Supplementary Information

MORF: Multi-View Oblique Random Forest for Hepatotoxicity Prediction

1. Detailed explanation of datasets

Rdkit is used to calculate 2D molecular representation of the compounds, including molecular fingerprint(fp), physico-chemical properties of the compounds (phychem)[1].The characteristic data of chemical compounds based on chemical molecular representation are very popular in computational models, which not only have high prediction performance of molecular properties, but also are not limited by experimental data[2], [3], [4]. Among them, PubChem belongs to the substructural bond fingerprint, containing the information of compound atom, bond and substructure.

DrugBank[5] is a bioinformatics and cheminformatics resource database, which contains freely accessible information on drugs and drug targets. Over 16000 compounds with more than 5763 related targets were collected by DrugBank. Then, we matched the database targets by CID of compounds in the hepatotoxicity tag dataset and finally obtained 721 compound 2657 target-regulated gene matching records. At the same time, we converted the target protein UniprotID into EntrezID by Rstudio package.

LINCS 1000(L1000)[6], [7] database contains about 1 million expression profiles, mainly from 99 cell lines in vitro expression data of 978 important genes under 32855 small molecule perturbations. The transcriptional expression profiles in this paper were obtained from the L1000 dataset, which contained z-score values corresponding to the normalized differential expression between the drug treatment and control across different conditions. Therefore, LINCS data has 978 dimensional features, representing 978 landmark genes. The original L1000 expression profile dataset only includes the data of Level 5. On this basis, the expression profile data of level 6 are obtained by weighted average exposure time and concentration dose of compound perturbation, and then the data of level 7 are obtained by balancing each cell line with the same treatment. The L1000 expression profile data for level 7 form specific expression profiles for each compound. Finally, 571 hepatotoxicity datasets containing transcriptome expression profiles, target genes, molecular fingerprints, and physicochemical properties were obtained by matching with the original datasets.

2. Impact of the number of ODTs on MORF performance

Increasing the number of base learners would lead to a higher randomness, which improves the accuracy and generalization performance of random forest algorithm. We therefore explore the impact of the number of oblique decision trees on the performance of MORF to select an appropriate number of base learners. We still choose combination 1 to carry out the experiment, and the number of decision trees is set from 10 to 100 in 10 groups of equally divided numbers. Figure S1 and Figure S2 show impact of the number of ODT on MORF-N and MORF-R.

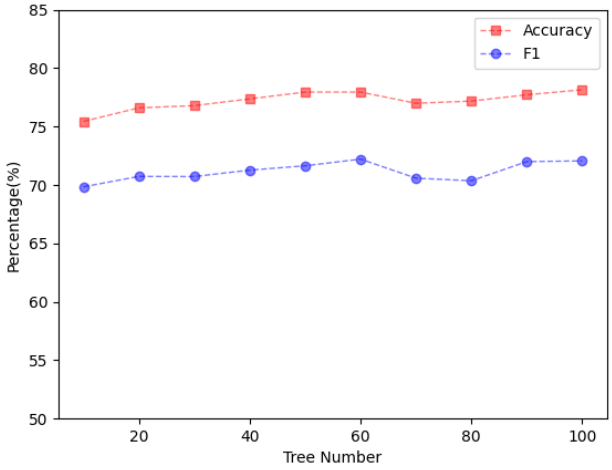

Figure S1 Impact of the number of ODT-Ns on MORF-N

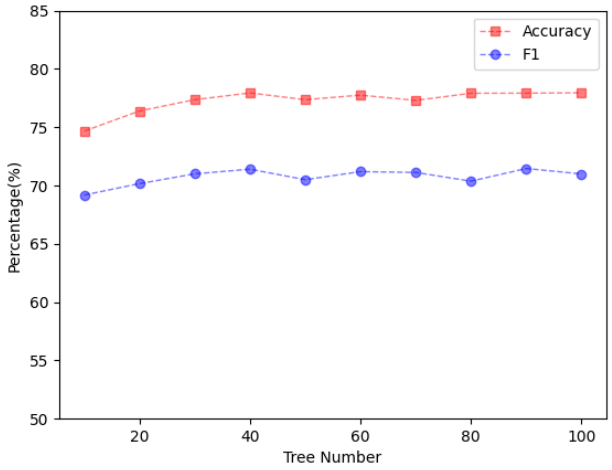

Figure S2 Impact of the number of ODT-Rs on MORF-R

As can be seen from these figures, Accuracy and F1 of the two algorithms gradually increase with the increase of the number of ODTs on MORF before 40. When the ensemble size is larger than 40, the performance of MORF-R tends to slightly increase stably. When containing more than 90 base learners, the performance of MORF-N was more stable. In conclusion, it is reasonable to set the number of MORF to 90 or more.

