ORIGINAL ARTICLE

In silico prediction of chemical-induced hematotoxicity with machine learning and deep learning methods

Yuqing Hua[1,2] · Yinping Shi[2] · Xueyan Cui[2] · Xiao Li[2,3]

Abstract

Chemical-induced hematotoxicity is an important concern in the drug discovery, since it can often be fatal when it happens. It is quite useful for us to give special attention to chemicals which can cause hematotoxicity. In the present study, we focused on in silico prediction of chemical-induced hematotoxicity with machine learning (ML) and deep learning (DL) methods. We collected a large data set contained 632 hematotoxic chemicals and 1525 approved drugs without hematotoxicity. Computational models were built using several different machine learning and deep learning algorithms integrated on the Online Chemical Modeling Environment (OCHEM). Based on the three best individual models, a consensus model was developed. It yielded the prediction accuracy of 0.83 and balanced accuracy of 0.77 on external validation. The consensus model and the best individual model developed with random forest regression and classification algorithm (RFR) and QNPR descriptors were made available at https://ochem.eu/article/135149, respectively. The relevance of 8 commonly used molecular properties and chemical-induced hematotoxicity was also investigated. Several molecular properties have an obvious differentiating effect on chemical-induced hematotoxicity. Besides, 12 structural alerts responsible for chemical hematotoxicity were identified using frequency analysis of substructures from Klekota–Roth fingerprint. These results should provide meaningful knowledge and useful tools for hematotoxicity evaluation in drug discovery and environmental risk assessment.

Graphic abstract
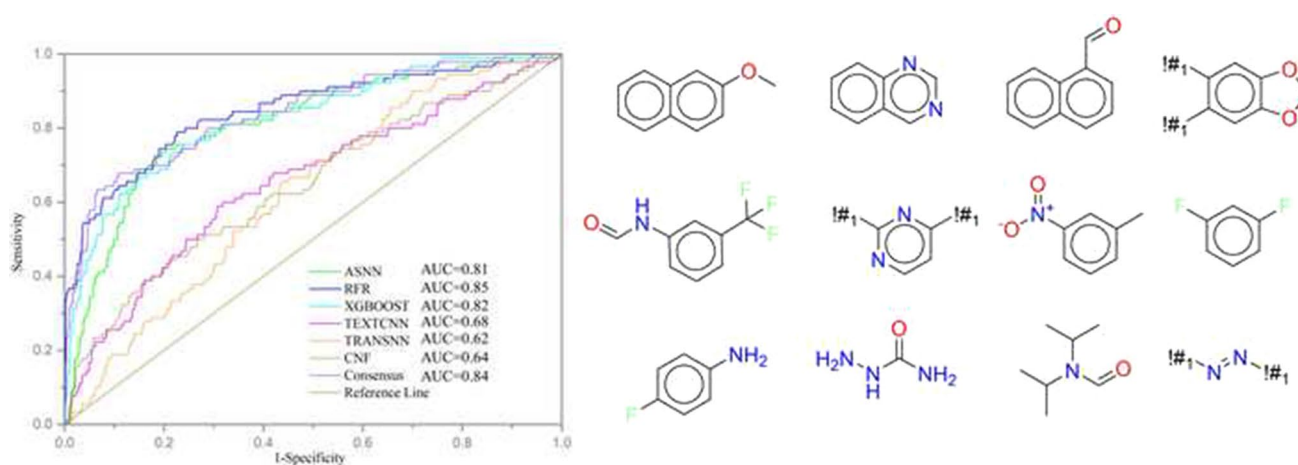
Keywords Chemical-induced hematotoxicity · Machine learning · Deep learning · Consensus model · Structural alert

✉ Xiao Li
  x.li@sdu.edu.cn; lixiao1688@163.com

Extended author information available on the last page of the article

Introduction

Chemical-induced hematotoxicity refers to adverse effects of chemicals on blood-forming organs such as bone marrow or the constituents of blood, including platelets, leukocytes and erythrocytes [1]. Hematotoxicity can result from environmental chemicals, food additives and medications. Chemical-induced hematotoxicity is rarely encountered in the past decades due to the implementation of conservative hygiene standards [2]. However, it can often be fatal when it happens. A number of drug candidates with good clinical activity failed during the Phase I due to the hematotoxicity, which has resulted in serious financial and time losses. Drug-induced hematotoxicity has even been the cause of several drugs being withdrawn from the market, such as remoxipride [3].

It is quite useful for us to give special attention to chemicals which can cause hematotoxicity, especially in the early stages of drug discovery. At present, in addition to the in vivo detection models with various animal species, several in vitro assays have also been used for the estimation of chemical-induced hematotoxicity, including CFU-GM assay, BFU-E assay and CFU-MK assay [4, 5]. The biological experiments using large numbers of animals are opposed and protested by animal rights groups, and these in vitro and in vivo test approaches are also time-consuming and costly. Therefore, the computational approaches should be applied to provide an alternative tool for estimation of chemical-induced hematotoxicity. In the past decades, machine learning (ML) and deep learning (DL) approaches have been widely used for the development of computational models for the toxicity assessment of chemicals, particularly for the drugs [6–9]. However, only few of the machine learning models for hematotoxicity have been reported up to now. Crivori, et al. [10] and Zhang, et al. [11] developed in silico regression and classification models for drug-induced myelotoxicity, respectively. These models have shown good predictive ability, while the chemicals with the other endpoints related to hematotoxicity were not included, the structural characteristics of toxic and non-toxic compounds were not analyzed in these studies, and the usefulness of the models was restricted because of poor availability.

In this study, we focused on the (1) development of in silico models for chemical-induced hematotoxicity with machine learning and deep learning methods based on structurally diverse organic chemicals; (2) the comparison of several common molecular properties between the hematotoxic compounds and non-hematotoxic drugs; and (3) analysis of the difference of structural characteristics between the chemicals with and without hematotoxicity.

