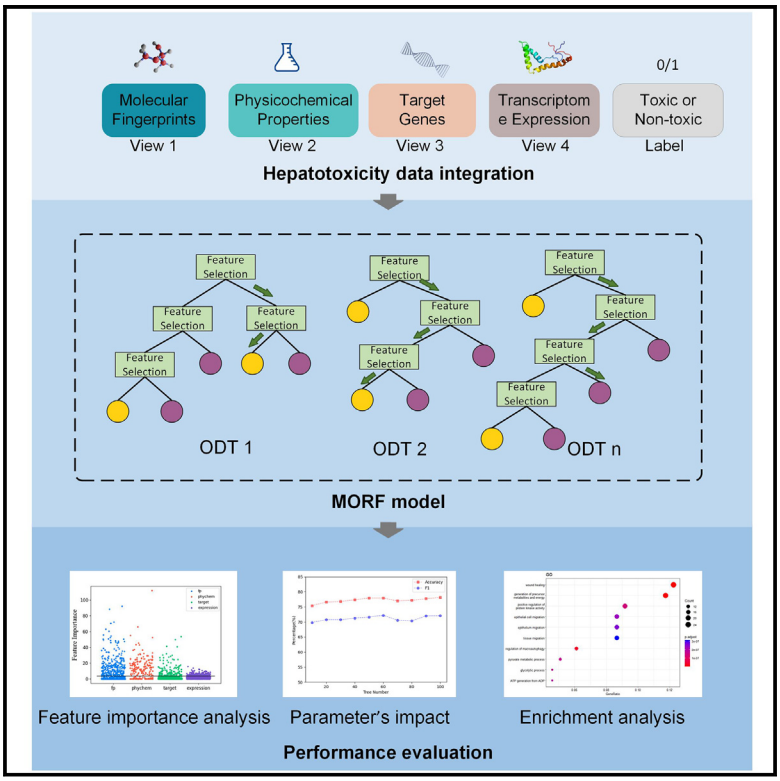
iScience

Article

MORF: Multi-view oblique random forest for hepatotoxicity prediction

Graphical abstract



Highlights

- Two versions of oblique decision trees are proposed
- The algorithms show high performance for the prediction of hepatotoxicity
- The results indicate that the model well explores the mechanism of hepatotoxicity

Authors

Binsheng Sui, Qingzhuo He, Bowei Yan, Kunhong Liu, Yong Xu, Song He, Xiaochen Bo

Correspondence

lkhqz@xmu.edu.cn (K.L.), hes1224@163.com (S.H.), boxc@bmi.ac.cn (X.B.)

In brief

Bioinformatics; Cancer

Sui et al., 2025, iScience *28*, 111755
February 21, 2025
https://doi.org/10.1016/j.isci.2025.111755

CellPress

iScience

CellPress
OPEN ACCESS

Article

MORF: Multi-view oblique random forest for hepatotoxicity prediction

Binsheng Sui,[1,7] Qingzhuo He,[2,7] Bowei Yan,[3] Kunhong Liu,[1,4,8,*] Yong Xu,[5] Song He,[6,*] and Xiaochen Bo[6,*]
[1]Department of Digital Media, Xiamen University, Xiamen 361005, China
[2]Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore
[3]Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China
[4]Xiamen Key Laboratory of Intelligent Storage and Computing, School of Informatics, Xiamen University, Xiamen 361005, China
[5]Xiamen Key Laboratory of Intelligent Fishery, Xiamen Ocean Vocational College, Xiamen 361100, China
[6]Institute of Health Service and Transfusion Medicine, Beijing 100850, China
[7]These authors contributed equally
[8]Lead contact
*Correspondence: lkhqz@xmu.edu.cn (K.L.), hes1224@163.com (S.H.), boxc@bmi.ac.cn (X.B.)
https://doi.org/10.1016/j.isci.2025.111755

SUMMARY

Hepatotoxicity prediction is significant in drug development, so the experts expect to get effective and reliable references. Based on the consideration that the hepatotoxicity data involve multiple types of features, this paper proposes a multi-view oblique random forest (MORF) for hepatotoxicity prediction by considering each type of feature as an independent view. The Householder transformation is employed to get the inclined cut hyperplane in each view. Two versions of the multi-view oblique decision tree (ODT) algorithms were designed by generating optimal nodes based on selecting proper hyperplanes from different views, named ODT-N and ODT-R. These two types of ODT algorithms serve as the base learners to construct two MORF. Experiments conducted on the hepatotoxicity data provide performance comparisons among different algorithms, and the results confirm that our algorithms can fully utilize the information in different views.

INTRODUCTION

Identification of hepatotoxicity is a major hurdle in drug discovery. The liver is vital for detoxification, especially when medications are taken orally. Drug-induced liver injuries contribute to a high rate of drug withdrawals globally. An accurate and efficient hepatotoxicity prediction method is urgently needed to reduce drug attrition rates during the clinical development process. Therefore, researchers have proposed a large number of algorithms to handle this problem.[1]

Recently, different machine learning approaches have been proposed for the hepatotoxicity prediction. For example, Ancuceanu[2] used DILIrank to predict liver toxicity by a variety of machine learning algorithms, and Kim[3] employed several machine learning models to identify risk factors of hepatotoxicity occurrence on patients using nilotinib. As a widely employed model, decision tree (DT)-based methods had been applied to the hepatotoxicity prediction task.[4] However, the exploration of this research topic is still limited. For example, despite the successful applications of DT-based algorithms in hepatotoxicity prediction, it is found that most DT algorithms partition the feature space using axis parallel splits. When the true decision boundaries are not aligned with the feature axes, the traditional DT tends to gain poor performance because of its complicated boundary structure. In contrast, oblique decision trees (ODTs) use oblique decision boundaries to simplify the boundary structure and maintain higher discriminative ability.

Unlike the traditional DT that searches rectangular partitions of the covariate space, the ODT tries to get oblique partitions by maximizing the interclass variance of the linear combination of multiple features.[5] It is an effective way to find a hyperplane that splits a set of examples into groups with minimizing misclassified samples on both sides of the hyperplane for complex problems. Although the ODT provides a new way to better fit the data distribution and reduce the model complexity, it usually requires a higher computational cost.

Random forest (RF) is a classical ensemble framework with fusing multiple DTs obtained by self-sampling. To improve the diversity among different DTs, RF enables the random feature subset selection to train each tree with a subset of attributes. In this way, RF can support high generalization ability by introducing randomness with low computational cost. In addition, RF had also been successfully applied to the hepatotoxicity prediction task.[6]

However, the difficulty in hepatotoxicity prediction lies in that its data contain different types of features, such as fingerprint and physicochemical properties of the compounds. Inspired by the fact that different types of features are not closely correlated, the multi-view learning technique was applied to treat these feature sets as multiple views, to promise high diversity among

Table 1. Basic information on the human hepatotoxicity dataset

Name	Number of samples	Views	Number of features
Hepatotoxicity dataset	521	fp	881
		phychem	200
		target	1,023
		expression	978

base learners in an ensemble. Researchers also put forward some advanced combination methods of views, such as co-training,[7] clustering,[8] and subspace learning,[9] to make the learning methods fit the data characteristics across multiple views.