3. Comparison between DT and ODT

In order to verify the effectiveness of the oblique decision tree (ODT), this subsection compares the performance of ODT-N and ODT-R with the traditional ID3 decision tree(DT) and the results are given in Table S1.

It can be observed from Table S1 that under most of multi-views, ODTs achieve a higher Accuracy score compared with DT. The score gap varies according to the splicing method, most of which exceed DT by 1-3 percentage points, with the largest gap reaching 7 percentage points. For F1-score, ODTs achieve a higher score under eight feature combinations, with a maximum difference of 6 percentage points. From the above comparison results, it can be seen that ODTs have better classification effects on human hepatotoxicity dataset than DT.

Table S1 Accuracy and F1 comparisons between DT and ODTs

Method / Combination	Accuracy			F1-score		
	DT	ODT-N	ODT-R	DT	ODT-N	ODT-R
1	70.02	**72.73**	70.39	65.68	**67.15**	61.39
2	68.75	**70.40**	69.43	**65.00**	64.93	62.71
3	58.43	**68.07**	65.17	52.48	**62.72**	59.69
4	71.08	**72.73**	72.14	67.09	**67.73**	67.10
5	61.51	**68.47**	64.05	55.97	**62.52**	59.24
6	60.05	68.27	**69.85**	53.88	62.48	**63.36**
7	71.57	**72.34**	**72.34**	67.68	**68.55**	61.96
8	66.14	**69.05**	68.29	61.22	**63.77**	63.08
9	60.16	**69.24**	64.58	53.37	**59.79**	57.83
10	65.56	**69.83**	68.07	60.59	**65.15**	63.02
11	64.98	**71.37**	70.40	60.61	**67.33**	64.15

4. GOBP, KEGG pathway enrichment analysis

In order to further explore the toxigenic mechanism of hepatotoxicity caused by compounds, GOBP and KEGG pathway enrichment analysis were extracted based on the important features of the first 200 dimensions of the target gene and transcriptome expression profile, respectively. Figure S3 and Figure S4-6 show the biological processes and signaling pathways involved in the hepatotoxicity caused by different compounds.

Despite the diverse etiologies of hepatotoxicity, innate immunity activation is a common feature involved in drug induced hepatotoxicity progression[8]. Among the hepatotoxicity caused by drugs, acetaminophen (APAP)-induced hepatotoxicity is the most cause of acute liver failure (AFL) in United States[9,10]. APAP induced hepatotoxicity includes three pathological stages successively: injury initiation, injury amplification, and liver regeneration and repair[11]. As can be seen from Figure S4, a large number of genes were enriched in the GO biological process "wound healing" ($P<0.01$). The 200 most important genes obtained from the 978 genes in our model were massively enriched in this term. In fact, studies have shown that once ingested, excess APAP can be catalyzed by CYP2E1 to form the reactive metabolite N-acetyl-p-benzoquinimide (NAPQI), which leads to glutathione (GSH) depletion and mitochondrial dysfunction, thereby triggering damage[12,13]. Dysfunctional mitochondria then lead to cellular ATP depletion, DNA fragmentation, cell necrosis, and activation of resident liver macrophages (Kupffer cells (KCs)), further amplifying damage. Finally, cytokines and chemokines secreted by activated KCs further attract the homing and migration of neutrophils and macrophages to promote hepatic inflammation resolution, regeneration and repair[8, 12, 14]. Currently, N-acetylcysteine (NAC), a precursor of GSH, is the only approved treatment for AILI. This is sufficient to prove that our model identifies important pathways related to the hepatotoxicity of existing compounds. Understanding the mechanism of action of APAP can lead to the treatment of APAP-induced hepatotoxicity[15]. For example, C-C chemokine motif ligand 5 (CCL5) inhibition may be a promising treatment strategy for APAP overdose[8]. Therefore, our model can not only reveal the mechanism of hepatotoxicity caused by compounds through analysis, but can also help identify effective therapeutic measures

against this mechanism. Similarly, other terms with GOBP enrichment based on expression profile were "positive regulation of protein kinase and energy" (P<0.01)[16], "pyruvate metabolic process"[17] (P<0.01) and "Glycolytic Process"[38] (P<0.01), which has been shown to be associated with the development of hepatotoxicity caused by compounds.