Materials and methods

Data source and preparation

The hematotoxic chemicals were collected from three different sources: (1) drugs with hematotoxicity related side effects extracted from the SIDER database [12]; (2) chemicals be able to induce blood disorders extracted from the ChemIDplus database [13]; and (3) drugs which can cause blood disorders reported in the literature. The chemical structures were carefully prepared in the following steps: (1) removing mixtures, inorganic and organometallic compounds; (2) salts were converted to their corresponding acidic or basic forms, and water molecules were removed from the hydrates; (3) removing compounds with molecular weights less than 30 or more than 1000; and (4) only one stereoisomer was retained. The final data set contained 632 chemicals with hematotoxicity. The hematotoxic structures were randomly divided into a training set and a validation set with the ratio of 6:1. Because of the lack of negative data, we used the approved drugs in DrugBank [14] as negative data in this study. The withdrawn, nutraceutical and illicit drugs were dropped, and duplicates with hematotoxic chemicals were removed. Then the approved drugs were prepared using the same methods as hematotoxic structures. A total of 1525 entities remained, which were used as non-hematotoxic structures. The non-hematotoxic chemicals are much more than the hematotoxic chemicals. To avoid the bias caused by the imbalanced class data sets, the non-hematotoxic chemicals were randomly partitioned into training and validation set in 3:2.

Principal component analysis for the definition of the chemical space of data sets

For the global QSAR models, it is important to have a data set with sufficiently structural diversity to ensure a reasonable predictive accuracy [15]. Several different approaches for the assessment of chemical diversity have been used in the previous literature. In the present study, we analyzed the chemical space of data sets with the first two principle components [16] of CDK (Chemistry Development Kit) Descriptors. Principal component analysis (PCA) is a technique for reducing the dimensionality and increasing interpretability, which can also minimize information loss at the same time. It reduced to solving an eigenvector problem by creating new uncorrelated variables that successively maximize variance. Herein, the PCA was performed using SPSS Statistics.

Machine learning and deep learning methods

The Online Chemical Modeling Environment (OCHEM) is a web-based platform that aims to automate and simplify the typical steps required for quantitative structure activity relationship (QSAR) modeling [17]. OCHEM integrates a number of the state-of-the-art machine learning (ML) and deep learning (DL) methods used for the development of QSAR models [18–22]. Machine learning is an application of artificial intelligence (AI) which focuses on the development of computer programs that can access data and use it to learn for themselves. Among the many ML methods, we applied five algorithms which have been widely used and performed well in the previous work. They are associative neural network (ASNN) [23], k-nearest neighbors (kNN) [24], support vector machine (SVM) [25], extreme gradient boosting (XGBoost) [26] and random forest regression and classification (RFR) [27]. Deep learning is an extension and expansion of the artificial neural network (ANN) research, which has excellent ability of feature learning. DL algorithms can characterize data and learn statistical laws from massive training samples and thus can make a more accurate prediction on new unknown data. In this study, we used three DL approaches, including TEXTCNN algorithm available from DeepChem [28], convolutional neural network fingerprint (CNF) [29] and transformer convolutional neural network (TRANSNN) [12], for the DL model building.

The detailed descriptions of the ML and DL methods can be found in the corresponding literature, and the parameters of these algorithms were optimized with the default setting at OCHEM.

Molecular description and descriptor selection

Many different descriptors contributed by academic groups and commercial enterprises can be used in OCHEM. For the ML modeling, 7 descriptor packages were calculated individually for molecular description, including Chemistry Development Kit (CDK, 256 descriptors), Dragon v.6 (4885 descriptors), InductiveDescriptors (54 descriptors), PyDescriptor (1624 descriptors), QNPR descriptors (length 1–3), RDKIT (blocks: 1–11 15–16) and MAP4 fingerprint. All of the descriptors that were filtered decrease the interference of multicollinearity. The pairwise decorrelation method for descriptor selection was applied, and the duplicated descriptors with pairwise correlation of more than 0.95 were eliminated.

Without descriptors, the whole molecule as a SMILES-string served as the input to the DL algorithms.

Consensus modeling

The term of consensus modeling has been used in many scientific disciplines to define methods by which a group of individuals can come to an agreement [30]. In QSAR community, this term was used for methodologies that aggregate the predictions of several individual QSAR models to arrive at a single prediction. It is believed that a single QSAR model always cannot guarantee the best quality of predictions of chemical toxicity, while the consensus models can benefit from various representations of chemical structures, which characterized molecules from different perspectives. Many consensus models reported in the literature use a naïve approach, which calculates the average value among all the individual model predictions [31–33]. In this study, the best performed individual models would be selected, and based on them, the consensus model would be developed with simple average of predictions from the individual models.

Model validation and evaluation

All models were validated by fivefold cross-validation and a diverse external validation set. The performances of models were assessed by several statistical parameters. The sensitivity (SE) and specificity (SP) are two measures applicable to binary classifications, which denote the ratios of positive instances and negative instances correctly identified, respectively. The accuracy (ACC) is the total correct predictive accuracy of samples, and the balanced accuracy (BAC) is defined as the arithmetic mean of SE and SP, which can account for this problem and be the correct classification performance metric for imbalanced sets [34]. The Matthews correlation coefficient (MCC) is generally regarded as a balanced measure which can be used even if the classes are of very different sizes [35]. These parameters can be calculated with Eqs. 1, 2, 3, 4, 5.

$$SE = \frac{TP}{TP + FN} \tag{1}$$

$$SP = \frac{TN}{TN + FP} \tag{2}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$BAC = \frac{SE + SP}{2} \tag{4}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

In these equations, TP, TN, FP and FN are the numbers of true positives, true negatives, false positives and false negatives, respectively.

Additionally, we also plotted the receiver operating characteristic (ROC) curve and computed the values of area under the ROC curve (AUC).

Analysis of the differences of molecular properties between hematotoxic and non-hematotoxic chemicals

The molecular properties of chemicals can make a significant difference in their biological activities. In the present study, we calculated eight commonly molecular properties of hematotoxic and non-hematotoxic compounds, and compared the differences between the two classes. The purpose of this analysis is to investigate the relevance of these molecular properties with chemical-induced hematotoxicity. These eight properties have been widely adopted in the analysis for other endpoints. They are molecular weight (MW), molecular polar surface area (MPSA), AlogP, molecular solubility (LogS), the number of hydrogen bond acceptors (nHAcc) and donors (nHDon), the number of rotatable bonds (nRotB) and the number of aromatic rings (nAR). These properties are relative to the molecular size, lipophilicity and solubility, hydrogen bonding ability and complexity, respectively.