This study proposes a multi-view oblique random forest (MORF)-based framework for the hepatotoxicity prediction. By treating the multiple types of features as independent views, our algorithm builds ODTs by searching oblique decision boundaries based on the Householder transformation in a view independently, to construct RFs. Experiments are conducted on a hepatotoxicity dataset, and the results confirm the advantages of our MORF. The main contributions of this paper include the following.

(1) Each type of features serves as an independent view, so as to form a multi-view learning scheme. A new oblique decision boundary search strategy is also introduced by setting the Householder space as the optimal hyperplane across multiple views.
(2) Two versions of ODTs are proposed by two different strategies to generate non-leaf nodes in the multiple views.
(3) Two versions of RF algorithms are designed with the random feature subspace technique, providing high performance for the prediction of hepatotoxicity.

RESULTS

In this section, the two algorithms are applied to the human hepatotoxicity dataset and the experimental results are compared with other machine learning algorithms.

The hepatotoxicity datasets

The hepatotoxicity dataset of compounds comes from 7 original datasets: Mulliner,[10] Greene,[11] DILIrank,[12] Xu,[13] and LiverTox.[14] According to The library of integrated network-based cellular signatures (LINCS) 1000 (L1000)[15] database, DrugBank,[16] and Rdkit,[17] the hepatotoxicity dataset containing transcriptome expression profiles (*expression*), target genes (*target*), molecular fingerprints (*fp*), and physicochemical properties (*phychem*) was obtained by matching with the original dataset, as shown in Table 1. The target of the hepatotoxicity dataset is to predict liver toxicity. There are 521 samples in this dataset, including 352 toxic samples and 169 non-toxic samples.

In the experiments, there are the combinations of different types of features to form the multi-view learning scheme, and no. 1 to 11 represent the eleven combinations of different views: *fp* + *phychem*, *fp* + *target*, *fp* + *expression*, *phychem* + *target*, *phychem* + *expression*, *target* + *expression*, *fp* + *phychem* + *target*, *fp* + *phychem* + *expression*, *fp* + *target* + *expression*, *phychem* + *target* + *expression*, and *fp* + *phychem* + *target* + *expression.*

Table 2. The accuracy results of various methods applied to a single view

Views	RF	ET	GBDT	XGB	SVC	KNN	R-E-GA	MORF
fp	77.37	77.17	75.63	78.72	74.66	70.79	76.00	76.98
phychem	76.99	75.63	77.57	77.76	76.99	74.28	75.55	78.72
target	74.86	73.11	75.24	74.08	75.05	72.92	71.25	76.54
expression	64.12	67.25	66.66	65.49	67.84	64.52	63.12	68.61

Table 3. The F1-score results of various methods applied to a single view

Views	RF	ET	GBDT	XGB	SVC	KNN	R-E-GA	MORF
fp	71.44	71.38	70.02	74.00	64.62	66.08	69.60	69.00
phychem	70.17	69.38	72.18	72.14	70.11	69.27	68.35	73.22
target	66.36	66.64	65.04	65.53	63.37	65.02	65.58	69.23
expression	54.15	48.54	52.34	53.83	40.37	43.93	48.05	48.93

Comparison between different models

All algorithms in experiments are tested 10 times by the 5-fold cross-validation. As for our algorithms, the minimum number of samples required for node subdivision is set to 2. Six machine learning algorithms are employed for comparison: RF, ExtraTrees, Gradient Boosting Decision Tree (GBDT), XGBoost, Support Vector Classification (SVC), and K-Nearest Neighbor (KNN). These algorithms are supported by the Scikit-Learn library with the optimal parameter settings. Another algorithm applied to hepatotoxicity prediction, Rotation-Ensemble-GA (R-E-GA),[18] is also used for comparison.

The experiments are conducted on four single views of the human hepatotoxicity dataset. Because only a view is used to construct a DT, MORF-N and MORF-R are the same. As shown in the Tables 2 and 3, MORF outperforms the other algorithms on the *phychem* and *target* views, and the accuracy and F1 are not optimal on the other two views.

Tables 4 and 5 show the experimental results of accuracy and F1-score indices. It is observed that the two versions of MORF algorithms show obvious advantages in accuracy. As for the 11 different combinations, our algorithms achieve the highest accuracy scores in 10 combinations, about 1%–9% higher than other algorithms. MORF algorithms also have obvious advantages in F1-score of human hepatotoxicity dataset and get the highest F1-score indices in 9 out of 11 combinations. While it is found that MORF-R attains better performance in terms of the accuracies scores, MORF-N gets close performance in the F1-score indices.

Contribution of different views

To investigate whether different contributions of various views are reflected in the multi-view algorithm, the contribution rate of each view in the MORF-N algorithm is calculated by selecting a view to generate a node, and the results are shown in Table 6.

It is found that *fp* contributes most among different views, as it is of the highest selection frequency in the tree construction

Table 4. The accuracy results of various methods applied to multiple views

Combinations	RF	ET	GBDT	XGB	SVC	KNN	R-E-GA	MORF-N	MORF-R
1	77.76	76.79	76.21	78.25	75.23	71.95	76.95	78.15	77.95
2	76.20	77.56	76.99	77.75	75.43	72.15	75.65	77.75	77.76
3	65.96	69.06	68.28	69.05	68.09	65.18	64.27	72.15	73.69
4	76.60	78.14	75.63	77.18	76.40	74.86	76.08	78.53	78.16
5	69.24	72.14	72.92	71.57	67.89	65.38	67.54	74.47	74.18
6	67.25	66.27	67.83	66.66	67.84	63.55	64.93	74.26	74.17
7	76.59	77.18	76.99	77.56	75.62	72.54	76.09	78.73	79.02
8	69.83	73.88	71.56	71.76	69.25	66.53	67.15	75.63	76.89
9	67.12	70.60	68.48	69.25	68.28	63.05	65.27	75.05	76.70
10	69.24	72.15	72.73	75.25	67.89	64.01	67.14	76.02	77.96
11	70.79	73.31	72.73	72.73	69.63	65.56	67.94	77.37	78.11

process. When there are 2–3 different types of views, the contribution rates of *fp* view are between 49% and 78%; when there are 4 types of views, the contribution rate of *fp* view also reaches 42.68%. Therefore, it is obvious that *fp* is selected at much higher frequency than the average rate.

In the multi-views involving the *phychem* view, its contribution rates are also higher than others and ranked the second in most cases. In contrast, the contribution rates of *target* are the lowest among all views and even less than 30%.

In short, the contribution rates of different views reflect their importance in the hepatotoxicity prediction.

Comparison between DT and ODT

In order to verify the effectiveness of the ODT, this subsection compares the performance of ODT-N and ODT-R with the traditional Iterative Dichotomiser 3 (ID3) DT, and the results are given in Table 7.

It can be observed from Table 7 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 aforementioned comparison results, it can be seen that ODTs have better classification effects on human hepatotoxicity dataset than DT.

Feature importance analysis

We analyzed the importance of features in each view for MORF-N. This section uses the 11 groups comprising four views to carry out experiments and calculates the weight sum of each feature. The results are shown in Figure 1.