KEGG pathway enrichment analysis based on expression profile showed that the genes were significantly enriched in the "Epstein-Barr virus infection"(P<0.01) pathway, as shown in Figure S4 Clinical cases have shown that Epstein-Barr virus(EBV) infection can cause various forms of liver enzyme abnormalities and even liver failure[19], which can also cause autoimmune hepatitis (AIH)[20]. In this section of experiments, we also found that KEGG based on expression profile was enriched in "pancreatic cancer" pathway (P<0.01). It was found that some of the drugs used to treat pancreatic cancer or diseases may cause hepatotoxicity. For example, Erlotinib is a drug used for the treatment of pancreatic cancer. Severe hepatotoxicity was observed in 4% to 31% of patients receiving erlotinib treatment prompting delay or termination of treatment [21]. In the same way, p110γ inhibitor deficiency protects against pancreatic carcinogenesis yet predisposes to diet-induced hepatotoxicity. Related mechanisms and phenomena have also been found in other signaling pathways[21].

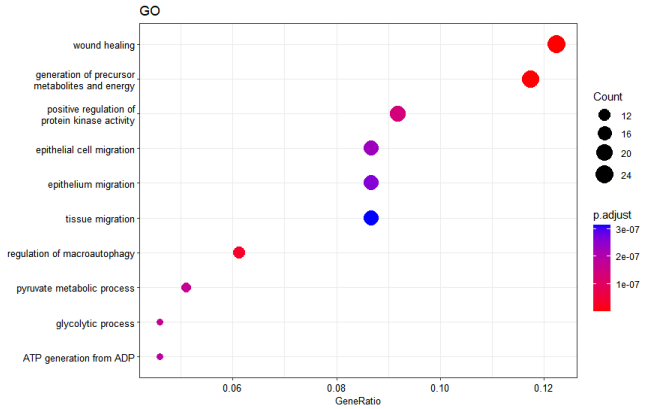

Figure S3 The top 10 GOBP terms enriched with important feature of L1000 expression profile

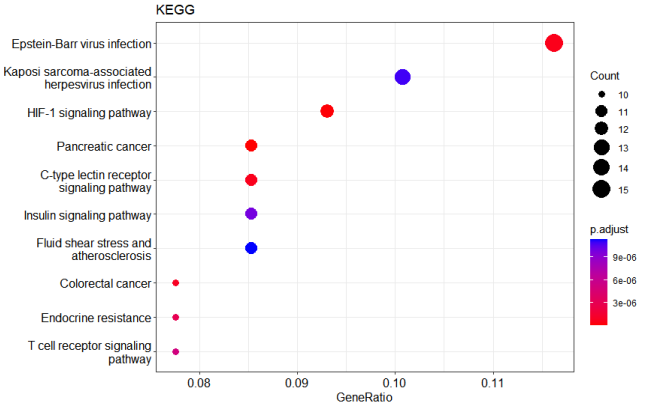

Figure S4 The top 10 KEGG pathways enriched with important feature of L1000 expression profile

GOBP enrichment analysis of target genes was as shown in Figure S5, which was mainly enriched in the "regulation of membrane potential" (P<0.01) pathway. The most important signaling pathways affecting hepatotoxicity are oxidative stress and mitochondrial damage[22-25], and the loss of membrane potential is one of the main causes of the mitochondrial loss, which may lead to

the mitochondrial malfunction[26]. KEGG pathway analysis based on target genes revealed that this important target was mainly enriched in the cAMP signaling pathway. It has been documented that cAMP is a key second messenger molecule that regulates various cellular functions, including inflammation and cell injury, by affecting the expression and function of genes[27].