The molecular properties were calculated with PaDEL-Descriptor package [36].

Identification of structural alerts responsible for chemical-induced hematotoxicity

Structural alert (SA) has been widely accepted in toxicity prediction studies, which can be defined as the structural fragments result in toxicity of the chemical compounds [37–39]. Due to the direct derivation from mechanistic knowledge, possible toxicological effect can be alerted with the presence of SA in a chemical compound.

In this study, the structural alerts responsible for hematotoxicity were identified with the appearance frequencies among hematotoxic and non-hematotoxic chemicals of each substructure from Klekota–Roth fingerprint (KRFP) [40]. If a structural fragment presented far more frequently in hematotoxic chemicals than non-hematotoxic chemicals, the presence of such a fragment would alert to hematotoxicity. This fragment would be regarded as a structural alert responsible for chemical-induced hematotoxicity.

Results and discussion

Data collection and diversity analysis

After careful filtering and preparation, 632 hematotoxic structures were collected. Meanwhile, 1525 non-hematotoxic structures were extracted from approved drugs. There is a severe imbalance in the number of hematotoxic and non-hematotoxic compounds. As we know, the imbalance of training samples should make it difficult to develop well-performed binary classification models [41, 42]. When we determine the training and validation sets, we randomly divided the hematotoxic structures into training and validation set with the ratio of 6:1, and the non-hematotoxic structures were randomly divided into training and validation set with the ration of 3:2. Thus, as shown in Table 1, the training set contained 541 hematotoxic chemicals and 915 non-hematotoxic drugs, and the validation set contained 91 hematotoxic chemicals and 610 non-hematotoxic drugs. The structures for modeling are given in Table S1 of Supporting Information.

As is known to all, the structural diversity of the compounds is very important for machine learning models. Herein, we calculated the CDK (Chemistry Development Kit) Descriptors [43] of chemicals; then, the principle component analysis (PCA) was performed to find principle components (PCs) which more technically address the different descriptors for analyzing the chemical space of those compounds. PCA can simplify the complexity in high-dimensional data while retaining trends and patterns. It does this by transforming the data into fewer dimensions, which act as summaries of features [44]. The first two PCs were kept and used for the definition of the chemical space of compound in the training and validation sets in this study. As shown in Fig. 1, the distribution scatter diagram illustrated that the data sets shared a similar chemical space.

Results of model building

Combined with five machine learning algorithms and 7 molecule descriptor packages implemented at the OCHEM platform, a total of 35 ML models for hematotoxicity prediction were generated. Besides, three deep learning models were

Table 1 Detailed statistical number of chemicals used in the data set

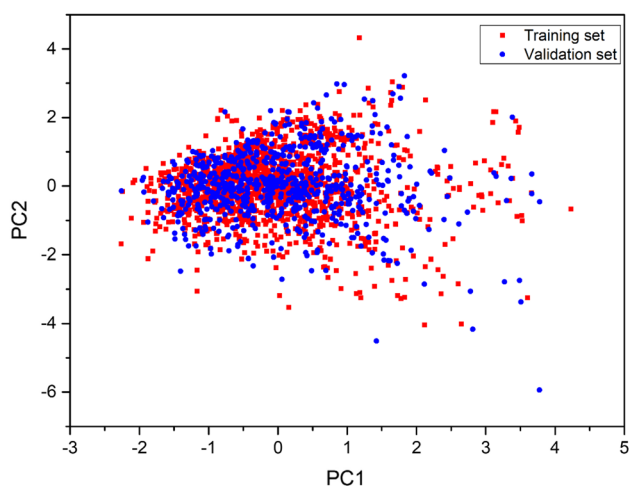|	Hematotoxic chemicals	Non-hematotoxic chemicals	Total
Training set	541	915	1456
Validation set	91	610	701
Total	632	1525	2157

Fig. 1 Chemical space defined by the first two principal components of CDK descriptors. Red squares stand for the training set, and blue circles stand for the validation set

established with different DL methods based on chemical SMILES-string. The performances of these models on five-fold cross-validation are given in Table 2.

Although the performance of most models failed to meet our expectations, the machine learning models developed with QNPR descriptors showed good predictive performance. These descriptors are used for QNPR (Quantitative Name Property Relationship), thus giving their name. The descriptors are derived directly from the compounds name or SMILES strings. For each molecule, either canonical SMILES or IUPAC name is split into fragments of a specified length, which is determined by the configuration [45]. The ASNN_QNPR model showed the best performance with ACC 0.79 and BAC 0.76, and the values of SE, SP, MCC and AUC were 0.67, 0.86, 0.54 and 0.81, respectively. The XGBOOST_QNPR and RFR_QNPR models also provided good results with high ACC (0.76 and 0.78), BAC (0.73 and 0.76), MCC (0.47 and 0.53) and AUC (0.81 and 0.84). The deep learning methods did not perform well for the model building. The best DL model was built with TEXTCNN. It showed good SP with 0.82 and AUC with 0.73, but the SE was only 0.44.

It has been proved that the consensus approach could improve the predictive ability for models. Since three ML models (ASNN_QNPR, XGBOOST_QNPR and RFR_QNPR) provided much better predictive performance than the others in this study, a consensus model was developed as a simple average of the individual models based on these three best ML models. Because the DL models did not show comparable predictive ability, none of the DL models were used for the development of the consensus model. As shown in Table 2, the consensus model performed better than individual models. It achieved the ACC value of 0.79 and the

BAC of 0.76, and the values of SP, SE, MCC and AUC were 0.66, 0.87, 0.54 and 0.85, respectively.

Model validation

The external validation set was completely independent from the model training, so it could test the robustness and predictive ability of models objectively. The performances of the models on external validation are given in Table 3. The representation of receiver operating characteristics (ROC) curve for models on external validation set is shown in Fig. 2. The three ML models with QNPR descriptors provided good performances on external validation, too. The RFR_QNPR model performed as the best individual model, with BAC at 0.76 and AUC at 0.85. For DL models, only the TEXTCNN model gave a barely satisfactory performance on validation. The ACC value was 0.68, and the BAC was 0.66. The consensus model based on the best ML models performed best on most of the statistical parameters, with the highest ACC at 0.83, BAC at 0.77 and MCC at 0.44. It should be noted that the consensus model did not achieved higher AUC value (0.84) than individual model RFR_QNPR. Combined with the results of internal and external validation, we considered the consensus model as the best model, while the RFR_QNPR model was the best individual model.

