Hindawi
BioMed Research International
Volume 2020, Article ID 4795140, 10 pages
https://doi.org/10.1155/2020/4795140

*Research Article*

# Machine-Learning Prediction of Oral Drug-Induced Liver Injury (DILI) via Multiple Features and Endpoints

**Xiaobin Liu,[1,2] Danhua Zheng,[3] Yi Zhong,[3] Zhaofan Xia,[1,2] Heng Luo,[2] and Zuquan Weng[2,3]**

[1] *Department of Burns, Changhai Hospital, Second Military Medical University, Shanghai, China*
[2] *The Centre for Big Data Research in Burns and Trauma, Fuzhou University, Fujian Province, China*
[3] *College of Biological Science and Engineering, Fuzhou University, Fujian Province, China*

Correspondence should be addressed to Zhaofan Xia; xiazf@fzu.edu.cn, Heng Luo; hengluo88@gmail.com, and Zuquan Weng; wengzq@fzu.edu.cn

Received 21 January 2020; Accepted 17 April 2020; Published 19 May 2020

Academic Editor: Despina Deligianni

Drug discovery is a costly process which usually takes more than 10 years and billions of dollars for one successful drug to enter the market. Despite all the safety tests, drugs may still cause adverse reactions and be restricted in use or even withdrawn from the market. Drug-induced liver injury (DILI) is one of the major adverse drug reactions, and computational models may be used to predict and reduce it. To assess the computational prediction performance of DILI, we curated DILI endpoints from three databases and prepared drug features including chemical descriptors, therapeutic classifications, gene expressions, and binding proteins. We trained machine-learning models to predict the various DILI endpoints using different drug features. Using the optimal feature sets, the top-performing models obtained areas under the receiver operating characteristic curve (AUC) around 0.8 for some DILI endpoints. We found that some features, including therapeutic classifications and proteins, have good prediction performance towards DILI. We also discovered that the severity of DILI endpoints as well as the selection of negative samples may significantly affect the prediction results. Overall, our study provided a comprehensive collection, curation, and prediction of DILI endpoints using various drug features, which may help the drug researchers to better understand and prevent DILI during the drug discovery process.

## 1. Introduction

The drug discovery process is both time-consuming and costly. It typically takes 10-17 years and costs $2.6 billion to develop a new drug [1, 2]. Even after a drug passes all the clinical trials and enters the market, it can still cause adverse drug reactions, which may result in restricted uses or even withdrawal [3, 4]. In the history of drug development, drug-induced liver injury (DILI) is one of the major factors to cause withdrawal of new drugs [5–7]. As an effort to reduce DILI, researchers have developed computational models to predict it [8, 9]. Machine learning is a method that utilizes computing systems to learn from the data and make predictions without the need of explicit programming [10]. Various machine-learning algorithms have been used to predict DILI, including $k$-nearest neighbor (KNN) [11, 12], Bayesian models [13, 14], linear discriminant analysis (LDA) [15], random forest (RF) [11, 16], support vector machine (SVM) [11], and artificial neural networks(ANN) [15]. Since predicting DILI may help to improve drug safety and reduce loss, this field is attracting interests from both the academia and the pharmaceutical industry.

However, predicting DILI is a challenging task since DILI involves different types of mechanisms such as direct hepatotoxicity, immune reactions, and mechanisms that are not completely understood [17, 18]. Besides, there are several limitations regarding the current approaches of DILI prediction. First, many studies focused on predicting either a single DILI endpoint or a superset of endpoints such as liver enzyme disorders, cytotoxic injury, cholestasis and jaundice, bile duct disorders [19], and liver steatosis [20]. Second, many studies focused on drug structural features [9, 12, 21, 22],

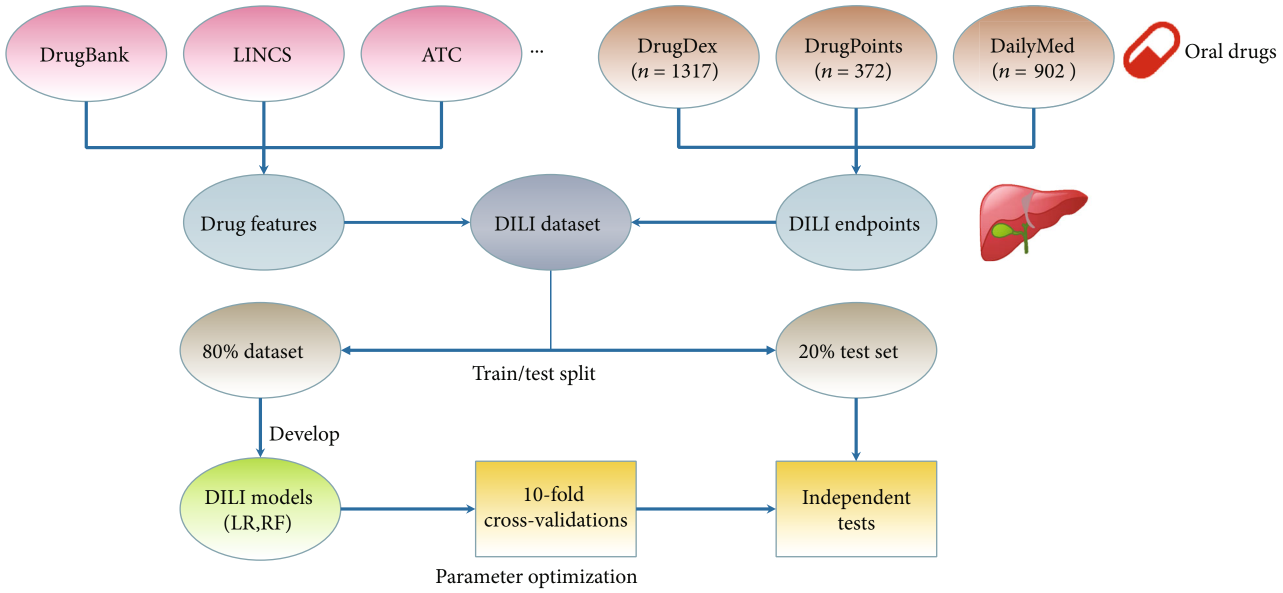


Figure 1: The workflow of this study. We collected drug features from various databases including DrugBank, LINCS, and WHO's ATC database and curated DILI labels from DrugDex, DrugPoints, and DailyMed for oral drugs. We split 20% of the dataset as an independent test set and used the remaining 80% for ten-fold cross-validations. We generated or collected the drug features and developed two types of models, logistic regression (LR) and random forest (RF), using different combinations of parameters and used the best parameters for independent tests.

while many additional types of data, such as binding assays [23], genomics [11], and postmarket surveillance data [19], are available. In this study, we collected a comprehensive dataset across different label sources (Micromedex DrugDex, Micromedex DrugPoints, and DailyMed), different feature types (chemical structure, protein binding, gene expression, and therapeutic classifications), and different DILI endpoints (such as liver failure, jaundice, biomarker increase, hepatomegaly, and hepatitis) for oral drugs. We investigated and evaluated model performance using different features to predict various DILI endpoints. We believe our results provide useful insights regarding DILI prediction and may potentially help to improve drug safety.

## 2. Methods

*2.1. Feature Collection and Processing.* The workflow of this study is shown in Figure 1. We collected multiple types of drug features from a variety of databases. The molecular weights and structures (SMILES format) of the drug molecules were collected from the PubChem database [24]. For structural features, we calculated five types of molecular descriptors including constitutional descriptors, electronic descriptors, geometrical descriptors, hybrid descriptors, and topological descriptors and three types of commonly used chemical fingerprints, including ECFP6 (1024 bits), PubChem fingerprints (881 bits), and standard fingerprints (1024 bits) using the rcdk package [25]. We collected the Anatomical Therapeutic Chemical (ATC) classification and Defined Daily Dose (DDD) codes from the World Health Organization (WHO). For protein binding features, the drug targets, enzymes, transporters, and carriers were collected from the DrugBank database [26]. For gene expression features, the drug-induced gene expression data for 978 landmark genes were collected from Wang et al. [27] based on the NIH Library of Integrated Network-Based Cellular Signatures (LINCS) database.