It can be seen from Figure 1 that the feature importance of *fp* and *phychem* view is more than 20, and the feature contribution degree is higher. Among the top 100 in importance ranking, features in *fp* view account for the highest proportion, indicating that not only are *fp* view selected more times but also its main features have a great impact on the prediction results, and view contribution is the highest.

The impact of parameter settings on model performance

Increasing the number of base learners would lead to a higher randomness, which improves the accuracy and generalization performance of the RF algorithm. We therefore explore the impact of the number of ODTs 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

Table 5. The F1-score results of various methods applied to multiple views

Combinations	RF	ET	GBDT	XGB	SVC	KNN	R-E-GA	MORF-N	MORF-R
1	71.34	70.07	70.46	75.22	67.14	66.43	70.59	72.06	70.99
2	70.10	71.52	70.13	73.36	65.34	67.81	70.09	70.71	70.50
3	52.63	57.45	55.39	56.72	43.25	54.21	48.33	58.48	61.02
4	69.98	71.76	68.84	71.61	68.46	70.66	68.98	72.70	71.07
5	58.88	63.60	64.72	63.26	42.12	49.08	53.51	66.54	62.86
6	53.48	50.10	55.31	50.93	40.42	44.97	48.06	63.06	63.74
7	70.45	70.89	71.30	72.57	67.11	68.06	69.43	72.77	72.24
8	57.47	65.18	62.82	63.55	48.49	55.97	52.91	67.27	68.13
9	54.65	60.09	56.94	57.34	43.83	53.40	50.18	64.83	67.54
10	57.29	60.96	64.35	68.30	42.12	48.01	53.01	67.99	69.15
11	59.48	63.75	65.02	65.04	49.08	56.63	53.52	69.85	67.69

Table 6. Contribution rates of different views in MORF-N

Combinations	fp	phychem	target	expression
1	63.10%	36.90%	–	–
2	78.76%	–	21.24%	–
3	62.67%	–	–	37.33%
4	–	79.71%	20.29%	–
5	–	62.24%	–	37.76%
6	–	–	26.26%	73.74%
7	53.43%	30.63%	15.95%	–
8	49.58%	24.46%	–	25.96%
9	51.68%	–	17.84%	30.48%
10	–	51.78%	17.85%	30.37%
11	42.68%	22.71%	14.50%	20.11%

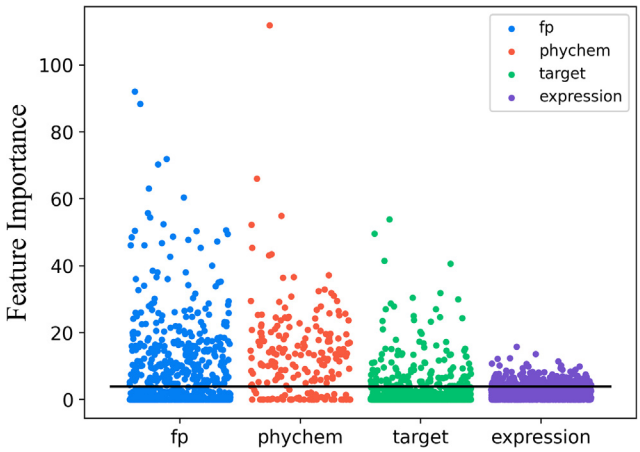


Figure 1. Feature importance of MORF-N

DTs is set from 10 to 100 in 10 groups of equally divided numbers. Figure 2 shows the impact of the number of ODTs on MORF-N and MORF-R.

As can be seen from these figures that 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.

We also conducted experiments on the main parameters that affect XGB and R-E-GA. For XGB, the main parameters to be adjusted are the learning rate and number of estimators for the evolutionary generation of adjusted parameters for R-E-GA. Figures 3 and 4 show the impact of parameter adjustment results on XGB and R-E-GA.

Toxicity mechanisms validation via the enrichment analysis

In order to further explore the toxigenic mechanism of hepatotoxicity caused by compounds, Gene Ontology biological process (GOBP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were extracted based on the important features of the first 200 dimensions of the transcriptome expression profile. Figures 5 and 6 show the biological processes and signaling pathways involved in the hepatotoxicity caused by different compounds.

Table 7. Accuracy and F1 comparisons between DT and ODTs

	Accuracy			F1-score		
Combinations	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

Despite the diverse etiologies of hepatotoxicity, innate immunity activation is a common feature involved in drug-induced hepatotoxicity progression.[19] Among the hepatotoxicity caused by drugs, acetaminophen (APAP)-induced hepatotoxicity is the most common cause of acute liver failure in the United States.[20] APAP-induced hepatotoxicity includes three pathological stages successively: injury initiation, injury amplification, and liver regeneration and repair.[21] As can be seen from Figure 5, a large number of genes were enriched in the GOBP "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, which leads to glutathione depletion and mitochondrial dysfunction, thereby triggering damage.[22]

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 6. Clinical cases have shown that Epstein-Barr virus (EBV) infection can cause various forms of liver enzyme abnormalities and even liver failure,[23] which can also cause autoimmune hepatitis.[24] In this section of experiments, we also found that KEGG based on expression profile was enriched in the "pancreatic cancer" pathway ($p < 0.01$). It was found that some of the drugs used to treat pancreatic cancer or diseases may cause hepatotoxicity.[25] For example, erlotinib is a drug used for the treatment of pancreatic cancer. Severe hepatotoxicity was observed in 4%–31% of patients receiving erlotinib treatment prompting delay or termination of treatment.[26]

The aforementioned 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

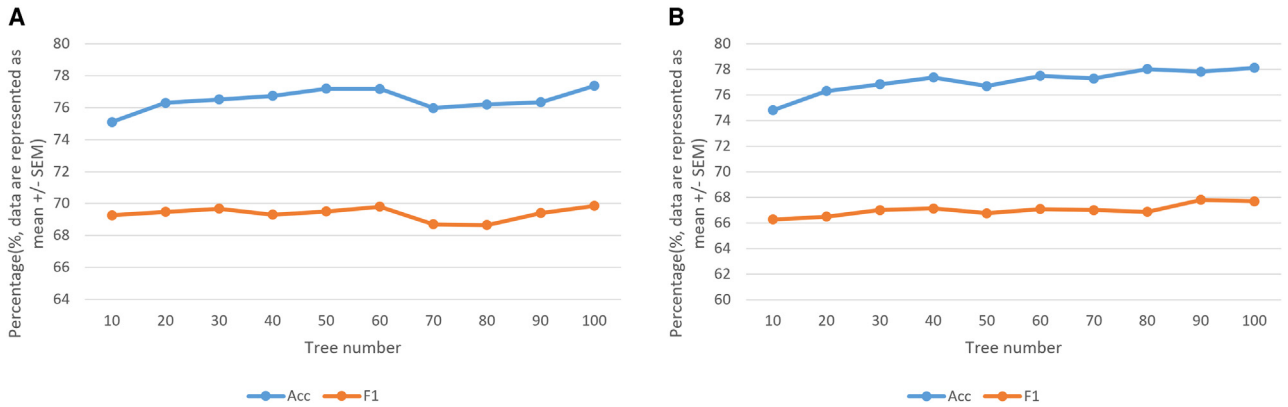


Figure 2. Comparison of parameter adjustment results on MORF-N and MORF-R
(A) Results on MORF-N; (B) results on MORF-R.

hepatotoxicity. It also potentially illustrates the relationship between disease and toxicity.