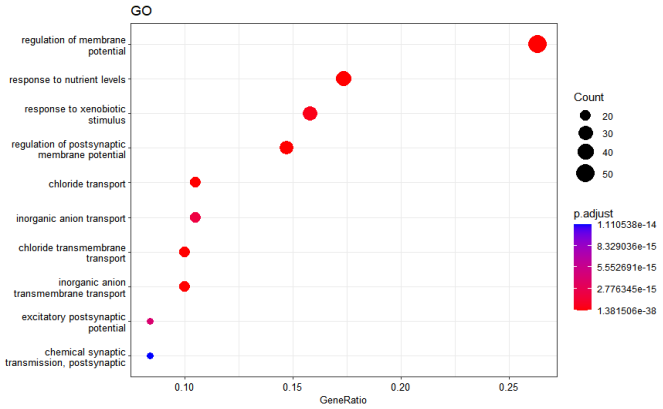

Figure S5 The top 10 GOBP terms enriched with important feature of DrugBank compound related target genes

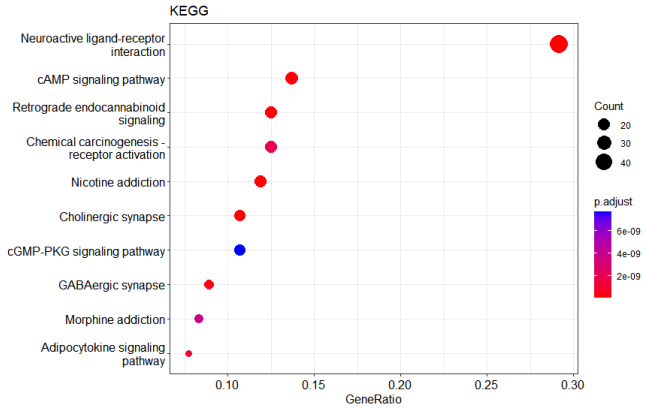

Figure S6 The top 10 GOBP pathways enriched with important feature of DrugBank compound-related target genes

The above enrichment results based on GOBP and KEGG fully demonstrate that the model can explore the potential hepatotoxicity mechanism and the results based on different dimensions of data are coordinated with each other. Our model also provides effective solutions for the treatment of hepatotoxicity. It also potentially illustrates the relationship between disease and toxicity.

Reference

[1] G. Landrumetal, "Rdkit : Open-source cheminformatics," 2016.

[2] Yan B, Ye X, Wang J, et al., "An Algorithm Framework for Drug-Induced Liver Injury Prediction Based on Genetic Algorithm and Ensemble Learning," Molecules, vol. 27, no. 10, p. 3112, 2022.

[3] Jiang D, Wu Z, Hsieh CY, et al., "Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models," J Cheminformatics, vol. 13, no. 1, p. 12, 2021.

[4] He S, Ye T, Wang R, et al., "An In Silico Model for Predicting Drug-Induced Hepatotoxicity," Int J Mol Sci, vol. 20, no. 8, p. 1897, 2019.