Availability of data and models

The performance of the consensus model suggested it could be a useful tool for estimation of the hematotoxicity for chemical compounds. For ease of use, this model was made available at https://ochem.eu/article/135149. The best individual model RFR_QNPR was also freely available at https://ochem.eu/article/135149. The user can open the models and access the statistical parameters, and the individual models for the consensus model building were also available with the corresponding model IDs. These models can be used for chemical hematotoxicity prediction by "Apply the model to new compounds" link. Furthermore, the training and validation sets could be download with the "Export this basket" link.

Performance of the ML and DL algorithms used in model building

From the results of model building and validation, there is no significant difference between the various ML approaches for most descriptor types. Interestingly, ML models built using the QNPR descriptor perform far better than other models. Except for the QNPR descriptor, models with other descriptors performed quite poorly.

The DL models did not provide better predictive ability than the ML models built using the QNPR descriptor, but

Table 2 Performance of models on fivefold cross-validation

Model	SE	SP	ACC	BAC	MCC	AUC
ASNN_CDK	0.33	0.83	0.64	0.58	0.18	0.64
KNN_CDK	0.22	0.87	0.63	0.55	0.12	0.63
LibSVM_CDK	0.35	0.80	0.64	0.58	0.18	0.58
XGBOOST_CDK	0.40	0.77	0.63	0.58	0.17	0.63
RFR_CDK	0.32	0.88	0.67	0.60	0.24	0.67
ASNN_Dragon	0.29	0.71	0.50	0.50	0.00	0.65
KNN_Dragon	0.14	0.93	0.64	0.54	0.12	0.61
LibSVM_Dragon	0.23	0.89	0.65	0.56	0.16	0.56
XGBOOST_Dragon	0.40	0.78	0.64	0.59	0.19	0.65
RFR_Dragon	0.29	0.88	0.66	0.59	0.21	0.68
ASNN_IDescriptors	0.09	0.96	0.64	0.53	0.10	0.59
KNN_IDescriptors	0.20	0.88	0.63	0.54	0.11	0.61
LibSVM_IDescriptors	0.27	0.82	0.62	0.55	0.11	0.55
XGBOOST_IDescriptors	0.36	0.76	0.61	0.56	0.13	0.61
RFR_IDescriptors	0.26	0.84	0.62	0.55	0.12	0.63
ASNN_MAP4	0.42	0.81	0.67	0.62	0.25	0.66
KNN_MAP4	0.48	0.82	0.69	0.65	0.32	0.72
LibSVM_MAP4	0.49	0.82	0.70	0.65	0.32	0.65
XGBOOST_MAP4	0.44	0.83	0.69	0.64	0.30	0.72
RFR_MAP4	0.30	0.96	0.71	0.63	0.36	0.76
ASNN_PyDescriptor	0.18	0.89	0.62	0.53	0.09	0.62
KNN_PyDescriptor	0.12	0.92	0.63	0.52	0.07	0.6
LibSVM_PyDescriptor	0.36	0.80	0.64	0.58	0.17	0.58
XGBOOST_PyDescriptor	0.40	0.77	0.64	0.59	0.19	0.64
RFR_PyDescriptor	0.34	0.86	0.66	0.60	0.23	0.67
ASNN_QNPR	0.67	0.86	0.79	0.76	0.54	0.81
KNN_QNPR	0.25	0.94	0.68	0.59	0.27	0.7
LibSVM_QNPR	0.60	0.86	0.76	0.73	0.48	0.73
XGBOOST_QNPR	0.64	0.83	0.76	0.73	0.47	0.81
RFR_QNPR	0.65	0.86	0.78	0.76	0.53	0.84
ASNN_RDKIT	0.11	0.93	0.63	0.52	0.08	0.62
KNN_RDKIT	0.03	0.97	0.62	0.50	0.00	0.59
LibSVM_RDKIT	0.30	0.82	0.63	0.56	0.14	0.56
XGBOOST_RDKIT	0.39	0.76	0.62	0.57	0.15	0.63
RFR_RDKIT	0.38	0.79	0.64	0.59	0.19	0.64
TEXTCNN	0.44	0.82	0.68	0.63	0.28	0.73
CNF	0.24	0.86	0.63	0.55	0.13	0.67
TRANSNN	0.28	0.84	0.63	0.56	0.14	0.63
Consensus	0.66	0.87	0.79	0.76	0.54	0.85

Table 3 Performance of models on external validation

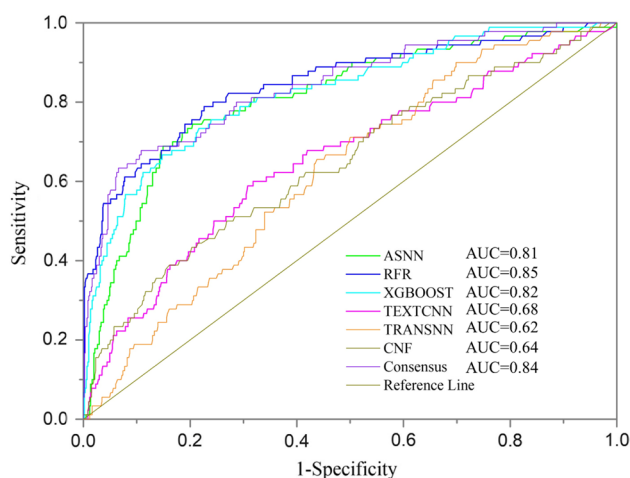| Model	SE	SP	ACC	BAC	MCC	AUC
ASNN_QNPR	0.69	0.83	0.81	0.76	0.42	0.81
RFR_QNPR	0.66	0.85	0.83	0.76	0.42	0.85
XGBOOST_QNPR	0.67	0.84	0.81	0.75	0.40	0.82
TEXTCNN	0.64	0.69	0.68	0.66	0.23	0.68
CNF	0.41	0.81	0.76	0.61	0.18	0.64
TRANSNN	0.22	0.87	0.78	0.54	0.09	0.62
Consensus	0.68	0.86	0.83	0.77	0.44	0.84

Fig. 2 ROC curve of models on external validation. Each color line represents a model

they performed better than the ML models with the other descriptors. As well known, deep learning algorithms could derive high-level features automatically by utilizing large training data. Thus, the main advantage of DL algorithms is reflected in big data problems. However, they should not perform well when data are small, since deep learning algorithms need a lot of data to fully understand it. Traditional machine learning algorithms always perform better than DL algorithms when the size of data is small. We presume that with more training data, most of the DL algorithms would outperform the traditional machine learning approaches.