External dataset validation experiment

We conducted experiments on another binary prediction dataset for drug toxicity. This hepatotoxicity data used were obtained from a comprehensive database of toxicological data and benchmarks (TOXRIC.)[26] This binary class classification dataset comprises four different types of features: drug-induced gene expression profile, Morgan fingerprint, PubChem fingerprint, and drug targets, which comprise the four views too. The detailed descriptions of feature types are presented in the Table 8. There are 630 samples in this dataset, including 380 toxic samples and 250 non-toxic samples.

We conducted experiments on this dataset, and all algorithms in experiments were tested 10 times by the 5-fold cross-validation. All algorithms were optimized for optimal parameter settings. As shown in Table 9, the experimental results of MORF in multiple views of this dataset also demonstrate superior performance. Although the effect did not exceed the original Multi-view uncertainty deep forest (MVU-DF), both MORF-N and MORF-R outperformed other compared algorithms. The MORF framework is shown in Figure 7. The difference in feature selection between two decision trees is shown in Figure 8.

DISCUSSION

In this paper, two versions of the multi-view RF algorithms are proposed for the hepatotoxicity prediction task. By treating different types of features as different views and generating different branches in various views, our algorithm produces two types of multi-view ODTs by conducting the Householder transformation to get optimal axis in optimal views. In this way, the ensemble of multiple oblique trees forms MORF-N and MORF-R, and the difference between them mainly lies in whether an ODT is constructed in multiple views or in a single view. Experimental results show that MORF algorithms have obvious advantages in the human hepatotoxicity dataset. Finally, the enrichment analysis based on targets and transcriptional expression profiles fully demonstrated that this study not only further improved the prediction accuracy of hepatotoxicity but also deeply explored the potential mechanism of hepatotoxicity. In the future, we will try to explore more methods for generating ODTs by utilizing the diversity among different data.

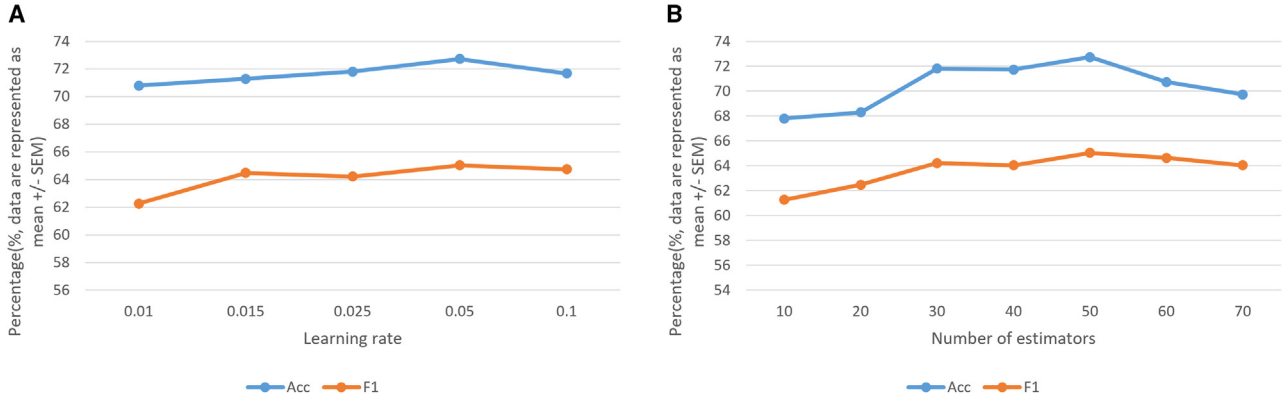


Figure 3. Comparison of parameter adjustment results on XGB
(A) Learning rate adjustment results on XGB; (B) N_estimators adjustment results on XGB.

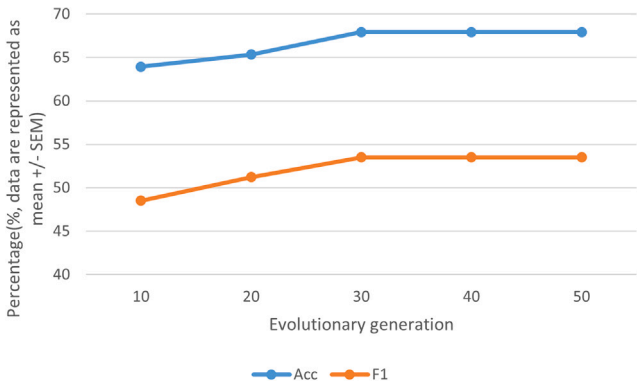


Figure 4. Comparison of parameter adjustment results on R-E-GA

Limitations of the study

Despite the outstanding performance of our proposed model, there are still some limitations. First, while significant efforts have been made to expand our dataset, the sample size is still relatively small. More data will be collected in future studies. Second, our approach did not utilize multimodal data. If multimodal features can be used again for liver toxicity prediction, it may be a future research direction.

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to the lead contact, Dr. Kunhong Liu (lkhqz@xmu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The hepatotoxicity dataset containing transcriptome expression profiles (*expression*), target genes (*target*), molecular fingerprints (*fp*), and physicochemical properties (*phychem*) was obtained by matching with the original dataset, which can be found in the key resources table.
- All original code can be found in GitHub as of the date of publication, which is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2023YFC2604400), the National Natural Science Foundation of China (grant no. 62103436), and The Public technology service platform project of Xiamen City (no. 3502Z20231043).

AUTHOR CONTRIBUTIONS

S.H., X.B., and K.L. conceived the project and supervised the project. B.S. and Q.H. designed the algorithm, conducted computational experiments, and wrote the manuscript. B.Y. drew the pictures. All authors approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