For feature processing, we categorized some continuous features into bins referring previous studies [28]. For example, the drug daily doses (DDD) were binned into DDD < 10 mg, 10 mg ≤ DDD < 100 mg, and DDD ≥ 100 mg. The solubility AlogP values were grouped into AlogP < 1, 1 ≤ AlogP < 3, and AlogP ≥ 3.

*2.2. Endpoint Data Collection.* The relationship between oral drugs and different types of DILI endpoints was extracted and curated from three databases, DrugDex, DrugPoints, and DailyMed, referring the extraction methods and criteria from previous studies [28]. For DrugDex, we extracted seven types of hepatic adverse drug reaction (hADR) endpoints including fatal hADRs, hADRs causing acute liver failure (liver failure), hADRs resulting in liver transplantation (liver transplantation), jaundice, biomarker increase, hepatomegaly, and hepatitis. The seven hADR endpoints were then categorized into severe hADRs (including fatal hADRs, liver failure, liver transplantation, and hADRs complying with Hy's law [29]) and less severe hADRs (including the rest hADRs). We ended up collecting 1,317 drugs from DrugDex for the above DILI endpoints (Supplementary Table 1). For DrugPoints, we collected endpoints including fatal hADRs, liver failure, jaundice, liver enzymes abnormal, bilirubin, hepatomegaly, and hepatitis for 372 drugs (Supplementary Table 2). The seven endpoints were also categorized into severe hADRs (including fatal hADRs and liver failure) and less severe hADRs (including the rest hADRs). For DailyMed, drugs were categorized into three groups: most concern, less concern, and no concern regarding DILI outcomes [30]. A drug is categorized as most concern for DILI when it was withdrawn from the market or given a warning, such as a black box warning or a precaution section of DILI; a drug is considered less concern for DILI if its label mentioned other DILI risks less severe than the previous criteria; and a drug with no concern for DILI does not have a DILI-related description in its label. We collected 902 drugs and 104, 235, and 563 of these drugs were categorized as most concern, less concern, and no concern for DILI, respectively (Supplementary Table 3).

For each endpoint, we defined two types of negative samples, NSap1 and NSap2. For a given hADR endpoint, NSap1 is defined as drugs that have no reported hepatotoxic reaction for the specific endpoint, while NSap2 is defined as drugs that have no reported hepatotoxic reaction across all endpoints. According to these definitions, NSap2 is a “cleaner” subset of NSap1.

*2.3. Model Training and Assessment.* For each dataset, we randomly held 20% as an independent test set and used the remaining 80% for training and validation. In this study, we trained two classifiers, logistic regression and random forest, using the scikit-learn package in Python. To minimize the data imbalance problem, the “class weight” parameter of each model was set to “balanced.” For each classifier, the best model parameters were selected by grid search based on areas under the receiver operating characteristic curve (AUC) during 10-fold cross-validations. Then, the model with the best parameters was evaluated on the independent test set.

Since we have two types of negative samples, NSap1 and NSap2, to find out whether the two types of negative samples had an impact on the model performance, we performed paired $t$-tests on the AUC scores of all features. We also ran paired $t$-tests specifically for the protein and ATC code features to find out whether they had any impact on the model performance.

## 3. Results and Discussion

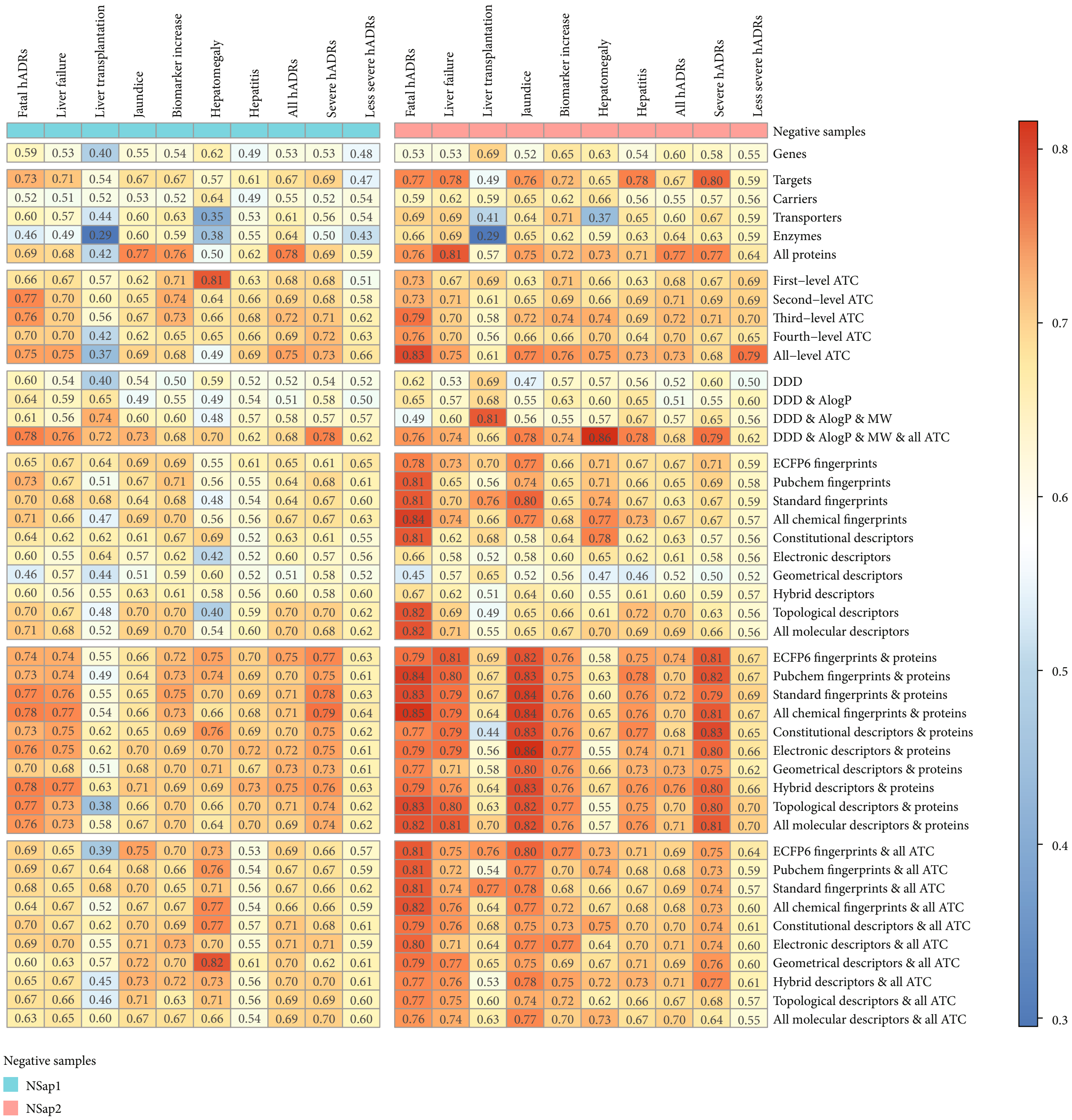
*3.1. Different Features and Model Performance.* We trained two types of classifiers, logistic regression and random forest, to predict different DILI endpoints using different types of features for drugs in the DrugDex, DrugPoints, and DailyMed databases. 10-fold cross-validations and independent tests were conducted to estimate model performance on the three databases. The AUC values of 10-fold cross-validations on the datasets using best parameters were visualized by heat map in Figure 2 and Supplementary Figs. 1–5. The results of the independent tests are in Supplementary Tables 4-6. Since some endpoints have very few or zero positive samples during the independent test and produced abnormally high or zero AUC values, we focused our analysis based on the results of 10-fold cross-validations and provided the independent test results as additional references in Supplementary Tables 4-6.

Like the previous study [31], we used different types of chemical fingerprints to predict DILI. While the logistic regression models showed random performance ($\mathrm{AUC} = 0.5$) on most endpoints using chemical fingerprints as features, the models got slightly better performance for the “All hADR” endpoint on either the NSap1 or NSap2 dataset with AUC values mostly larger than 0.6 (Supplementary Figure 1). For random forest models, the performance is generally better than logistic regression models using chemical fingerprints, especially for endpoints like fatal hADRs and severe hADRs, which have AUC values close to 0.8 (Figure 2). Similar results were also found for endpoints in DrugPoints and DailyMed. Since random forest is an ensemble model with a more complex structure, it is expected that it exceeded the performance of logistic regression. The models showed similar performance patterns using molecular descriptors as features, with a few exceptions.