The differences of molecular properties between hematotoxic and non-hematotoxic chemicals

It is well known that molecular properties could provide useful information for chemical properties. In this study, we took a systemic analysis on the relationships between drug-induced hematotoxicity and 8 commonly used molecular properties. The distributions of these descriptors for hematotoxic and non-hematotoxic structures are shown in Fig. 3.

MW and MPSA are two important characteristics simply assess the size and complexity of compounds. The values of MW were distributed between 30.03 and 994.19, with a mean of 363.66 for the entire data set. The mean value was 372.52 for hematotoxic chemicals and 359.99 for non-hematotoxic drugs. The difference between the mean MW of hematotoxic and non-hematotoxic drugs was not significantly different, with a p-value of 0.11. The values of MPSA were distributed from 0 to 427.26, with a mean of 93.07 for all the entire data set. The mean value was 103.07 for hematotoxic chemicals and 88.92 for non-hematotoxic drugs, with $p < 0.001$. These results suggested that there was

no significant difference in structure size between hematotoxic and non-hematotoxic compounds, while hematotoxic chemicals had significantly larger polar surface area.

The lipophilicity of a chemical is commonly represented by AlogP. For chemicals in the entire data set, the values of AlogP ranged from -11.10 to 20.10, with a mean of 2.01. The mean value of AlogP was 1.73 for hematotoxic chemicals and 2.12 for non-hematotoxic drugs. As shown in Fig. 3, the distributions were significantly different between hematotoxic and non-hematotoxic drugs with p value 0.01. The result indicated that chemical-induced hematotoxicity may be associated with chemical lipophilicity.

Chemical hydrogen bonding ability also is an important character for its activity and toxicity, and it was usually in terms of nHBA and nHBD. In this data set, mean values of nHBA for hematotoxic and non-hematotoxic drugs were 5.74 and 5.01, respectively, with a $p < 0.001$. Meanwhile, the hematotoxic and non-hematotoxic drugs had the mean values of nHBD with 2.15 and 2.00, respectively, and the difference is not significant with a p value of 0.14. The data indicated that nHBA was obviously associated with chemical-induced hematotoxicity while nHBD was not.

As shown in Fig. 3, chemical-induced hematotoxicity was also obviously associated with nRotB (the mean values were 5.56 for hematotoxic chemicals and 5.80 for non-hematotoxic drugs, with p-value 0.23), and nAR (the mean values were 1.45 for hematotoxic chemicals and 1.28 for non-hematotoxic drugs, with p value 0.002).

LogS is an estimation of molecular solubility in water. Herein, the values of LogS were distributed from -20.78 to 4.55, and the mean value was -4.23 for the entire data set. For hematotoxic structures, the mean value of LogS was -4.07 and it is -4.29 for non-hematotoxic drugs. The difference proved no statistical significance with a p value 0.09. This result demonstrated that there is not remarkable difference of molecular solubility between hematotoxic and non-hematotoxic drugs.

Through the analysis of molecular properties, we found several molecular properties have obvious differentiating effect on drug-induced hematotoxicity. In fact, chemical-induced hematotoxicity is a complex endpoint. It is not easy to explain the mechanism of chemical-induced hematotoxicity with individual simple chemical descriptors.

Structural alerts responsible for chemical-induced hematotoxicity

We identified the structural alerts by investigation the appearance frequency of each substructure from KRFP fingerprint. Only the substructures presented in more than 6 compounds were kept. It was found that there was no universal substructure shown in hematotoxic chemicals or non-hematotoxic chemicals. However, we identified several
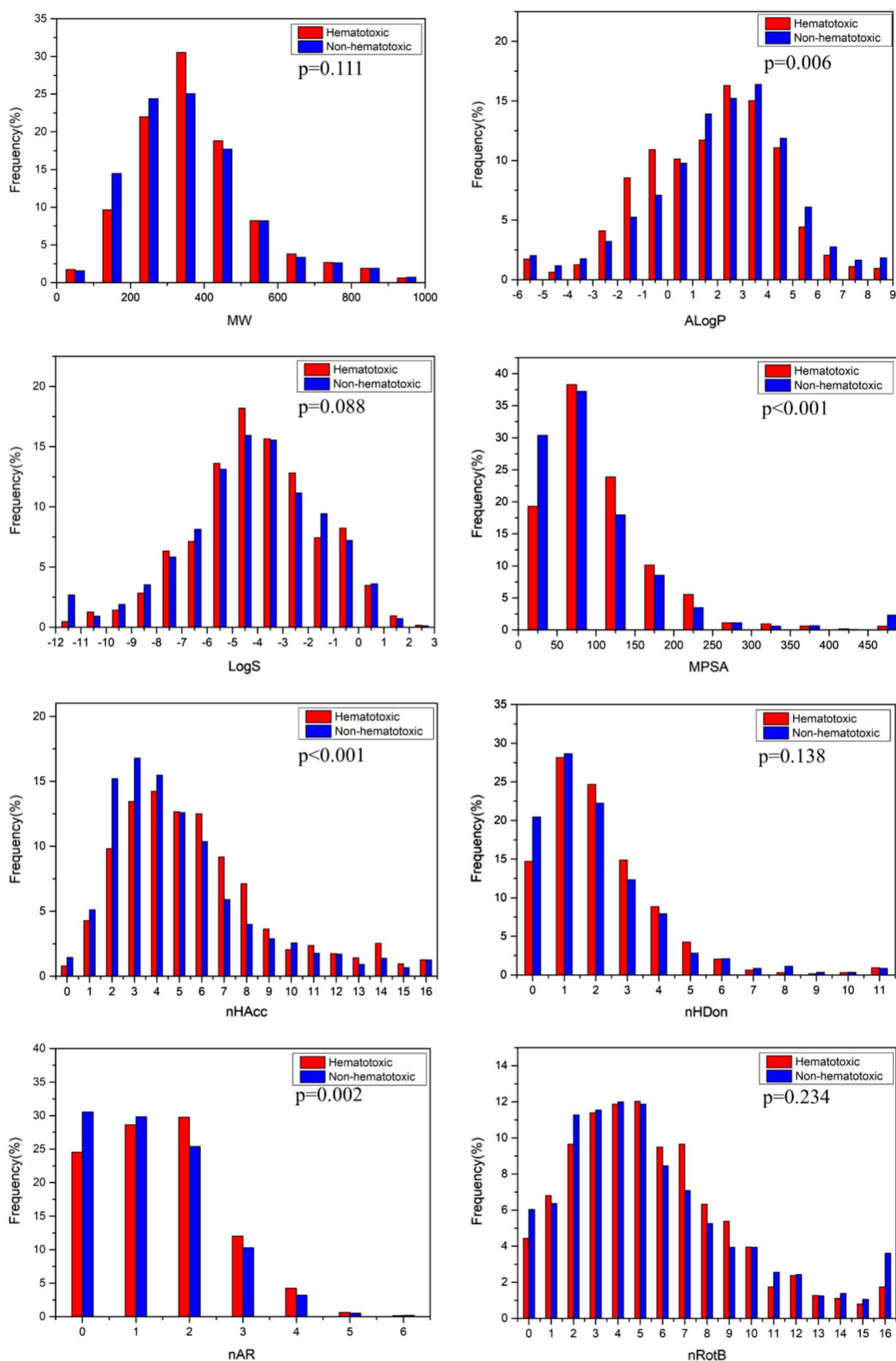
Fig. 3 Distributions of the molecular properties for hematotoxic and non-hematotoxic chemicals