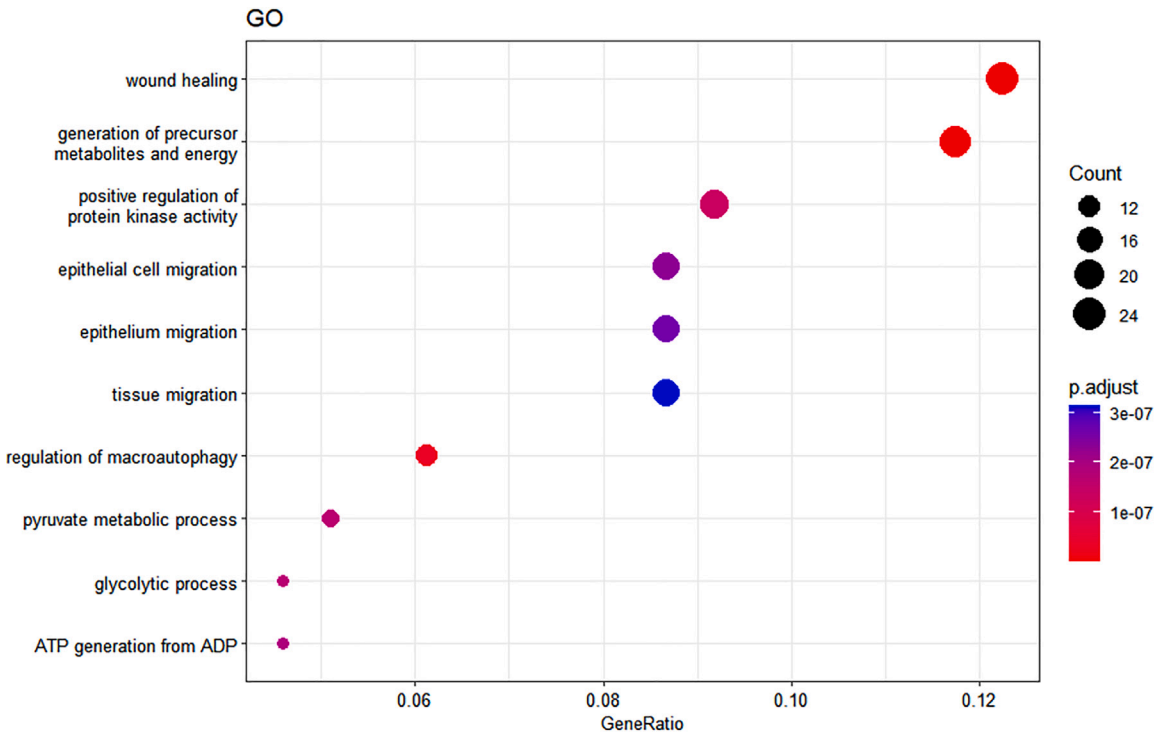


Figure 5. The top 10 GOBP terms enriched with important feature of L1000 expression profile

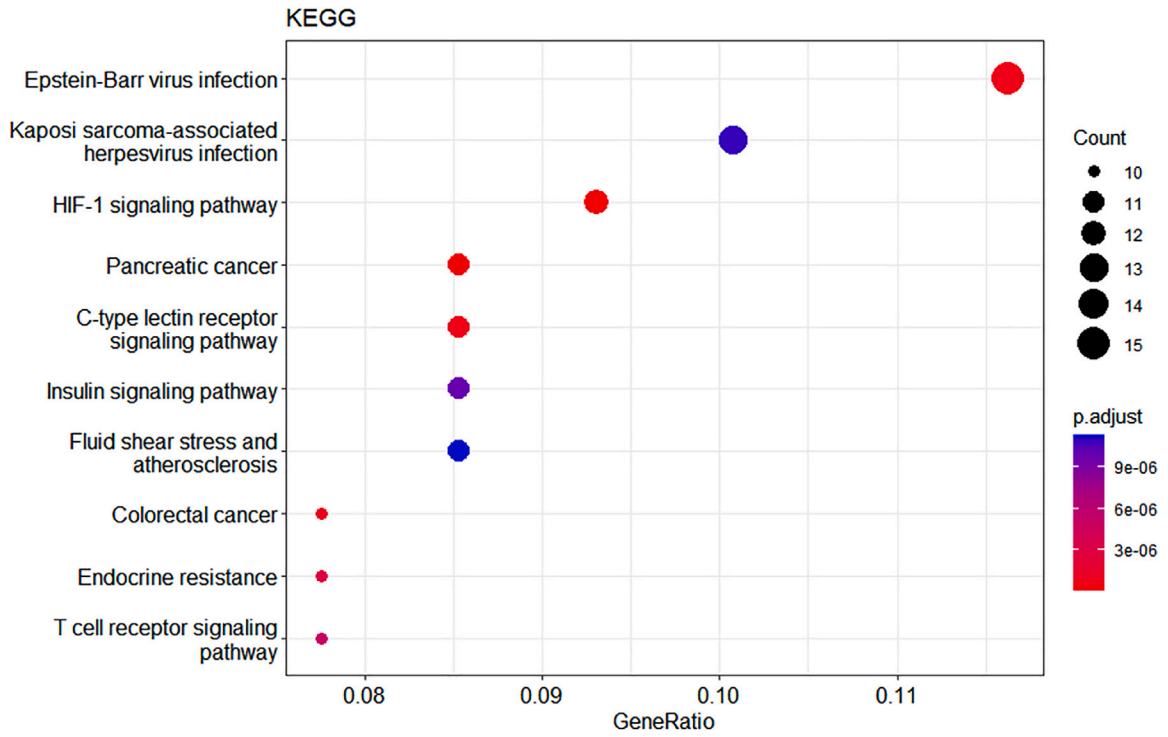


Figure 6. The top 10 KEGG pathways enriched with important feature of L1000 expression profile

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Design of the oblique decision tree
  - Feature subspace selection
  - The framework of two MORF models
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.111755.

Received: August 17, 2024
Revised: October 9, 2024
Accepted: January 3, 2025
Published: January 6, 2025

Table 8. Dataset description

Name	Number of samples	Views	Number of features
Hepatotoxicity dataset	630	Morgan fingerprint	1,024
		PubChem fingerprint	881
		target matrix	3,705
		gene expression	978

REFERENCES

1. Mostafa, F., and Chen, M. (2023). Computational models for predicting liver toxicity in the deep learning era. Front. Toxicol. *5*, 1340860.
2. Ancuceanu, R., Hovanet, M.V., Anghel, A.I., Furtunescu, F., Neagu, M., Constantin, C., and Dinu, M. (2020). Computational Models Using Multiple Machine Learning Algorithms for Predicting Drug Hepatotoxicity with the DILIrank Dataset. Int. J. Mol. Sci. *21*, 2114.
3. Kim, J.S., Han, J.M., Cho, Y.S., Choi, K.H., and Gwak, H.S. (2021). Machine Learning Approaches to Predict Hepatotoxicity Risk in Patients Receiving Nilotinib. Molecules *26*, 3300.
4. Yokoyama, Y., Ono, A., Yoshida, M., Matsumoto, K., and Saito, M. (2021). Refinement of decision tree to assess the consequences of increased serum ALP in dogs: Additional analysis on toxicity studies of pesticides evaluated recently in Japan. Regul. Toxicol. Pharmacol. *124*, 104963.
5. Brodley, C.E., and Utgoff, P.E. (1995). Multivariate Decision Trees. Mach. Learn. *19*, 45–77.
6. Hu, Q., Wang, H., and Xu, T. (2023). Predicting Hepatotoxicity Associated with Low-Dose Methotrexate Using Machine Learning. J. Clin. Med. *12*, 1599.
7. Lu, R.K., Liu, J.W., Wang, Y.F., Xie, H.J., and Zuo, X. (2019). Auto-encoder Based Co-training Multi-view Representation Learning. In 23rd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Macau, Peoples Republic of China, *11441* (Lecture Notes in Artificial Intelligence), pp. 119–130.
8. Chen, Z., Lin, P., Chen, Z., Ye, D., and Wang, S. (2022). Diversity embedding deep matrix factorization for multi-view clustering. Inform. Sci. *610*, 114–125.
9. Fan, R., Luo, T., Zhuge, W., Qiang, S., and Hou, C. (2020). Multi-view subspace learning via bidirectional sparsity. Pattern Recogn. *108*, 107524.