ATC codes are hierarchical therapeutic classifications of drugs. A previous study has identified associations between drug indications and side effects [32]; thus, we assumed that the therapeutic classifications might also be helpful in predicting DILI. From the results, we can see that ATCs have better performance for predicting most DILI endpoints compared to chemical fingerprints. The logistic regression and random forest models using the second level to fourth level of ATC codes were able to obtain AUC values around or larger than 0.7 in most DILI endpoints. However, the first level of ATC codes had worse performance due to a lack of therapeutic classification details. We also combined ATC codes with other features, including the chemical fingerprints and molecular descriptors. We found that the combination generally improved the model performance than using a single type of features, indicating the usefulness of combining various types of features (Figure 2 and Supplementary Figs. 1-5).

According to the DILIN prospective study [33], drugs in specific categories may have a higher association with DILI, as the authors indicated 45% of the 899 investigated DILI cases were caused by antimicrobials. To find out if similar patterns can be observed in our data, we took drugs collected from DrugDex as an example and calculated the odds ratio (OR) and Fisher’s exact test $p$ values between their top-level ATC codes and different DILI endpoints. The results are shown in Supplementary Table 7. We observed that for anti-infective drugs for systemic use, their odds ratios against all DILI endpoints are above 2.5 with $p$ values < 0.01, indicating a significant positive association. We also analyzed the feature importance for prediction (Supplementary Table 8) and found this category was relatively important to predict various DILI endpoints, which is consistent with the previous study. Additionally, we observed that antineoplastic and immunomodulating agents and drugs for the musculoskeletal system may also have a higher association with DILI compared to drugs in other categories. We believe such data and analysis can provide valuable information to understand and prevent DILI.

The gene expression features used in this study [27] represent gene expression changes of the LINC L1000 978 landmark genes aggregated from a variety of cell lines before and after treatment by drugs. The results showed that their AUC values ranged mostly between 0.5 and 0.6 in all three databases. This indicates that the processed dataset of LINCS gene expression profiles may not be good enough to predict