[5] Law V, Knox C, Djoumbou Y, et al., "DrugBank 4.0: shedding new light on drug metabolism," Nucleic Acids Res, vol. 42, no. D1, pp. D1091-D1097, 2014.
[6] Subramanian A, Narayan R, Corsello SM, et al., "A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles," Cell, vol. 171, no. 6, pp. 1437-1452, 2017.
[7] McDermott MBA, Wang J, Zhao WN, et al., "Deep learning benchmarks on L1000 gene expression data," IEEE/ACM transactions on computational biology and bioinformatics, vol. 17, no. 6, pp. 1846-1857, 2019.
[8] Li M, Sun X, Zhao J, et al., "CCL5 deficiency promotes liver repair by improving inflammation resolution and liver regeneration through M2 macrophage polarization," Cell Mol Immunol, vol. 17, no. 7, pp. 753-764, 2020.
[9] M. J. Hodgman and A. R. Garrard, "A review of acetaminophen poisoning," Crit Care Clin, vol. 28, no. 4, pp. 499-516, 2012.
[10] E. Yoon, A. Babar, M. Choudhary, M. Kutner, and N. Pyrsopoulos, "Acetaminophen-Induced Hepatotoxicity: a Comprehensive Update," J Clin Transl Hepatol, vol. 4, no. 2, pp. 131-42, 2016.
[11] H. Jaeschke, C. D. Williams, A. Ramachandran, and M. L. Bajt, "Acetaminophen hepatotoxicity and repair: the role of sterile inflammation and innate immunity," Liver Int, vol. 32, no. 1, pp. 8-20, 2012.
[12] H. Jaeschke, "Acetaminophen Hepatotoxicity: Not as Simple as One Might Think! Introductory Comments on the Special Issue-Recent Advances in Acetaminophen Hepatotoxicity," Livers, vol. 2, no. 3, pp. 105-107, 2022.
[13] D. C. Dahlin, G. T. Miwa, A. Y. Lu, and S. D. Nelson, "N-acetyl-p-benzoquinone imine: a cytochrome P-450-mediated oxidation product of acetaminophen," Proc Natl Acad Sci U S A, vol. 81, no. 5, pp. 1327-31, 1984.
[14] H. Jaeschke, L. Duan, J. Y. Akakpo, A. Farhood, and A. Ramachandran, "The role of apoptosis in acetaminophen hepatotoxicity," Food Chem Toxicol, vol. 118, pp. 709-718, 2018.
[15] Tran T, Lee WM., "DILI: New Insights into Diagnosis and Management.," Curr. Hepat. Rep. , vol. 12, pp. 53–58, 2013.
[16] Aboulhoda BE, El-Din SS, Khalifa MM, et al., "Histological, immunohistochemical, and molecular investigation on the hepatotoxic effect of potassium dichromate and the ameliorating role of Persea americana mill pulp extract," Microsc Res Tech, vol. 84, no. 10, pp. 2434-2450, 2021.
[17] Araújo AM, Enea M, Fernandes E, et al., "3,4-Methylenedioxymethamphetamine Hepatotoxicity under the Heat Stress Condition: Novel Insights from in Vitro Metabolomic Studies," J Proteome Res, vol. 19, no. 3, pp. 1222-1234, 2020.
[18] A. Kobayashi, Y. Suzuki, and S. Sugai, "Specificity of transaminase activities in the prediction of drug-induced hepatotoxicity," J Toxicol Sci, vol. 45, no. 9, pp. 515-537, 2020.
[19] Kimura H, Hoshino Y, Kanegane H, et al., "Clinical and virologic characteristics of chronic active Epstein-Barr virus infection," Blood, vol. 98, no. 2, pp. 280-6, 2001.
[20] Wada Y, Sato C, Tomita K, et al., "Possible autoimmune hepatitis induced after chronic active Epstein-Barr virus infection," Clin J Gastroenterol, vol. 7, no. 1, pp. 58-61, 2014.
[21] M. K. Kim, J. Yee, Y. S. Cho, H. W. Jang, J. M. Han, and H. S. Gwak, "Risk factors for erlotinib-induced hepatotoxicity: a retrospective follow-up study," BMC Cancer, vol. 18, no. 1, p. 988, 2018.

[22] L. Guo, H. Gong, T. L. Tang, B. K. Zhang, L. Y. Zhang, and M. Yan, "Crizotinib and Sunitinib Induce Hepatotoxicity and Mitochondrial Apoptosis in L02 Cells via ROS and Nrf2 Signaling Pathway," Front Pharmacol, vol. 12, p. 620934, 2021.

[23] D. S. Umbaugh, N. T. Nguyen, H. Jaeschke, and A. Ramachandran, "Mitochondrial Membrane Potential Drives Early Change in Mitochondrial Morphology After Acetaminophen Exposure," Toxicol Sci, vol. 180, no. 1, pp. 186-195, 2021.

[24] M. R. McGill, M. R. Sharpe, C. D. Williams, M. Taha, S. C. Curry, and H. Jaeschke, "The mechanism underlying acetaminophen-induced hepatotoxicity in humans and mice involves mitochondrial damage and nuclear DNA fragmentation," J Clin Invest, vol. 122, no. 4, pp. 1574-83, 2012.

[25] Larson AM., "Acetaminophen hepatotoxicity," Clin Liver Dis, vol. 11, no, 1, pp.525-48, 2007.

[26] H. M. Ni, J. A. Williams, H. Jaeschke, and W. X. Ding, "Zonated induction of autophagy and mitochondrial spheroids limits acetaminophen-induced necrosis in the liver," Redox Biol, vol. 1, pp. 427-32, 2013.

[27] O'Brien AJ, Fullerton JN, Massey KA, et al., "Immunosuppression in acutely decompensated cirrhosis is mediated by prostaglandin E2. ," Nat Med., vol. 20, no. 5, pp. 518–523, 2014.