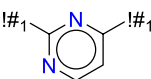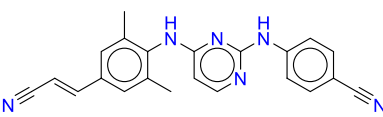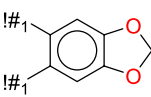**Table 4** Structural alerts responsible for chemical-induced hematotoxicity and the representative structures
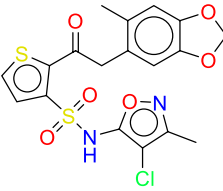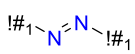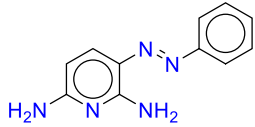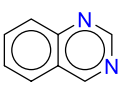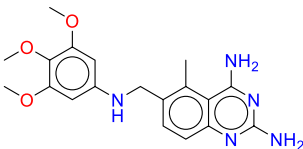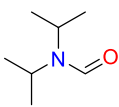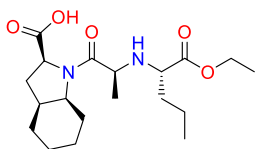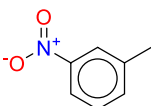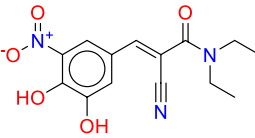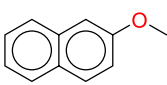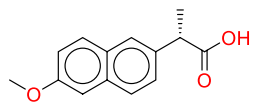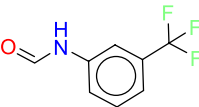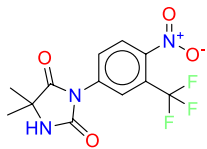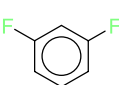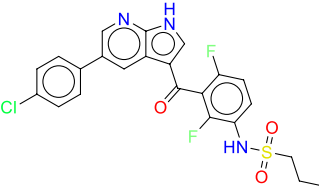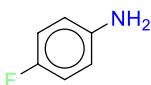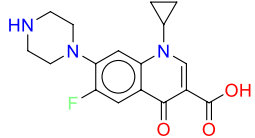
No.	structure	Num _P	Freq _P	Num _N	Freq _N	representative structure
1		6	0.95%	4	0.26%	
2		6	0.95%	4	0.26%	
3		8	1.27%	0	0.00%	
4		8	1.27%	4	0.26%	
5		7	1.11%	3	0.20%	
6		6	0.95%	1	0.07%	
7		9	1.42%	6	0.39%	
8		7	1.11%	1	0.07%	
9		12	1.90%	8	0.52%	
10		15	2.37%	9	0.59%	

Table 4 (continued)

11		13	2.06%	8	0.52%	
12		6	0.95%	3	0.20%	

substructures that appeared far more often in hematotoxic structures than in non-hematotoxic compounds. On the basis of the analysis, 12 structural alerts and the representative structures were filtered and are listed in Table 4. These structural fragments exist much more commonly in hematotoxic chemicals than in non-hematotoxic compounds. To some extent, they may be considered as the structural alerts which can be used to alert the potential of chemical-induced hematotoxicity. If a compound contains one or more such substructures, it is more likely to be hematotoxic than non-hematotoxic. The selected structural alerts were concerned with several different hematotoxic mechanisms. For instance, quinazoline (No. 4) is an important nitrogen-containing heterocyclic compound, which is widely used in many antitumor drugs. Quinazoline and derivatives can interfere with DNA synthesis through a variety of ways, such as DNA methyltransferase [46], leading to blocked blood cell proliferation in the blood system. Aminobenzene and nitrobenzene derivatives (No. 6, No. 8, No. 10) can destroy the red blood cell markedly by increasing methemoglobin formation when incubated with native hemoglobin [47]. As alkylating agents, 1,2-dimethyldiazene (No. 3) and hydrazinecarboxamide (No. 11) can cause myelosuppression by binding DNA base pairs and cross-linking the DNA double strands [48].