| | Fatal hADRs | Liver failure | Liver transplantation | Jaundice | Biomarker increase | Hepatomegaly | Hepatitis | All hADRs | Severe hADRs | Less severe hADRs | Fatal hADRs | Liver failure | Liver transplantation | Jaundice | Biomarker increase | Hepatomegaly | Hepatitis | All hADRs | Severe hADRs | Less severe hADRs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Negative samples | 🟦 | 🟦 | 🟦 | 🟦 | 🟦 | 🟦 | 🟦 | 🟦 | 🟦 | 🟦 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 |
| Genes | 0.59 | 0.53 | 0.40 | 0.55 | 0.54 | 0.62 | 0.49 | 0.53 | 0.53 | 0.48 | 0.53 | 0.53 | 0.69 | 0.52 | 0.65 | 0.63 | 0.54 | 0.60 | 0.58 | 0.55 |
| Targets | 0.73 | 0.71 | 0.54 | 0.67 | 0.67 | 0.57 | 0.61 | 0.67 | 0.69 | 0.47 | 0.77 | 0.78 | 0.49 | 0.76 | 0.72 | 0.65 | 0.78 | 0.67 | 0.80 | 0.59 |
| Carriers | 0.52 | 0.51 | 0.52 | 0.53 | 0.52 | 0.64 | 0.49 | 0.55 | 0.52 | 0.54 | 0.59 | 0.62 | 0.59 | 0.65 | 0.62 | 0.66 | 0.56 | 0.55 | 0.57 | 0.56 |
| Transporters | 0.60 | 0.57 | 0.44 | 0.60 | 0.63 | 0.35 | 0.53 | 0.61 | 0.56 | 0.54 | 0.69 | 0.69 | 0.41 | 0.64 | 0.71 | 0.37 | 0.65 | 0.60 | 0.67 | 0.59 |
| Enzymes | 0.46 | 0.49 | 0.29 | 0.60 | 0.59 | 0.38 | 0.55 | 0.64 | 0.50 | 0.43 | 0.66 | 0.69 | 0.29 | 0.65 | 0.62 | 0.59 | 0.63 | 0.64 | 0.63 | 0.59 |
| All proteins | 0.69 | 0.68 | 0.42 | 0.77 | 0.76 | 0.50 | 0.62 | 0.78 | 0.69 | 0.59 | 0.76 | 0.81 | 0.57 | 0.75 | 0.72 | 0.73 | 0.71 | 0.77 | 0.77 | 0.64 |
| First–level ATC | 0.66 | 0.67 | 0.57 | 0.62 | 0.71 | 0.81 | 0.63 | 0.68 | 0.68 | 0.51 | 0.73 | 0.67 | 0.69 | 0.63 | 0.71 | 0.66 | 0.63 | 0.68 | 0.67 | 0.69 |
| Second–level ATC | 0.77 | 0.70 | 0.60 | 0.65 | 0.74 | 0.64 | 0.66 | 0.69 | 0.68 | 0.58 | 0.73 | 0.71 | 0.61 | 0.65 | 0.69 | 0.66 | 0.69 | 0.71 | 0.69 | 0.69 |
| Third–level ATC | 0.76 | 0.70 | 0.56 | 0.67 | 0.73 | 0.66 | 0.68 | 0.72 | 0.71 | 0.62 | 0.79 | 0.70 | 0.58 | 0.72 | 0.74 | 0.74 | 0.69 | 0.72 | 0.71 | 0.70 |
| Fourth–level ATC | 0.70 | 0.70 | 0.42 | 0.62 | 0.65 | 0.65 | 0.66 | 0.69 | 0.72 | 0.63 | 0.76 | 0.70 | 0.56 | 0.66 | 0.66 | 0.70 | 0.64 | 0.70 | 0.67 | 0.65 |
| All–level ATC | 0.75 | 0.75 | 0.37 | 0.69 | 0.68 | 0.49 | 0.69 | 0.75 | 0.73 | 0.66 | 0.83 | 0.75 | 0.61 | 0.77 | 0.76 | 0.75 | 0.73 | 0.73 | 0.68 | 0.79 |
| DDD | 0.60 | 0.54 | 0.40 | 0.54 | 0.50 | 0.59 | 0.52 | 0.52 | 0.54 | 0.52 | 0.62 | 0.53 | 0.69 | 0.47 | 0.57 | 0.57 | 0.56 | 0.52 | 0.60 | 0.50 |
| DDD & AlogP | 0.64 | 0.59 | 0.65 | 0.49 | 0.55 | 0.49 | 0.54 | 0.51 | 0.58 | 0.50 | 0.65 | 0.57 | 0.68 | 0.55 | 0.63 | 0.60 | 0.65 | 0.51 | 0.55 | 0.60 |
| DDD & AlogP & MW | 0.61 | 0.56 | 0.74 | 0.60 | 0.60 | 0.48 | 0.57 | 0.58 | 0.57 | 0.57 | 0.49 | 0.60 | 0.81 | 0.56 | 0.55 | 0.57 | 0.67 | 0.57 | 0.65 | 0.56 |
| DDD & AlogP & MW & all ATC | 0.78 | 0.76 | 0.72 | 0.73 | 0.68 | 0.70 | 0.62 | 0.68 | 0.78 | 0.62 | 0.76 | 0.74 | 0.66 | 0.78 | 0.74 | 0.86 | 0.78 | 0.68 | 0.79 | 0.62 |
| ECFP6 fingerprints | 0.65 | 0.67 | 0.64 | 0.69 | 0.69 | 0.55 | 0.61 | 0.65 | 0.61 | 0.65 | 0.78 | 0.73 | 0.70 | 0.77 | 0.66 | 0.71 | 0.67 | 0.67 | 0.71 | 0.59 |
| Pubchem fingerprints | 0.73 | 0.67 | 0.51 | 0.67 | 0.71 | 0.56 | 0.55 | 0.64 | 0.68 | 0.61 | 0.81 | 0.65 | 0.56 | 0.74 | 0.65 | 0.71 | 0.66 | 0.65 | 0.69 | 0.58 |
| Standard fingerprints | 0.70 | 0.68 | 0.68 | 0.64 | 0.68 | 0.48 | 0.54 | 0.64 | 0.67 | 0.60 | 0.81 | 0.70 | 0.76 | 0.80 | 0.65 | 0.74 | 0.67 | 0.63 | 0.67 | 0.59 |
| All chemical fingerprints | 0.71 | 0.66 | 0.47 | 0.69 | 0.70 | 0.56 | 0.56 | 0.67 | 0.67 | 0.63 | 0.84 | 0.74 | 0.66 | 0.77 | 0.68 | 0.77 | 0.73 | 0.67 | 0.67 | 0.57 |
| Constitutional descriptors | 0.64 | 0.62 | 0.62 | 0.61 | 0.67 | 0.69 | 0.52 | 0.63 | 0.61 | 0.55 | 0.81 | 0.62 | 0.68 | 0.58 | 0.64 | 0.78 | 0.62 | 0.63 | 0.57 | 0.56 |
| Electronic descriptors | 0.60 | 0.55 | 0.64 | 0.57 | 0.62 | 0.42 | 0.52 | 0.60 | 0.57 | 0.56 | 0.66 | 0.58 | 0.52 | 0.58 | 0.60 | 0.65 | 0.62 | 0.61 | 0.58 | 0.56 |
| Geometrical descriptors | 0.46 | 0.57 | 0.44 | 0.51 | 0.59 | 0.60 | 0.52 | 0.51 | 0.58 | 0.52 | 0.45 | 0.57 | 0.65 | 0.52 | 0.56 | 0.47 | 0.46 | 0.52 | 0.50 | 0.52 |
| Hybrid descriptors | 0.60 | 0.56 | 0.55 | 0.63 | 0.61 | 0.58 | 0.56 | 0.60 | 0.58 | 0.60 | 0.67 | 0.62 | 0.51 | 0.64 | 0.60 | 0.55 | 0.61 | 0.60 | 0.59 | 0.57 |
| Topological descriptors | 0.70 | 0.67 | 0.48 | 0.70 | 0.70 | 0.40 | 0.59 | 0.70 | 0.70 | 0.62 | 0.82 | 0.69 | 0.49 | 0.65 | 0.66 | 0.61 | 0.72 | 0.70 | 0.63 | 0.56 |
| All molecular descriptors | 0.71 | 0.68 | 0.52 | 0.69 | 0.70 | 0.54 | 0.60 | 0.70 | 0.68 | 0.62 | 0.82 | 0.71 | 0.55 | 0.65 | 0.67 | 0.70 | 0.69 | 0.69 | 0.66 | 0.56 |
| ECFP6 fingerprints & proteins | 0.74 | 0.74 | 0.55 | 0.66 | 0.72 | 0.75 | 0.70 | 0.75 | 0.77 | 0.63 | 0.79 | 0.81 | 0.69 | 0.82 | 0.76 | 0.58 | 0.75 | 0.74 | 0.81 | 0.67 |
| Pubchem fingerprints & proteins | 0.73 | 0.74 | 0.49 | 0.64 | 0.73 | 0.74 | 0.69 | 0.70 | 0.75 | 0.61 | 0.84 | 0.80 | 0.67 | 0.83 | 0.75 | 0.63 | 0.78 | 0.70 | 0.82 | 0.67 |
| Standard fingerprints & proteins | 0.77 | 0.76 | 0.55 | 0.65 | 0.75 | 0.70 | 0.69 | 0.71 | 0.78 | 0.63 | 0.83 | 0.79 | 0.67 | 0.84 | 0.76 | 0.60 | 0.76 | 0.72 | 0.79 | 0.69 |
| All chemical fingerprints & proteins | 0.78 | 0.77 | 0.54 | 0.66 | 0.73 | 0.66 | 0.68 | 0.71 | 0.79 | 0.64 | 0.85 | 0.79 | 0.64 | 0.84 | 0.76 | 0.65 | 0.76 | 0.70 | 0.81 | 0.67 |
| Constitutional descriptors & proteins | 0.73 | 0.75 | 0.62 | 0.65 | 0.69 | 0.76 | 0.69 | 0.70 | 0.75 | 0.62 | 0.77 | 0.79 | 0.44 | 0.83 | 0.76 | 0.67 | 0.77 | 0.68 | 0.83 | 0.65 |
| Electronic descriptors & proteins | 0.76 | 0.75 | 0.62 | 0.70 | 0.69 | 0.70 | 0.72 | 0.72 | 0.75 | 0.61 | 0.79 | 0.79 | 0.56 | 0.86 | 0.77 | 0.55 | 0.74 | 0.71 | 0.80 | 0.66 |
| Geometrical descriptors & proteins | 0.70 | 0.68 | 0.51 | 0.68 | 0.70 | 0.71 | 0.67 | 0.73 | 0.73 | 0.61 | 0.77 | 0.71 | 0.58 | 0.80 | 0.76 | 0.66 | 0.73 | 0.73 | 0.75 | 0.62 |
| Hybrid descriptors & proteins | 0.78 | 0.77 | 0.63 | 0.71 | 0.69 | 0.69 | 0.73 | 0.75 | 0.76 | 0.63 | 0.79 | 0.76 | 0.64 | 0.83 | 0.76 | 0.67 | 0.76 | 0.76 | 0.80 | 0.66 |
| Topological descriptors & proteins | 0.77 | 0.73 | 0.38 | 0.66 | 0.70 | 0.66 | 0.70 | 0.71 | 0.74 | 0.62 | 0.83 | 0.80 | 0.63 | 0.82 | 0.77 | 0.55 | 0.75 | 0.70 | 0.80 | 0.70 |
| All molecular descriptors & proteins | 0.76 | 0.73 | 0.58 | 0.67 | 0.70 | 0.64 | 0.70 | 0.69 | 0.74 | 0.62 | 0.82 | 0.81 | 0.70 | 0.82 | 0.76 | 0.57 | 0.76 | 0.71 | 0.81 | 0.70 |
| ECFP6 fingerprints & all ATC | 0.69 | 0.65 | 0.39 | 0.75 | 0.70 | 0.73 | 0.53 | 0.69 | 0.66 | 0.57 | 0.81 | 0.75 | 0.76 | 0.80 | 0.77 | 0.73 | 0.71 | 0.69 | 0.75 | 0.64 |
| Pubchem fingerprints & all ATC | 0.69 | 0.67 | 0.64 | 0.68 | 0.66 | 0.76 | 0.54 | 0.67 | 0.67 | 0.59 | 0.81 | 0.72 | 0.54 | 0.77 | 0.70 | 0.74 | 0.68 | 0.68 | 0.73 | 0.59 |
| Standard fingerprints & all ATC | 0.68 | 0.65 | 0.68 | 0.70 | 0.65 | 0.71 | 0.56 | 0.67 | 0.66 | 0.62 | 0.81 | 0.74 | 0.77 | 0.78 | 0.68 | 0.66 | 0.67 | 0.69 | 0.74 | 0.57 |
| All chemical fingerprints & all ATC | 0.64 | 0.67 | 0.52 | 0.67 | 0.67 | 0.77 | 0.54 | 0.66 | 0.66 | 0.59 | 0.82 | 0.76 | 0.64 | 0.77 | 0.72 | 0.67 | 0.68 | 0.68 | 0.73 | 0.60 |
| Constitutional descriptors & all ATC | 0.70 | 0.67 | 0.62 | 0.70 | 0.69 | 0.77 | 0.57 | 0.71 | 0.68 | 0.61 | 0.79 | 0.76 | 0.68 | 0.75 | 0.73 | 0.75 | 0.70 | 0.70 | 0.74 | 0.61 |
| Electronic descriptors & all ATC | 0.69 | 0.70 | 0.55 | 0.71 | 0.73 | 0.70 | 0.55 | 0.71 | 0.71 | 0.59 | 0.80 | 0.71 | 0.64 | 0.77 | 0.77 | 0.64 | 0.70 | 0.71 | 0.74 | 0.60 |
| Geometrical descriptors & all ATC | 0.60 | 0.63 | 0.57 | 0.72 | 0.70 | 0.82 | 0.61 | 0.70 | 0.62 | 0.61 | 0.79 | 0.77 | 0.65 | 0.75 | 0.69 | 0.67 | 0.71 | 0.69 | 0.76 | 0.60 |
| Hybrid descriptors & all ATC | 0.65 | 0.67 | 0.45 | 0.73 | 0.72 | 0.73 | 0.56 | 0.70 | 0.70 | 0.61 | 0.77 | 0.76 | 0.53 | 0.78 | 0.75 | 0.72 | 0.73 | 0.71 | 0.77 | 0.61 |
| Topological descriptors & all ATC | 0.67 | 0.66 | 0.46 | 0.71 | 0.63 | 0.71 | 0.56 | 0.69 | 0.69 | 0.60 | 0.77 | 0.75 | 0.60 | 0.74 | 0.72 | 0.62 | 0.66 | 0.67 | 0.68 | 0.57 |
| All molecular descriptors & all ATC | 0.63 | 0.65 | 0.60 | 0.67 | 0.67 | 0.66 | 0.54 | 0.69 | 0.70 | 0.60 | 0.76 | 0.74 | 0.63 | 0.77 | 0.70 | 0.73 | 0.67 | 0.70 | 0.64 | 0.55 |