Table 9. Performance comparison between MORF and other models

Method	ACC ± std	F1 ± std
RF	0.6095 ± 0.041	0.4943 ± 0.0222
ET	0.6848 ± 0.0212	0.6477 ± 0.0229
GBDT	0.5381 ± 0.0223	0.5053 ± 0.0232
XGB	0.5705 ± 0.0266	0.5284 ± 0.0145
SVC	0.6032 ± 0.0122	0.3762 ± 0.0165
KNN	0.5937 ± 0.0169	0.5609 ± 0.0198
R-E-GA	0.6031 ± 0.0331	0.4845 ± 0.021
MORF-N	0.7122 ± 0.0251	0.6856 ± 0.0267
MORF-R	0.7052 ± 0.0264	0.6677 ± 0.0274

10. Mulliner, D., Schmidt, F., Stolte, M., Spirkl, H.P., Czich, A., and Amberg, A. (2016). Computational Models for Human and Animal Hepatotoxicity with a Global Application Scope. Chem. Res. Toxicol. *29*, 757–767.

11. Greene, N., Fisk, L., Naven, R.T., Note, R.R., Patel, M.L., and Pelletier, D.J. (2010). Developing Structure-Activity Relationships for the Prediction of Hepatotoxicity. Chem. Res. Toxicol. *23*, 1215–1222.

12. Chen, M., Vijay, V., Shi, Q., Liu, Z., Fang, H., and Tong, W. (2011). FDA-approved drug labeling for the study of drug-induced liver injury. Drug Discov. Today *16*, 697–703.

13. Xu, Y., Dai, Z., Chen, F., Gao, S., Pei, J., and Lai, L. (2015). Deep Learning for Drug-Induced Liver Injury. J. Chem. Inf. Model. *55*, 2085–2093.

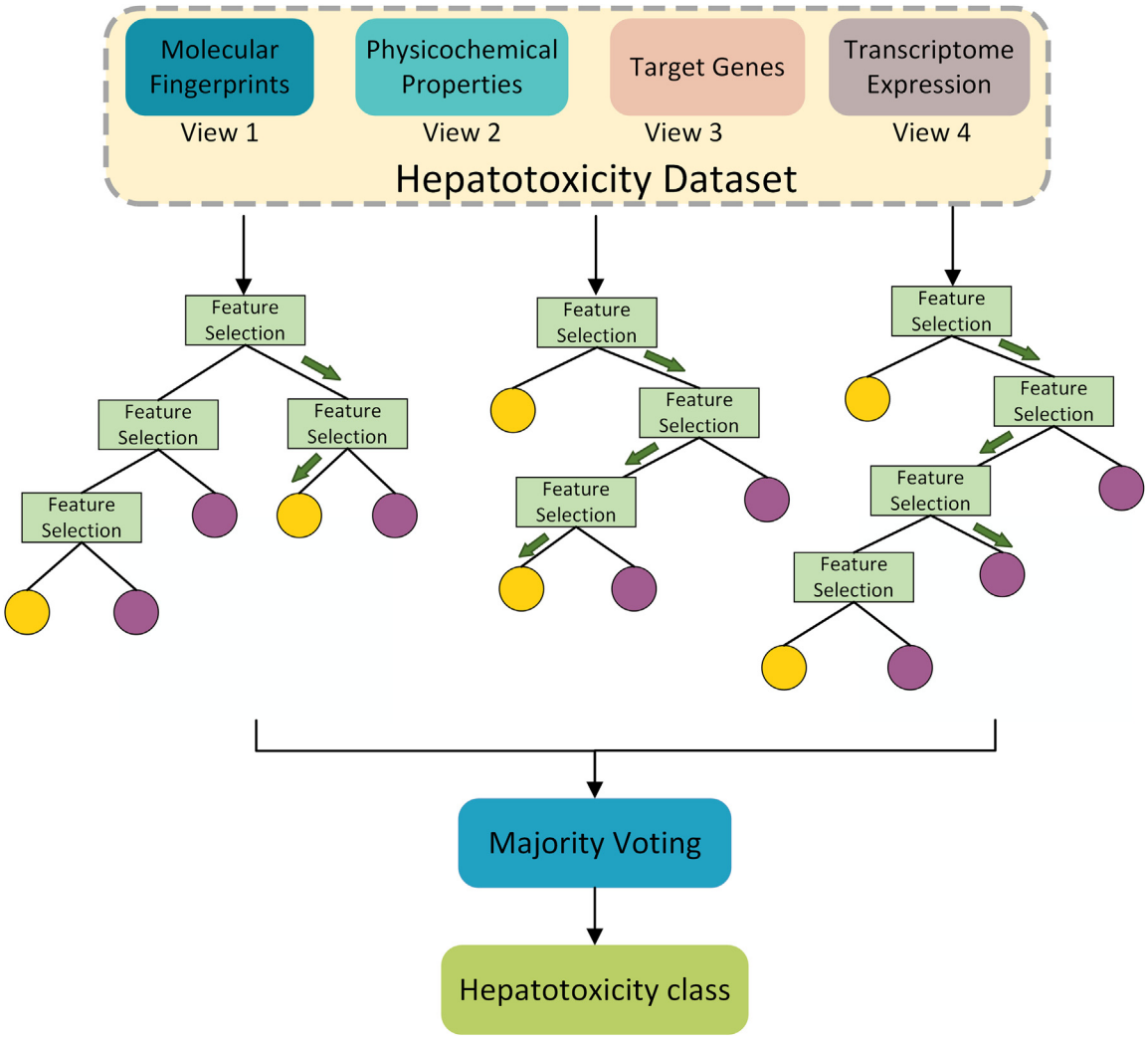


Figure 7. The framework of the MORF model

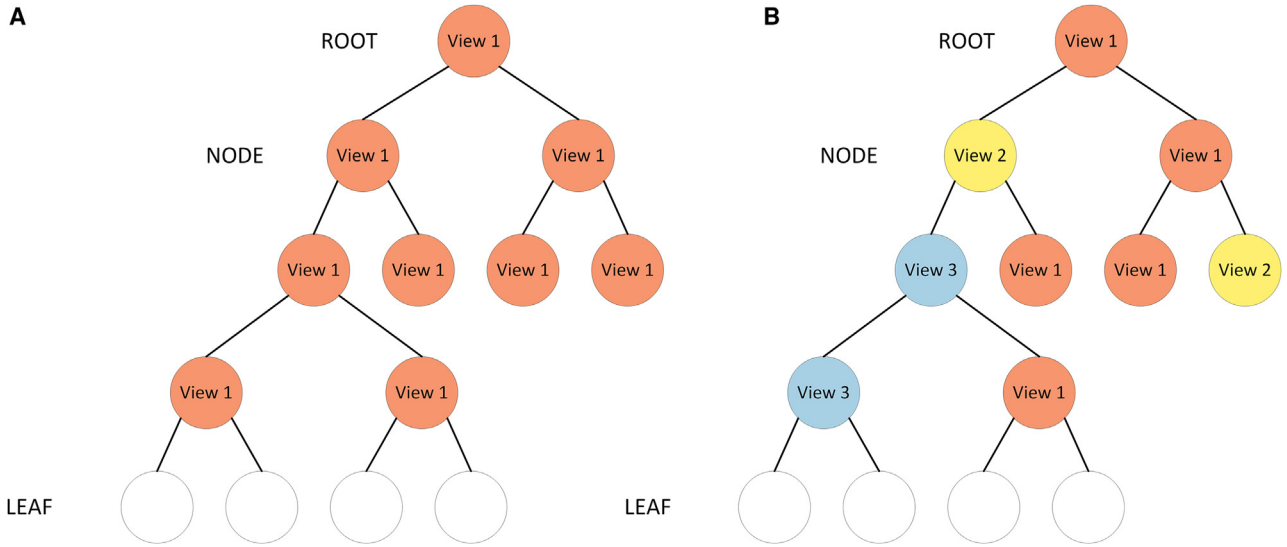


Figure 8. Comparison of two versions of oblique decision trees
(A) A typical tree by ODT-R; (B) a typical tree by ODT-N.