The structural alerts responsible for chemical-induced hematotoxicity were identified with frequency analysis, which has been widely used for the discovery of many other endpoints. Admittedly, this method is not perfect, since it is not capable of characterizing the spatial arrangement of identified fragments, and it cannot make a good distinction when two or more structural alerts presented in the same compound. Despite the deficiencies, these identified structural alerts are still very useful. They should be a useful tool for the quick estimation of chemical hematotoxicity with visual alert function in early stage of drug discovery.

Conclusions

In this study, we proposed a high-quality data set containing 632 hematotoxic chemicals and 1525 approved drugs without hematotoxicity. To our knowledge, this should be the largest data set for chemical-induced hematotoxicity until now. On the basis of the data set, computational models were developed with different machine learning and deep learning approaches. The analysis of structural characteristics of hematotoxic chemicals was also performed. Combined with the different machine learning algorithms and descriptors, a total of 35 ML models were developed. Besides, 3 deep learning methods were also applied for model building. Several ML models performed better both on fivefold cross-validation and external validation then the others. Among them, the RFR_QNPR model performed as the best individual model, with AUC at 0.85. Based on these best individual models, we developed a consensus model, which provided the best prediction accuracy of 0.83 and balanced accuracy of 0.77 on external validation. We made the consensus model and the best individual model RFR_QNPR available at https://ochem.eu/article/135149, which can be used freely. Moreover, the relevance of 8 commonly used molecular properties and chemical-induced hematotoxicity was investigated. The results indicated that several key molecular properties, including ALogP, number of hydrogen bond acceptors (nHAcc), molecular polar surface area (MPSA) and the number of rotatable bonds (nRotB), have an obvious differentiating effect on drug-induced hematotoxicity. Thus, these molecular descriptors could play an important part in the identification of hematotoxic chemicals. Finally, the structural alerts responsible for chemical-induced hematotoxicity were identified with frequency analysis. Twelve substructures from KRFP fingerprint were identified as structural alerts for hematotoxicity. A chemical compound would be regarded as hematotoxic if it contains one or more such substructures. We hope the ML models and the structural alerts could provide useful tools for in silico estimation of chemical hematotoxicity in drug discovery.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Rich IN (2003) In vitro hematotoxicity testing in drug development: a review of past, present and future applications. Curr Opin Drug Discov Devel 6(1):100–109

2. Budinsky RA Jr (2000) Hematotoxicity: chemically induced toxicity of the blood: principles of toxicology. Wiley, New York, pp 87–109

3. Cox A (2007) Recognition and management of drug-induced blood disorders. Prescriber 18(3):51–56. https://doi.org/10.1002/psb.22

4. Goto K, Goto M, Ando-Imaoka M et al (2017) Evaluation of drug-induced hematotoxicity using novel in vitro monkey CFU-GM and BFU-E colony assays. J Toxicol Sci 42(4):397–405. https://doi.org/10.2131/jts.42.397

5. Ng P, Belgur C, Barthakur S et al (2019) Organs-on-chips: a new paradigm for safety assessment of drug-induced thrombosis. Cur Opinion Toxicol 17:1–8. https://doi.org/10.1016/j.cotox.2019.08.004

6. Jiao Z, Hu P, Xu H et al (2020) Machine learning and deep learning in chemical health and safety: a systematic review of techniques and applications. ACS Chem Health Safety 27(6):316–334. https://doi.org/10.1021/acs.chas.0c00075

7. Vo AH, Van Vleet TR, Gupta RR et al (2020) An overview of machine learning and big data for drug toxicity evaluation. Chem Res Toxicol 33(1):20–37. https://doi.org/10.1021/acs.chemrestox.9b00227

8. Wang MWH, Goodman JM, Allen TEH (2021) Machine learning in predictive toxicology: recent applications and future directions for classification models. Chem Res Toxicol 34(2):217–239. https://doi.org/10.1021/acs.chemrestox.0c00316

9. Yang H, Lou C, Sun L et al (2019) admetSAR 2.0 web-service for prediction and optimization of chemical ADMET properties. Bioinformatics 35(6):1067–1069

10. Crivori P, Pennella G, Magistrelli M et al (2011) Predicting myelosuppression of drugs from in silico models. J Chem Inf Model 51(2):434–445. https://doi.org/10.1021/ci1003834

11. Zhang H, Yu P, Zhang T-G et al (2015) In silico prediction of drug-induced myelotoxicity by using Naïve Bayes method. Mol Diversity 19(4):945–953. https://doi.org/10.1007/s11030-015-9613-3

12. Kuhn M, Letunic I, Jensen LJ et al (2016) The SIDER database of drugs and side effects. Nucleic Acids Res 44(D1):D1075–D1079. https://doi.org/10.1093/nar/gkv1075

13. Tomasulo P (2002) ChemIDplus-super source for chemical and drug information. Med Ref Serv Q 21(1):53–59. https://doi.org/10.1300/J115v21n01_04

14. Wishart DS, Knox C, Guo AC et al (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucl Acids Res. https://doi.org/10.1093/nar/gkm958

15. Ancuceanu R, Dinu M, Neaga I et al (2019) Development of QSAR machine learning-based models to forecast the effect of substances on malignant melanoma cells. Oncol Lett 17(5):4188–4196. https://doi.org/10.3892/ol.2019.10068

16. Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci 374(2065):20150202. https://doi.org/10.1098/rsta.2015.0202

17. Sushko I, Novotarskyi S, Körner R et al (2011) Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J Comput Aided Mol Des 25(6):533–554. https://doi.org/10.1007/s10822-011-9440-2

18. Cui X, Liu J, Zhang J et al (2019) In silico prediction of drug-induced rhabdomyolysis with machine-learning models and structural alerts. J Appl Toxicol 39(8):1224–1232. https://doi.org/10.1002/jat.3808

19. Cui X, Yang R, Li S et al (2020) Modeling and insights into molecular basis of low molecular weight respiratory sensitizers. Mol Diversity. https://doi.org/10.1007/s11030-020-10069-3

20. Karpov P, Godin G, Tetko IV (2020) Transformer-CNN: Swiss knife for QSAR modeling and interpretation. J Cheminform 12(1):17. https://doi.org/10.1186/s13321-020-00423-w