Negative samples

🟦 NSap1

🟥 NSap2

Figure 2: AUC values of different sets of features and DILI endpoints using random forest for drugs in the DrugDex database during 10-fold cross-validations. In the table, each row represents a set of drug features, each column represents a DILI endpoint and the negative sample set (NSap1 vs. NSap2), and each cell represents an AUC value (colored by its value). For DrugDex, there are seven DILI endpoints (fatal hADRs, liver failure, liver transplantation, jaundice, biomarker increase, hepatomegaly, and hepatitis). They were categorized as "severe hADRs" and "less severe hADRs.". "All hADRs" include all DILI endpoints.

DILI, possibly because the immortal cell lines in which drugs were tested may not necessarily represent the specific cell types of hepatocytes or liver tissues. Thus, the expression profiles aggregated from these experiments may not be predictive towards DILI endpoints.

To explore the importance of protein features in predicting DILI, we trained models to predict various DILI endpoints using drug-binding proteins including targets, carriers, transporters, and enzymes. We found that using a single type of protein features alone, the models obtained various results with the highest AUC value around 0.8. Meanwhile, combining all types of protein features could improve model performance even more. Additionally, we found combining the protein features with the chemical fingerprints or molecular descriptors could significantly improve the performance of just using chemical fingerprints or molecular descriptors in most cases of DrugDex and DrugPoints and some cases of DailyMed (Table 1). This

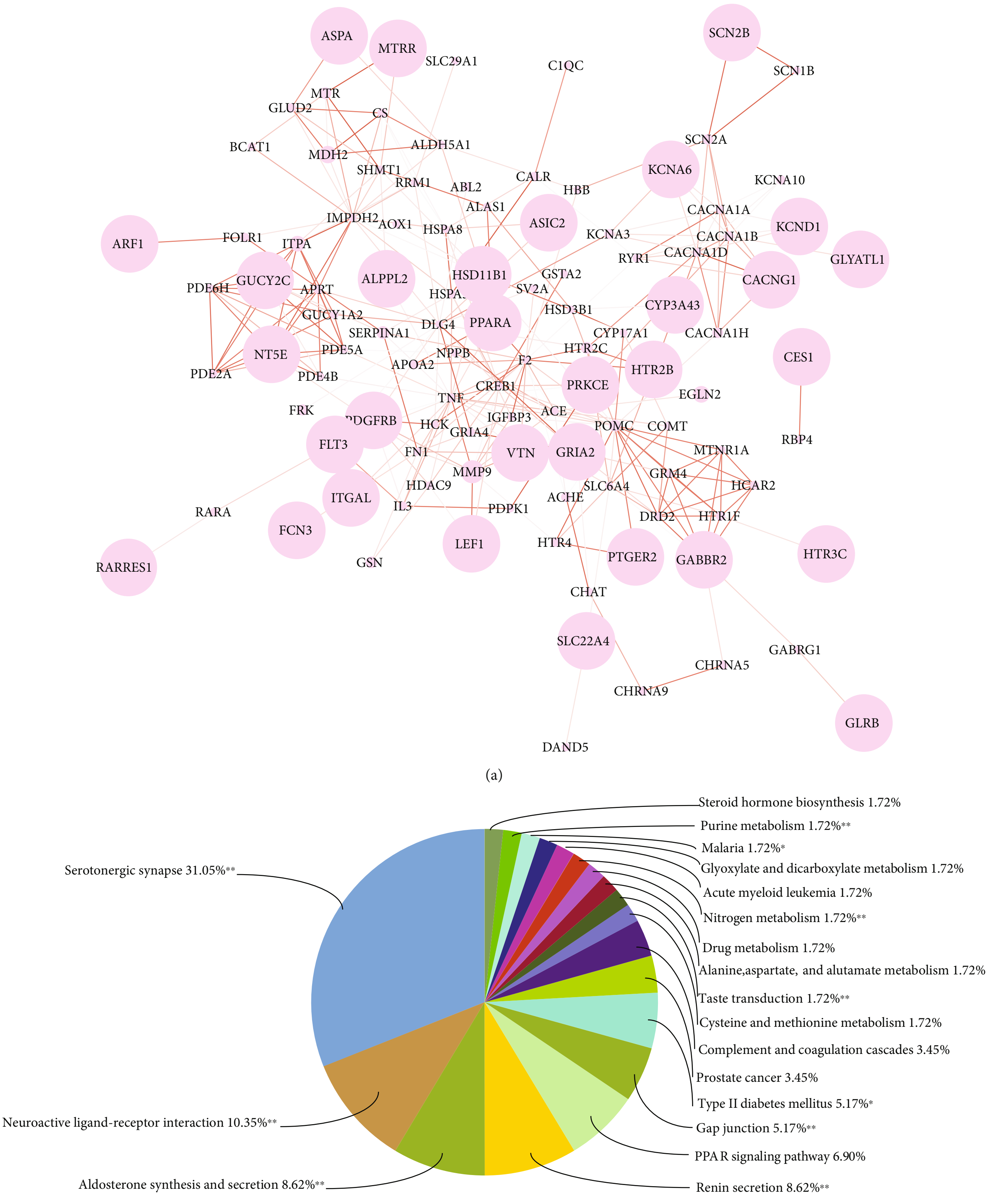
Table 1: Paired $t$-test results of AUC values during 10-fold cross-validations with or without using protein-binding features.