14. Hoofnagle, J.H., Serrano, J., Knoben, J.E., and Navarro, V.J. (2013). LiverTox: A website on drug-induced liver injury. Hepatol. Edit. Mater. *57*, 873–874.
15. Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell *171*, 1437–1452.e17.
16. Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., et al. (2014). DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. *42*, D1091–D1097.
17. Landrumetal, G. (2016). Rdkit : Open-source cheminformatics.
18. Yan, B., Ye, X., Wang, J., Han, J., Wu, L., He, S., Liu, K., and Bo, X. (2022). An Algorithm Framework for Drug-Induced Liver Injury Prediction Based on Genetic Algorithm and Ensemble Learning. Molecules *27*, 3112.
19. Li, M., Sun, X., Zhao, J., Xia, L., Li, J., Xu, M., Wang, B., Guo, H., Yu, C., Gao, Y., et al. (2020). CCL5 deficiency promotes liver repair by improving inflammation resolution and liver regeneration through M2 macrophage polarization. Cell. Mol. Immunol. *17*, 753–764.
20. Prescott, L.F. (2024). Paracetamol (acetaminophen) poisoning: The early years. Br. J. Clin. Pharmacol. *90*, 127–134.
21. Jaeschke, H., Williams, C.D., Ramachandran, A., and Bajt, M.L. (2012). Acetaminophen hepatotoxicity and repair: the role of sterile inflammation and innate immunity. Liver Int. *32*, 8–20.
22. Wang, L., Zhang, Y.W., Zhong, J.J., Zhang, Y., Zhou, S.S., and Xu, C.F. (2022). Mesenchymal Stem Cell Therapy for Acetaminophen-related Liver Injury: A Systematic Review and Meta-analysis of Experimental Studies In Vivo. Curr. Stem Cell Res. Ther. *17*, 825–838.
23. Kimura, H., Hoshino, Y., Kanegane, H., Tsuge, I., Okamura, T., Kawa, K., and Morishima, T. (2001). Clinical and virologic characteristics of chronic active Epstein-Barr virus infection. Blood *98*, 280–286.
24. Bhumi, S.A., and Wu, G.Y. (2023). Seronegative Autoimmune Hepatitis. J. Clin. Transl. Hepatol. Rev. *11*, 459–465.
25. Guo, L., Gong, H., Tang, T.L., Zhang, B.K., Zhang, L.Y., and Yan, M. (2021). Crizotinib and Sunitinib Induce Hepatotoxicity and Mitochondrial Apoptosis in L02 Cells via ROS and Nrf2 Signaling Pathway. Front. Pharmacol. *12*, 620934.
26. Wu, L., Yan, B., Han, J., Li, R., Xiao, J., He, S., and Bo, X. (2023). TOXRIC: a comprehensive database of toxicological data and benchmarks. Nucleic Acids Res. *51*, D1432–D1445.
27. Wickramarachchi, D.C., Robertson, B.L., Reale, M., Price, C.J., and Brown, J. (2016). HHCART: An oblique decision tree. Comput. Stat. Data Anal. *96*, 12–23.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Scikit-learn	Version 0.24.2	https://pypi.org/project/scikit-learn/0.24.2/
Python	Version 3.9.17	https://www.python.org/downloads/release/python-3917/
Matplotlib	Version 3.6.3	https://matplotlib.org/3.6.3/
Data and code	The code for training and test of our model.	https://github.com/MLDMXM2017/MORF

METHOD DETAILS

This section describes the design of MORF algorithm, as shown in Figure 7. The hepatotoxicity dataset containing four views serves as input for MORF, which is composed of multi-view oblique decision trees. Final class result is obtained through majority voting.

Design of the oblique decision tree

Similar to the traditional decision tree, our oblique decision tree tries to search for an optimal partition hyperplane to map original data to a new feature space at a non-leaf node.[27]

Assume that the number of features is m, and p^1 denotes the dominant feature vector of the covariance matrix of class 1. Then there exists a Householder matrix H that makes the following Equation 1 valid:

$$H = I - 2uu^T, where\ u = \frac{e_1 - p^1}{\|e_1 - p^1\|_2}\ and\ e_{1,m\times 1} = (1, 0, \ldots, 0)^T \quad \text{(Equation 1)}$$

At each non-leaf node, data is split based on the direction of the feature vector of each class covariance matrix. It is possible to find a suitable axis parallel partition in the new space, which was tilted in the original space by using the Householder matrix. Then the original dataset S is converted into the new dataset S' by the formula: $S' = H$. In the new space, the hyperplane parallel to the axis can be found along with the coordinate axis. Since Householder matrix H has symmetry and orthogonality, a transformed data in the new space can be reversed to the original space using the property $H \times H^T = I$. The axis found in the new space is parallel split and tilted after being reversed to the original space.

There are three main functions and advantages of using the Householder transform to map the original data to a new space. First, after the Householder transformation, the scope of searching for the best partition hyperplane is expanded, and a better partition hyperplane can be found in a larger search space. Second, the Householder transformation requires lower computational cost. Thirdly, compared with other oblique decision tree algorithms, an optimal partition hyperplane in both the original space and the new transformation space can be obtained at the same time, so as to reduce the complexity of searching for an inclined hyperplane while expanding the search scope.

Feature subspace selection

The random feature subspace method is employed in the division of all non-leaf nodes including the root node, with randomly selecting two-thirds of the features in each view. After the randomness is introduced, the variance values of normalized features based on the new feature subset is calculated, and the top 50% features are kept.

The operation of random selection introduces randomness into the algorithm to reduce the possibility of over-fitting problems and improve the generalization ability of the model. The results of variance statistics can be used to assist feature selection, and some features with small variance values and less connected to the class can be discarded to improve the discriminative ability of our algorithm.

The framework of two MORF models

Our MORF algorithm is proposed by combining the multi-view learning and random feature subspace selection for splitting each non-leaf node. It includes two kinds of ODT algorithms according to the different strategies to construct nodes in multiple views. The structure comparisons between the two ODT algorithms are illustrated in Figure 8. The MORF-ROOT algorithm (MORF-R for short) selects a view from different views at the root node and continues to use it in subsequent partitioning operations. It's base estimator is Oblique Decision Tree-ROOT(ODT-R), as shown in Figure 8A. The MORF-NODE algorithm (MORF-N for short) determines the best view among multiple views for each non-leaf node generation to partition. It's base estimator is Oblique Decision Tree-NODE(ODT-N), as shown in Figure 8B.