21. Kovalishyn V, Abramenko N, Kopernyk I et al (2018) Modelling the toxicity of a large set of metal and metal oxide nanoparticles using the OCHEM platform. Food Chem Toxicol 112:507–517. https://doi.org/10.1016/j.fct.2017.08.008

22. Li X, Zhang Y, Li H et al (2017) Modeling of the hERG K+ channel blockage using online chemical database and modeling environment (OCHEM). Mol Inf 36(12):1700074. https://doi.org/10.1002/minf.201700074

23. Tetko IV (2008) Associative neural network. In: Clifton NJ (ed) Methods in molecular biology. Springer, Berlin

24. P Indyk, R Motwani, (1998) Approximate nearest neighbors: towards removing the curse of dimensionality. Paper presented at the Proceedings of the thirtieth annual ACM symposium on Theory of computing, Dallas, Texas, USA https://doi.org/10.1145/276698.276876

25. Chang C-C, Lin C-J (2011) LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol. https://doi.org/10.1145/1961189.1961199

26. Chen T, Guestrin C (2016) *XGBoost: A Scalable Tree Boosting System*. Paper presented at the Proceedings of the 22nd ACM SIGKDD International conference on knowledge discovery and data mining, San Francisco, California, USA. https://doi.org/10.1145/2939672.2939785

27. Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

28. Wu Z, Ramsundar B, Feinberg Evan N et al (2018) Molecule Net: a benchmark for molecular machine learning. Chem Sci 9(2):513–530. https://doi.org/10.1039/C7SC02664A

29. Nogueira RF, Lotufo RdA, Machado RC (2016) Fingerprint liveness detection using convolutional neural networks. IEEE Trans Inf Forensics Secur 11(6):1206–1213. https://doi.org/10.1109/TIFS.2016.2520880

30. Hewitt M, Cronin MTD, Madden JC et al (2007) Consensus QSAR models: do the benefits outweigh the complexity? J Chem Inf Model 47(4):1460–1468. https://doi.org/10.1021/ci700016d

31. Lei T, Li Y, Song Y et al (2016) ADMET evaluation in drug discovery: 15 Accurate prediction of rat oral acute toxicity using relevance vector machine and consensus modeling. J Cheminformatics 8(1):6

32. Khan K, Benfenati E, Roy K (2019) Consensus QSAR modeling of toxicity of pharmaceuticals to different aquatic organisms: ranking and prioritization of the DrugBank database compounds. Ecotoxicol Environ Saf 168:287–297. https://doi.org/10.1016/j.ecoenv.2018.10.060

33. Valsecchi C, Grisoni F, Consonni V et al (2020) Consensus versus individual QSARs in classification: comparison on a large-scale case study. J Chem Inf Model 60(3):1215–1223. https://doi.org/10.1021/acs.jcim.9b01057

34. Abdelaziz A, Spahn-Langguth H, Schramm K-W et al (2016) Consensus modeling for HTS assays using In silico descriptors calculates the best balanced accuracy in Tox21 challenge. Front Environ Sci. https://doi.org/10.3389/fenvs.2016.00002

35. Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21(1):6. https://doi.org/10.1186/s12864-019-6413-7

36. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32(7):1466–1474. https://doi.org/10.1002/jcc.21707

37. Li X, Zhang Y, Chen H et al (2017) Insights into the molecular basis of the acute contact toxicity of diverse organic chemicals in the honey bee. J Chem Inf Model 57(12):2948–2957. https://doi.org/10.1021/acs.jcim.7b00476

38. Li X, Zhang Y, Chen H et al (2017) In silico prediction of chronic toxicity with chemical category approaches. RSC Adv 7(66):41330–41338. https://doi.org/10.1039/C7RA08415C

39. Yang H, Lou C, Li W et al (2020) Computational approaches to identify structural alerts and their applications in environmental toxicology and drug discovery. Chem Res Toxicol 33(6):1312–1322. https://doi.org/10.1021/acs.chemrestox.0c00006

40. Klekota J, Roth FP (2008) Chemical substructures that enrich for biological activity. Bioinformatics 24(21):2518–2525. https://doi.org/10.1093/bioinformatics/btn479

41. Korkmaz S (2020) Deep learning-based imbalanced data classification for drug discovery. J Chem Inf Model 60(9):4180–4190. https://doi.org/10.1021/acs.jcim.9b01162

42. Jing XY, Zhang X, Zhu X et al (2021) Multiset feature learning for highly imbalanced data classification. IEEE Trans Pattern Anal Mach Intell 43(1):139–156. https://doi.org/10.1109/TPAMI.2019.2929166

43. Willighagen EL, Mayfield JW, Alvarsson J et al (2017) The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching. J Cheminformatics. https://doi.org/10.1186/s13321-017-0220-4

44. Ringnér M (2008) What is principal component analysis? Nat Biotechnol 26(3):303–304. https://doi.org/10.1038/nbt0308-303

45. Thormann M, Vidal D, Almstetter M et al (2007) Nomen est omen: quantitative prediction of molecular properties directly from IUPAC names. The Open Applied Informatics J. https://doi.org/10.2174/1874136300701010028

46. Xu P, Hu G, Luo C et al (2016) DNA methyltransferase inhibitors: an updated patent review (2012–2015). Expert Opin Ther Pat 26(9):1017–1030. https://doi.org/10.1080/13543776.2016.1209488

47. Goldstein RS, Rickert DE (1985) Relationship between red blood cell uptake and methemoglobin production by nitrobenzene and dinitrobenzene in vitro. Life Sci 36(2):121–125. https://doi.org/10.1016/0024-3205(85)90090-6

48. Carey PJ (2003) Drug-induced myelosuppression. Drug Saf 26(10):691–706. https://doi.org/10.2165/00002018-200326100-00003

Authors and Affiliations

Yuqing Hua[1,2] · **Yinping Shi**[2] · **Xueyan Cui**[2] · **Xiao Li**[2,3]

1. School of Pharmacy, Shandong First Medical University, Taian 271000, China

2. Department of Clinical Pharmacy, The First Affiliated Hospital of Shandong First Medical University & Shandong Provincial Qianfoshan Hospital, Jinan 250014, China

3. Department of Clinical Pharmacy, Shandong Provincial Qianfoshan Hospital, Shandong University, Jinan 250014, China