| | | Logistic regression | | Random forest | |
|---|---|---|---|---|---|
| Database | Features | $t$ | $p$ | $t$ | $p$ |
| DrugDex | ECFP6 fingerprints | -3.51 | 1.96E-03** | -2.48 | 1.80E-02* |
| | PubChem fingerprints | -3.09 | 5.38E-03** | -2.56 | 1.48E-02* |
| | Standard fingerprints | -3.32 | 2.86E-03** | -2.26 | 2.94E-02* |
| | Constitutional descriptors | -2.12 | 4.35E-02* | -2.96 | 5.41E-03** |
| | Electronic descriptors | -4.44 | 1.14E-04** | -6.10 | 7.04E-07** |
| | Geometrical descriptors | -5.75 | 4.22E-06** | -8.30 | 6.47E-10** |
| | Hybrid descriptors | -3.50 | 1.90E-03** | -8.79 | 5.96E-10** |
| | Topological descriptors | -2.35 | 2.43E-02* | -1.93 | 6.11E-02 |
| | All fingerprints | -2.34 | 2.68E-02* | -1.94 | 5.95E-02 |
| | All descriptors | -2.63 | 1.29E-02* | -2.48 | 1.78E-02* |
| | All combined | -10.25 | 2.76E-21** | -10.56 | 3.79E-23** |
| DrugPoints | ECFP6 fingerprints | -2.06 | 5.60E-02 | -2.99 | 8.91E-03** |
| | PubChem fingerprints | -3.26 | 9.78E-03** | 0.10 | 9.19E-01 |
| | Standard fingerprints | -2.66 | 2.10E-02* | -2.49 | 2.51E-02* |
| | Constitutional descriptors | -3.20 | 4.97E-03** | -2.18 | 4.28E-02* |
| | Electronic descriptors | -3.31 | 5.00E-03** | -3.51 | 2.98E-03** |
| | Geometrical descriptors | -5.42 | 4.06E-05** | -5.21 | 6.70E-05** |
| | Hybrid descriptors | -4.80 | 9.79E-04** | -2.31 | 3.55E-02* |
| | Topological descriptors | -4.04 | 8.19E-04** | -3.04 | 7.08E-03** |
| | All fingerprints | -2.41 | 2.75E-02* | -2.03 | 5.80E-02* |
| | All descriptors | -4.61 | 3.56E-04** | -2.35 | 3.08E-02* |
| | All combined | -10.13 | 2.42E-19** | -7.30 | 1.04E-11** |
| DailyMed | ECFP6 fingerprints | -0.79 | 4.50E-01 | -0.31 | 7.62E-01 |
| | PubChem fingerprints | -2.24 | 7.56E-02 | -0.35 | 7.37E-01 |
| | Standard fingerprints | 0.00 | 1.00E+00 | -0.85 | 4.19E-01 |
| | Constitutional descriptors | -0.94 | 3.80E-01 | -1.56 | 1.53E-01 |
| | Electronic descriptors | -1.25 | 2.58E-01 | -1.65 | 1.30E-01 |
| | Geometrical descriptors | -2.10 | 8.66E-02 | -4.80 | 7.95E-04** |
| | Hybrid descriptors | -2.81 | 3.74E-02* | -1.49 | 1.79E-01 |
| | Topological descriptors | -0.27 | 7.97E-01 | -0.26 | 8.00E-01 |
| | All fingerprints | 0.10 | 9.26E-01 | -0.23 | 8.24E-01 |
| | All descriptors | -0.90 | 3.97E-01 | -0.56 | 5.87E-01 |
| | All combined | -3.16 | 2.06E-03** | -2.88 | 4.74E-03** |

For each $t$-test, the AUC score vectors of model performance on all endpoints were paired up and compared. $^{*}p < 0.05$; $^{**}p < 0.01$.

indicates the protein-binding profiles of drugs are potentially important indicators for DILI. Liu et al. [34] investigated the prediction of adverse drug reactions using chemical features, protein features, and phenotypic properties of drugs. They also found that the combination of both protein features and chemical features improved the prediction performance compared to using only one of them. As one family of adverse drug reactions, DILI has idiosyncratic and complicated mechanisms [18]. Since protein features provide important target-binding information in addition to chemical features, we believe the combination of such multidimensional data can improve the model prediction performance.

*3.2. Network and Pathway Analysis of Protein Features.* In this section, we did network and pathway analyses of the protein features using the DrugDex database as an example. To find out which proteins and pathways are important to DILI prediction, we calculated the Gini importance values for the protein features using ExtraTrees [35]. For each endpoint, we selected proteins with feature importance equal or larger



(a)

(b)

Figure 3: For fatal hADRs as the endpoint, (a) the network of proteins according to the feature importance and (b) KEGG pathway analysis of important protein features. In (a), each protein is represented by its gene symbol. The node size represents feature importance of protein to DILI models. The line thickness presents the combined score made by the STRING database. In (b), the important protein features were selected and analyzed by Cytoscape ClueGO using KEGG pathways. The stars indicate the significance levels for the enrichment tests.

Table 2: Paired $t$-test results of AUC values during 10-fold cross-validations between severe hADRs and less severe hADRs using NSap2 as negative examples.

| | Logistic regression | | Random forest | |
|---|---|---|---|---|
| Database | $t$ | $p$ | $t$ | $p$ |
| DrugDex | 2.51 | 1.77$E$-02∗ | 3.72 | 8.13$E$-04∗∗ |
| DrugPoints | 3.36 | 1.92$E$-03∗∗ | 1.73 | 9.18$E$-02 |
| DailyMed | -0.07 | 9.45$E$-01 | 5.16 | 2.41$E$-05∗∗ |

For each endpoint, the AUC score vectors of model performance on all features were paired up and compared. $^{*}p < 0.05$; $^{**}p < 0.01$.

than 0.001 and queried the STRING database [36] to find the protein-protein associations among them. The protein-protein association networks are visualized in Figure 3(a) and Supplementary Figure 6 indicating protein-protein binding, coexistence in the same functional pathway/process, or other indirect interactions. From Figure 3(a), we found that some highlighted genes, such as PPARA, HTR2B, and SLC22A4, were reported in the literature to be associated with DILI or liver diseases [37–39]. We believe this feature analysis may provide helpful insights to identify potential DILI-related genes and generate new hypotheses to be further tested in the wet lab.

We also used the ClueGO plugin in Cytoscape [40, 41] to explore which pathways are enriched among the proteins passing our feature importance criteria (Figure 3(b) and Supplementary Figure 6). We found that the serotonergic synapse pathway was significantly enriched for fatal hADRs and the dopaminergic synapse pathway was significantly enriched for a few other DILI endpoints. Studies showed that serotonin and dopamine may have an association with neuropsychiatric symptoms and neurobiology of liver failure [42, 43]. From our analysis, we believe the feature importance analysis and pathway enrichment analysis may help to generate new hypotheses and useful insights for the DILI mechanisms and thus aid in the understanding and prevention of DILI.

*3.3. Different Endpoints and Model Performance.* We compared the AUC values of all the features between the endpoints of severe hADRs and less severe hADRs and found the models mostly performed better on severe hADRs (Table 2). We also observed better performance on endpoints of fatal hADRs and liver failure compared to other endpoints (Figure 2 and Supplementary Figs. 1-5). It is suggested that these severe DILI endpoints are more predictable than less severe endpoints. Interestingly, as an exception, the jaundice endpoint which belongs to less severe hADRs was found to be predicted well using protein features. This finding is consistent with a previous study which showed the importance of transporters in the cholestasis model [44].

*3.4. Negative Sample Selection and Model Performance.* To elucidate the differences of selecting negative samples in DILI model performance, we prepared two types of negative drugs in three databases, NSap1 and NSap2. In general, the models performed better using NSap2 as negative samples compared to NSap1 (Figure 2 and Supplementary Figs 1-5). Paired $t$-test results of the AUC values in each endpoint between NSap1 and NSap2 are shown in Table 3. We found that for most endpoints in DrugDex, using NSap2 as negative samples had better results than using NSap1. Thus, the selection of negative samples could make a significant difference in predicting DILI endpoints.

Defining an accurate negative set is important to study DILI; however, different sources may lead to different negative sets. Zhu and Li [45] identified a set of 957 drugs without hepatotoxicity report from eHealthMe websites as the negative set, which was also used in the work of Bajzelj and Drgan [46]. DILIrank [47] contains a negative set of 312 no-DILI-concern drugs whose labels did not contain any DILI indication, and this set was later used in the study of Shin et al. [48]. He et al. [49] collected a negative set of 709 drugs without hepatotoxicity records from various literature sources. Note that all the above approaches are similar to our approach, which is to define drugs without reported hepatotoxic reaction as the negative set. However, since different research groups utilized different sources to determine their negative sets, it can be challenging to find a consistent gold standard. Taking DILIrank [47] as an example, while 38% of its no-DILI-concern drugs also exist in our negative set collected from DrugDex, a lower proportion (31%) was found in the negative set from Zhu and Li [45].

