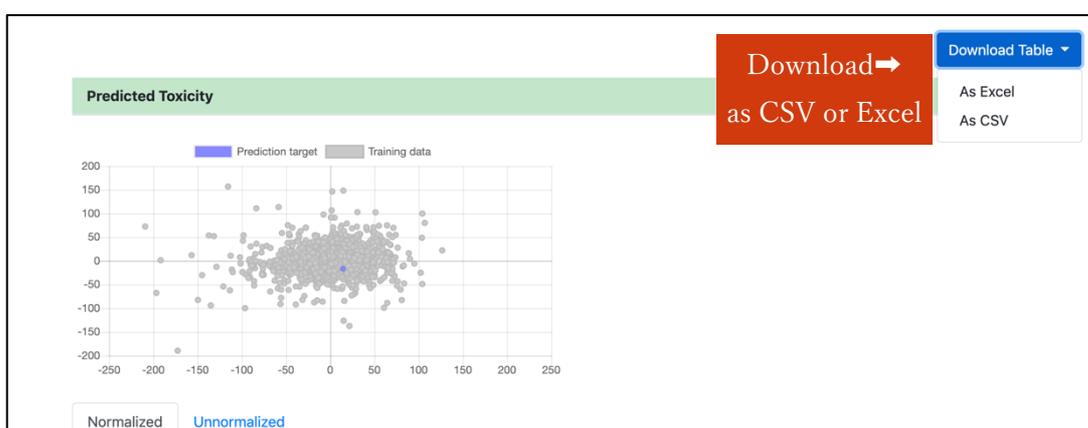
toxicity	Negative, Positive or Insignificant
ROR	Reported Odds Ratio
p-value	Reported significance in p-value
canonical smiles	canonical SMILES (not isomeric)
isomeric smiles	isomeric SMILES

For each DILI and DIPT, a volcano plot is shown as follows. In the volcano plots, a toxic area (toxicity = Positive) is shown as a red area and insignificant area (not enough p-value) is shown as a grey area.



3.2.6. Download prediction results

On the upper right of the page, you can click "Download Table" button to download result of the prediction in CSV or Excel table.



The contents of CSV are as follows.

Column name	Description
SMILES	SMILES of the input compound
<DILI/DIPT name>_normalized	Predicted toxicity for each DILI / DIPT after normalization.
<DILI/DIPT name>_unnormalized	Predicted toxicity for each DILI / DIPT before normalization.
<DILI/DIPT name>_reliability	Prediction reliability for each DILI / DIPT
<MIE name>_normalized_{1, 40}	Predicted value for each MIE after normalization. 1 or 40 stands for criteria.
<MIE name>_unnormalized_{1, 40}	Predicted value for each MIE before normalization. 1 or 40 stands for criteria.
<MIE name>_reliability	Prediction reliability for each MIE.

3.3. About prediction models

If you click the "About Models" menu on the top of each page, you can see the performance of prediction models.

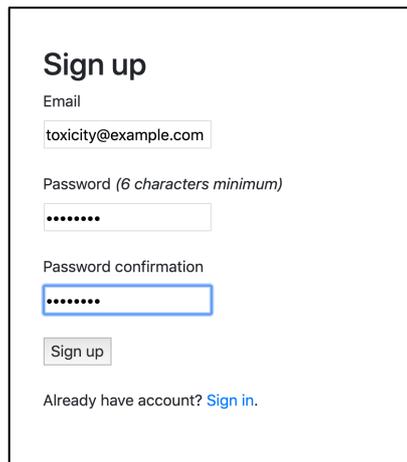
Performance of Prediction Models										
Hepatotoxicity Prediction Models										
Type	AUC	Sensitivity	Specificity	Cutoff						
Cholestatic Liver Injury	0.807	0.735	0.768	0.831						
Cytotoxic Liver Injury	0.836	0.765	0.764	0.752						
Liver Cancer	0.815	0.736	0.754	0.161						
Pulmonary toxicity Prediction Models										
Type	AUC	Sensitivity	Specificity	Cutoff						
Drug-Induced Pulmonary Toxicity	0.850	0.748	0.842	0.134						
MIE Prediction Models										
			Criteria: 1				Criteria: 40			
Index	Description	AID	AUC	Sensitivity	Specificity	Cutoff	AUC	Sensitivity	Specificity	Cutoff
1	ATAD5_ind (ATAD5 genotoxic inducer)	720516 PubChem	0.845	0.744	0.847	0.0370	0.840	0.750	0.843	0.0280
2	p53_ago (p53 agonist)	720552 PubChem	0.845	0.804	0.793	0.142	0.899	0.824	0.830	0.0269
3	MMP_disr (mitochondrial membrane potential disruptor)	720637 PubChem	0.795	0.698	0.788	0.368	0.919	0.845	0.846	0.0635

The description of the contents in this page are as the following table.

Name	Description
Type	Name of toxicity (for DILI / DIPT)
AUC	Area Under the Curve of ROC
Sensitivity	The value of sensitivity (TP / (TP + FN))
Specificity	The value of specificity (TN / (TN + FP))
Cutoff	Cutoff value on ROC based on Youden Index
Index	The index of MIE used in this system
Description	The abbreviated name and the detailed name of MIE
AID	Assay ID in PubChem and link to it

3.4. Sign up / Sign in

This system can be used as an anonymous user. But the history feature (see section 3.2.2) is only available for registered users. To sign up, please click "Sign up" menu on the header menu and input your email address and password. Sign up will be completed immediately after you click "Sign up" button.



Sign up

Email

Password (6 characters minimum)

Password confirmation

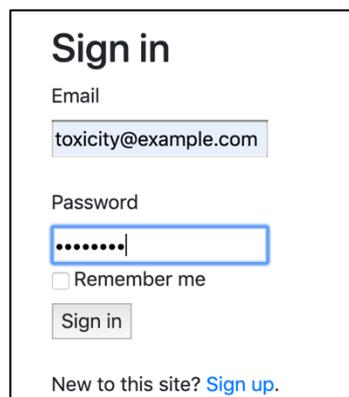
Already have account? [Sign in.](#)

After you signed in, you can see the your user name on the right side of the header menu. Please click "Sign out" if you want to sign out.



User: toxicity Sign out

If you want to sign in again, please click "Sign in" on the header menu and input your email address and password.



Sign in

Email

Password

Remember me

New to this site? [Sign up.](#)

Acknowledgements

This system was built using the adverse effect database JAPIC AERS.