In our algorithms, each non-leaf node selects the view that obtains parallel axis division with the maximal information gain in different feature subspaces. A larger information gain indicates the samples contained in the branch nodes is split as far as possible, leading to a higher purity. The information gain is calculated by:

$$Gain(S,h) = Ent(S) - \frac{|S^{h_l}|}{|S|} Ent\left(S^{h_l}\right) - \frac{|S^{h_r}|}{|S|} Ent\left(S^{h_r}\right) \quad \text{(Equation 2)}$$

$$Ent(S) = -\sum_{k=1}^{|y|} p_k \log_2 p_k \quad \text{(Equation 3)}$$

Function $Ent(S)$ represents the information entropy of data subset S, where S^{h_l} and S^{h_r} denote the sample subset assigned to the right and the left children based on the axis h respectively. y represents the number of classes, p_k represents the proportion of the samples of the k-th class in data subset S.

Assume dataset D contains N_D samples. The terminate condition of growing a branch includes: (1) more than l percentage of the samples included in node q belong to a class partition; (2) the number of samples left to be split is smaller than a threshold; (3) the depth of the current node reaches the maximum depth.

Algorithm 1 presents the detailed generation process of ODT-N, beginning from layer 0 to create the root node. Steps 2-5 judge whether a terminal condition is met to produce the leaf node. Steps 6-19 try to find an optimal view from various views for the selected random feature subspace. In this way, each node in the decision tree would be produced in different views to enhance the diversity of a tree. Step 7 randomly selects a feature subset, and Step 8 firstly finds the best axis parallel split h_v^1. Then Steps 9-18 search all eigenvectors of the estimated covariance matrix to get the dominant eigenvector of each class only. At Step 9, for each class, a Householder matrix is constructed for each eigenvector and axis parallel splits are performed along each coordinate axis in the reflected space. e_t is a base vector, p_i is an eigenvector, and τ is a preset parameter. Therefore, $\|e_t - p_i\| > \tau$ determines the parallelism

Algorithm 1. ODT-N (*D*, *V*, *d*)

Input: data set D, the view set V
Output: An oblique decision tree
Initialization: S=D, d=0;

1. Create a null node q for S;
2. If one of the terminal conditions meets **Then**
3. Set q as a leaf node by the majority class k in S;
4. Stop the current loop;
5. End If
6. For each view $v \in V$ **Do**
7. Randomly select 1/3 features f_T sorted by variance in descending order, forming the training data S_v;
8. Get the axis parallel division h_v^1 of S_v according to Equation 1;
9. **For** class $i \in [1, C]$ **Do**
10. Calculate the main eigenvector p_i using PCA;
11. **If** $\|e_t - p_i\| > \tau$ $(\forall t \in T)$ **Then**
12. Get e_t with max $(\|e_t - p_i\|)$;
13. Construct H_t according to Equation 1;
14. $S_{vc} = S_v \times H_t$;
15. Get optimal axis parallel division h_t^i of S_{vc} according to Equations 2 and 3;
16. End If
17. h_v =argmax$(Gain(S,h_t^i), Gain(S,h_v^1))$;
18. End For
19. End For
20. Split S into S^{h_l} and S^{h_r} based on h_v to generate q;
21. Get the left subtree of node q: ODT-N (S^{h_l}, V, d+1);
22. Get the right subtree of node q: ODT-N (S^{h_r}, V, d+1);

Algorithm 2. ODT -R(D, v, d)

Input: data set D, a randomly selected view v;
Output: An oblique decision tree
Initialization: S=D, d=0;

1. Randomly select 1/3 features in v, sort f_T sorted by variance in descending order, forming the training data S_v;
2. Get the axis parallel division h_v^1 of S_v by Equation 1;
3. For class $i \in [1, C]$ **Do**
4. Calculate the main eigenvector p_i using PCA;
5. **If** $\|e_t - p_i\| > \tau$ $(\forall t \in T)$ **Then**
6. Get e_t with max $(\|e_t - p_i\|)$;
7. Construct H_t according to Equation 1;
8. $S_{vc} = S_v \times H_t$;
9. Get optimal axis parallel division h_t^i of S_{vc} according to Equations 2 and 3;
10. End If
11. End For
12. h_v =argmax($Gain(S, h_t^i), Gain(S, h_v^1)$);
13. Split S into S^{h_l} and S^{h_r} based on h_v to generate node q;
14. Obtain the training data for the left-child and right-child nodes according to h_v ;
15. If one of the terminal conditions meets **Then**
16. Set q as a leaf node by the majority class k in S;
17. Stop the current loop;
18. Else
19. Get the left subtree of node q:
ODT-R (S^{h_l}, v, d+1);
20. Get the right subtree of node q:
ODT-R (S^{h_r}, v, d+1);
21. End If

between e_t and p_i. τ is set to 0.05 to filter eigenvectors that are nearly parallel, aiming to accelerate the parallel axis search in the reflected space. In this way, the best axis is obtained to split one class from another.

Step 20 splits S in q by assigning the two subsets, S^{h_l} and S^{h_r}, to the left and right child node based on h_v. By iterating the loops at Steps 21-22, an oblique decision tree ODT-N is grown with splitting data in different views.

In contrast, Algorithm 2 grows each non-leaf node in the oblique decision tree ODT-N similar to that of Algorithm 1. That is, Steps 2-12 searches for the optimal axis h_v and Steps 13-14 assign the two sample subsets to the left and right child node based on h_v. Steps 15-20 determine the iteration for tree generation to fully grow ODT-N tree.

Algorithm 3 works as the ensemble framework to form MORF-N and MORF-R by using ODT-N and ODT-R as base learners respectively. It should be noted without this ensemble, ODT-R only generates an oblique decision tree in a single random view and cannot fully utilize the advantage of the multiple views. The ensemble of ODT-R makes MORF-R include different trees generated in various views, so as to further enhance the diversity compared with the traditional Random Forest. While each ODT-N tree is constructed in multiple views, the ensemble of ODT-N trees (MORF-N) is of higher diversity compared with that of MORF-R, which may lead to higher generalization ability if the base learners are of similar performance.

Algorithm 3. Ensemble of Decision Trees

Input: data set D, ensemble size m
Output: MORF-N or MORF-R

1. For $i \in [1, m]$ **do**
2. Randomly select N_D training samples with replacement from D;
3. Generate an oblique decision tree based on ODT-N (D, V, d) or ODT-R (D, v, d)
4. End For
5. Return the ensemble of m oblique decision trees;

QUANTIFICATION AND STATISTICAL ANALYSIS

We calculated key metrics such as Accuracy and F1 score. We utilized Accuracy to evaluate the overall prediction accuracy of the model. We utilized the F1 score to assess the model's effectiveness in differentiating between positive and negative cases. The F1 score, ranging from 0 to 1, indicates that higher scores reflect the model's ability to correctly identify positive instances while minimizing false positives. The relationships between these parameters and the metrics are defined by the following equations.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{(Equation 4)}$$

$$F1\ score = \frac{2 * TP}{2 * TP+FP+FN} \quad \text{(Equation 5)}$$

ADDITIONAL RESOURCES

This study did not generate additional data.