## 4. Conclusions

In this study, we collected different types of drug features, including chemical fingerprints, molecular descriptors, binding proteins, gene expression, and therapeutic classifications, and collected the DILI endpoints from three databases, DrugDex, DrugPoints, and DailyMed. We trained machine-learning models to predict the DILI endpoints using the various features. The models were assessed via 10-fold cross-validations, and the results were analyzed by different types of features and endpoints. We found that

(1) the features of ATC codes or binding proteins may have significant implications for prediction performance. Analyzing the important protein features using networks and pathways may elicit potential insights regarding DILI mechanisms

(2) severe liver injury, such as fetal hADRs, severe hADRs, and liver failure, had better prediction performance compared to nonsevere endpoints

Table 3: Paired $t$-test results of AUC values during 10-fold cross-validations between NSap1 and NSap2 as negative examples.

| | | Logistic regression | | Random forest | |
|---|---|---|---|---|---|
| Database | Features | $t$ | $p$ | $t$ | $p$ |
| DrugDex | Fatal hADRs | -3.80 | 7.69E-04** | -2.83 | 7.53E-03** |
| | Liver failure | -3.33 | 2.46E-03** | -1.51 | 1.40E-01 |
| | Liver transplantation | -2.33 | 2.63E-02* | -2.50 | 1.69E-02* |
| | Jaundice | -3.10 | 4.04E-03** | -3.69 | 1.01E-03** |
| | Biomarker increase | -2.76 | 9.05E-03** | -0.59 | 5.60E-01 |
| | Hepatomegaly | -0.35 | 7.28E-01 | -0.72 | 4.77E-01 |
| | Hepatitis | -3.15 | 3.52E-03** | -3.00 | 4.70E-03** |
| | All hADRs | -0.12 | 9.02E-01 | -0.03 | 9.78E-01 |
| | Severe hADRs | -3.65 | 1.06E-03** | -0.68 | 5.00E-01 |
| | Less severe hADRs | -2.74 | 9.73E-03** | -0.58 | 5.65E-01 |
| DrugPoints | Liver failure | -0.82 | 4.20E-01 | 0.42 | 6.75E-01 |
| | Jaundice | -0.11 | 9.15E-01 | 1.18 | 2.47E-01 |
| | All hADRs | -0.81 | 4.21E-01 | 0.04 | 9.67E-01 |
| | Severe hADRs | -1.37 | 1.78E-01 | -0.03 | 9.74E-01 |
| | Less severe hADRs | 0.85 | 4.01E-01 | -0.41 | 6.81E-01 |
| DailyMed | All hADRs | 0.00 | 1.00E+00 | 0.00 | 1.00E+00 |
| | Severe hADRs | 5.22 | 6.75E-06** | -0.60 | 5.50E-01 |
| | Less severe hADRs | 1.41 | 1.72E-01 | 10.04 | 1.57E-10** |

For each endpoint, the AUC score vectors of model performance on all features were paired up and compared. $^{*}p < 0.05$; $^{**}p < 0.01$.

(3) the selection of negative samples had an impact on DILI prediction. Clean negative samples of drugs without any DILI information in their labels may produce better performance for DILI predictions

We also provided all the curated DILI labels from three databases. We believe our study provides valuable information and comprehensive evaluations for computational DILI prediction and may help researchers to better understand DILI and improve drug safety.

## Data Availability

The data used to support the findings of this study are available from the article and supplementary information file.

## Disclosure

Heng Luo present address is BenevolentAI, 1 Dock 72 Way, 7th Floor, Brooklyn, NY 11205 , USA.

## Conflicts of Interest

The authors declare that there is no conflict of interest.

## Authors' Contributions

Xiaobin Liu and Danhua Zheng contributed equally to the study.

## Acknowledgments

This work was supported by Funds of the Joint Plan for Health Education in Fujian (#WKJ2016-2-25), the Project in Fuzhou Science and Technology Bureau (#2018-G-49), and the National Natural Science Foundation of China (81971837).

## Supplementary Materials

Supplementary Figure 1: AUC values of different sets of features and DILI endpoints using logistic regression for drugs in the DrugDex database during 10-fold cross-validations. Supplementary Figure 2: AUC values of different sets of features and DILI endpoints using logistic regression for drugs in the DrugPoints database during 10-fold cross-validations. Supplementary Figure 3: AUC values of different sets of features and DILI endpoints using random forest for drugs in the DrugPoints database during 10-fold cross-validations. Supplementary Figure 4: AUC values of different sets of features and DILI endpoints using logistic regression for drugs in the DailyMed database during 10-fold cross-validations. Supplementary Figure 5: AUC values of different sets of features and DILI endpoints using random forest for drugs in the DailyMed database during 10-fold cross-validations. Supplementary Figure 6: for the other DILI endpoints in DrugDex, the network of proteins according to the feature importance (a), and KEGG pathway analysis of important protein features (b). Supplementary Table 1: DILI endpoints curated from DrugDex. Supplementary Table 2: DILI endpoints curated from DrugPoints. Supplementary Table 3: DILI endpoints curated from DailyMed. Supplementary Table 4: AUC values of different sets of features and DILI endpoints for drugs in the DrugDex database during an independent test. Supplementary Table 5: AUC values of different sets of features and DILI endpoints for drugs in the DrugPoints database during an independent test. Supplementary Table 6: AUC values of different sets of features and DILI endpoints for drugs in the DailyMed database during an independent test. Supplementary Table 7: association between DrugDex DILI endpoints and top-level ATC codes. Supplementary Table 8: feature importance of using top-level ATC codes to predict DrugDex DILI endpoints. *(Supplementary Materials)*

## References

[1] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nature Reviews. Drug Discovery*, vol. 3, no. 8, pp. 673–683, 2004.

[2] H. Luo, W. Mattes, D. L. Mendrick, and H. Hong, "Molecular docking for identification of potential targets for drug repurposing," *Current Topics in Medicinal Chemistry*, vol. 16, no. 30, pp. 3636–3645, 2016.

[3] R. A. Wilke, D. W. Lin, D. M. Roden et al., "Identifying genetic risk factors for serious adverse drug reactions: current progress and challenges," *Nature Reviews. Drug Discovery*, vol. 6, no. 11, pp. 904–916, 2007.

[4] H. Luo, T. Du, P. Zhou et al., "Molecular docking to identify associations between drugs and class I human leukocyte antigens for predicting idiosyncratic drug reactions," *Combinatorial Chemistry & High Throughput Screening*, vol. 18, no. 3, pp. 296–304, 2015.

[5] D. Schuster, C. Laggner, and T. Langer, "Why drugs fail–a study on side effects in new chemical entities," *Current Pharmaceutical Design*, vol. 11, no. 27, pp. 3545–3559, 2005.

[6] R. Andrade, M. Lucena, M. Fernandez et al., "Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period," *Gastroenterology*, vol. 129, no. 2, pp. 512–521, 2005.

[7] A. Regev, "Drug-induced liver injury and drug development: industry perspective," *Seminars in Liver Disease*, vol. 34, no. 2, pp. 227–239, 2014.

[8] A. Cheng and S. L. Dixon, "In silico models for the prediction of dose-dependent human hepatotoxicity," *Journal of Computer-Aided Molecular Design*, vol. 17, no. 12, pp. 811–823, 2003.

[9] R. D. Clark, P. R. N. Wolohan, E. E. Hodgkin, J. H. Kelly, and N. L. Sussman, "Modelling in vitro hepatotoxicity using molecular interaction fields and SIMCA," *Journal of Molecular Graphics & Modelling*, vol. 22, no. 6, pp. 487–497, 2004.

[10] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 44, no. 1.2, pp. 206–226, 2000.

[11] Y. Low, T. Uehara, Y. Minowa et al., "Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches," *Chemical Research in Toxicology*, vol. 24, no. 8, pp. 1251–1262, 2011.

[12] A. D. Rodgers, H. Zhu, D. Fourches, I. Rusyn, and A. Tropsha, "Modeling liver-related adverse effects of drugs using *k*nearest neighbor quantitative structure-activity relationship method," *Chemical Research in Toxicology*, vol. 23, no. 4, pp. 724–732, 2010.

[13] S. Ekins, A. J. Williams, and J. J. Xu, "A predictive ligand-based Bayesian model for human drug-induced liver injury," *Drug Metabolism and Disposition*, vol. 38, no. 12, pp. 2302–2308, 2010.

[14] Z. Liu, Q. Shi, D. Ding, R. Kelly, H. Fang, and W. Tong, "Translating clinical findings into knowledge in drug safety evaluation - drug induced liver injury prediction system (DILIps)," *Plos Computational Biology*, vol. 7, no. 12, article e1002310, 2011.

[15] M. Cruz-Monteagudo, M. N. D. S. Cordeiro, and F. Borges, "Computational chemistry approach for the early detection of drug-induced idiosyncratic liver toxicity," *Journal of Computational Chemistry*, vol. 29, no. 4, pp. 533–549, 2008.

[16] X. W. Zhu, A. Sedykh, and S. S. Liu, "Hybrid in silico models for drug-induced liver injury using chemical descriptors and in vitro cell-imaging information," *Journal of Applied Toxicology*, vol. 34, no. 3, pp. 281–288, 2014.

[17] N. Kaplowitz, "Drug-induced liver injury," *Clinical Infectious Diseases*, vol. 38, Supplement 2, pp. S44–S48, 2004.

[18] M. P. Holt and C. Ju, "Mechanisms of drug-induced liver injury," *The AAPS Journal*, vol. 8, no. 1, pp. E48–E54, 2006.

[19] C. J. Ursem, N. L. Kruhlak, J. F. Contrera, P. M. MacLaughlin, R. D. Benz, and E. J. Matthews, "Identification of structure–activity relationships for adverse effects of pharmaceuticals in humans. Part A: use of FDA postmarket reports to create a database of hepatobiliary and urinary tract toxicities," *Regulatory Toxicology and Pharmacology*, vol. 54, no. 1, pp. 1–22, 2009.

[20] I. Tsakovska, M. al Sharif, P. Alov et al., "Molecular modelling study of the PPAR$\gamma$ receptor in relation to the mode of action/adverse outcome pathway framework for liver steatosis," *International Journal of Molecular Sciences*, vol. 15, no. 5, pp. 7651–7666, 2014.

[21] D. Fourches, J. C. Barnes, N. C. Day, P. Bradley, J. Z. Reed, and A. Tropsha, "Cheminformatics analysis of assertions mined from literature that describe drug-induced liver injury in different species," *Chemical Research in Toxicology*, vol. 23, no. 1, pp. 171–183, 2010.

[22] K. Chan, N. S. Jensen, P. M. Silber, and P. J. O'Brien, "Structure–activity relationships for halobenzene induced cytotoxicity in rat and human hepatoctyes," *Chemico-Biological Interactions*, vol. 165, no. 3, pp. 165–174, 2007.

[23] C. Funk and A. Roth, "Current limitations and future opportunities for prediction of DILI from in vitro," *Archives of Toxicology*, vol. 91, no. 1, pp. 131–142, 2017.

[24] S. Kim, P. A. Thiessen, E. E. Bolton et al., "PubChem substance and compound databases," *Nucleic Acids Research*, vol. 44, no. D1, pp. D1202–D1213, 2016.

[25] R. Guha, "Chemical informatics functionality inR," *Journal of Statistical Software*, vol. 18, no. 5, pp. 1–16, 2007.

[26] D. S. Wishart, Y. D. Feunang, A. C. Guo et al., "DrugBank 5.0: a major update to the DrugBank database for 2018," *Nucleic Acids Research*, vol. 46, no. D1, pp. D1074–D1082, 2018.

[27] Z. Wang, N. R. Clark, and A. Ma'ayan, "Drug-induced adverse events prediction with the LINCS L1000 data," *Bioinformatics*, vol. 32, no. 15, pp. 2338–2345, 2016.

[28] Z. Weng, K. Wang, H. Li, and Q. Shi, "A comprehensive study of the association between drug hepatotoxicity and daily dose, liver metabolism, and lipophilicity using 975 oral medications," *Oncotarget*, vol. 6, no. 19, pp. 17031–17038, 2015.

[29] R. Temple, “Hy's law: predicting serious hepatotoxicity,” *Pharmacoepidemiology and Drug Safety*, vol. 15, no. 4, pp. 241–243, 2006.

[30] M. Chen, J. Borlak, and W. Tong, “A model to predict severity of drug-induced liver injury in humans,” *Hepatology*, vol. 64, no. 3, pp. 931–940, 2016.

[31] M. Hewitt, S. J. Enoch, J. C. Madden, K. R. Przybylak, and M. T. D. Cronin, “Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action,” *Critical Reviews in Toxicology*, vol. 43, no. 7, pp. 537–558, 2013.

[32] L. Yang and P. Agarwal, “Systematic drug repositioning based on clinical side-effects,” *PLoS One*, vol. 6, no. 12, article e28025, 2011.

[33] N. Chalasani, H. L. Bonkovsky, R. Fontana et al., “Features and outcomes of 899 patients with drug-induced liver injury: the DILIN prospective study,” *Gastroenterology*, vol. 148, no. 7, pp. 1340–1352.e7, 2015.

[34] M. Liu, Y. Wu, Y. Chen et al., “Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs,” *Journal of the American Medical Informatics Association*, vol. 19, no. e1, pp. e28–e35, 2012.

[35] C. Strobl, A. L. Boulesteix, A. Zeileis, and T. Hothorn, “Bias in random forest variable importance measures: illustrations, sources and a solution,” *BMC Bioinformatics*, vol. 8, p. 25, 2007.

[36] C. von Mering, L. J. Jensen, B. Snel et al., “STRING: known and predicted protein-protein associations, integrated and transferred across organisms,” *Nucleic Acids Research*, vol. 33, no. - Database issue, pp. D433–D437, 2004.

[37] V. Souza-Mello, “Peroxisome proliferator-activated receptors as targets to treat non-alcoholic fatty liver disease,” *World Journal of Hepatology*, vol. 7, no. 8, pp. 1012–1019, 2015.

[38] M. R. Ebrahimkhani, F. Oakley, L. B. Murphy et al., “Stimulating healthy tissue regeneration by targeting the 5-HT$_{2B}$ receptor in chronic liver disease,” *Nature Medicine*, vol. 17, no. 12, pp. 1668–1673, 2011.

[39] A. Anzai, R. R. Marcondes, T. H. Gonçalves et al., “Impaired branched-chain amino acid metabolism may underlie the non-alcoholic fatty liver disease-like pathology of neonatal testosterone-treated female rats,” *Scientific Reports*, vol. 7, no. 1, article 13167, 2017.

[40] G. Bindea, B. Mlecnik, H. Hackl et al., “ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks,” *Bioinformatics*, vol. 25, no. 8, pp. 1091–1093, 2009.

[41] J. Li, P. Zhao, Y. Li, Y. Tian, and Y. Wang, “Systems pharmacology-based dissection of mechanisms of Chinese medicinal formula Bufei Yishen as an effective treatment for chronic obstructive pulmonary disease,” *Scientific Reports*, vol. 5, no. 1, article 15290, 2015.

[42] V. Lozeva-Thomas, “Serotonin brain circuits with a focus on hepatic encephalopathy,” *Metabolic Brain Disease*, vol. 19, no. 3/4, pp. 413–420, 2004.

[43] K. J. Jensen, G. Alpini, and S. Glaser, “Hepatic nervous system and neurobiology of the liver,” *Comprehensive Physiology*, vol. 3, no. 2, pp. 655–665, 2013.

[44] E. Kotsampasakou and G. F. Ecker, “Predicting drug-induced cholestasis with the help of hepatic transporters-an in silico modeling approach,” *Journal of Chemical Information and Modeling*, vol. 57, no. 3, pp. 608–615, 2017.

[45] X. W. Zhu and S. J. Li, “In silico prediction of drug-induced liver injury based on adverse drug reaction reports,” *Toxicological Sciences*, vol. 158, no. 2, pp. 391–400, 2017.

[46] B. Bajzelj and V. Drgan, “Hepatotoxicity modeling using counter-propagation artificial neural networks: handling an imbalanced classification problem,” *Molecules*, vol. 25, no. 3, p. 481, 2020.

[47] M. Chen, A. Suzuki, S. Thakkar, K. Yu, C. Hu, and W. Tong, “DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans,” *Drug Discovery Today*, vol. 21, no. 4, pp. 648–653, 2016.

[48] H. K. Shin, M. G. Kang, D. Park, T. Park, and S. Yoon, “Development of prediction models for drug-induced cholestasis, cirrhosis, hepatitis, and steatosis based on drug and drug metabolite structures,” *Frontiers in Pharmacology*, vol. 11, p. 67, 2020.

[49] S. He, T. Ye, R. Wang et al., “An in silico model for predicting drug-induced hepatotoxicity,” *International Journal of Molecular Sciences*, vol. 20, no. 8, p. 1897, 2019.