Interdisciplinary Sciences: Computational Life Sciences
https://doi.org/10.1007/s12539-020-00407-2

ORIGINAL RESEARCH ARTICLE

MutagenPred-GCNNs: A Graph Convolutional Neural Network-Based Classification Model for Mutagenicity Prediction with Data-Driven Molecular Fingerprints

Shimeng Li[1] · Li Zhang[1,2,3] · Huawei Feng[1] · Jinhui Meng[1] · Di Xie[1] · Liwei Yi[4] · Isaiah T. Arkin[5] · Hongsheng Liu[1,2,3]

Received: 6 May 2020 / Revised: 24 November 2020 / Accepted: 3 December 2020

Abstract

An important task in the early stage of drug discovery is the identification of mutagenic compounds. Mutagenicity prediction models that can interpret relationships between toxicological endpoints and compound structures are especially favorable. In this research, we used an advanced graph convolutional neural network (GCNN) architecture to identify the molecular representation and develop predictive models based on these representations. The predictive model based on features extracted by GCNNs can not only predict the mutagenicity of compounds but also identify the structure alerts in compounds. In fivefold cross-validation and external validation, the highest area under the curve was 0.8782 and 0.8382, respectively; the highest accuracy (Q) was 80.98% and 76.63%, respectively; the highest sensitivity was 83.27% and 78.92%, respectively; and the highest specificity was 78.83% and 76.32%, respectively. Additionally, our model also identified some toxicophores, such as aromatic nitro, three-membered heterocycles, quinones, and nitrogen and sulfur mustard. These results indicate that GCNNs could learn the features of mutagens effectively. In summary, we developed a mutagenicity classification model with high predictive performance and interpretability based on a data-driven molecular representation trained through GCNNs.

Keywords Graph convolutional networks · Mutagenicity prediction · Deep learning

1 Introduction

The importance of compound toxicity prediction in the drug discovery process is obvious. The mutagenic potential of compounds is one of the safety parameters that should be considered [1, 2]. The Ames test, a bacterial assay, is now widely used to determine the ability of molecules to cause mutations [3, 4]. The Ames test takes only 2 days to generate results. From the perspective of in vivo experiments, the Ames test is the best choice considering its advantages. However, it still cannot meet the needs of predicting the mutagenicity of a large number of compounds before synthesizing the drug structure. The method of predicting mutagenicity based on molecular structures has gradually become the mainstream method in recent years.

Initially, experts identified mutagenicity by exploring the high-quality quantitative structure–toxicity relationship (QSTR) of compounds. There are many expert systems for mutagenicity prediction, the most representative of which are DEREK for Windows [5], Leadscope Model Applier (LSMA) [6], CASE Ultra [7–9], and Toxtree [10]. Hillebrecht et al. [4] built an Ames mutagenicity dataset containing

Shimeng Li and Li Zhang have contributed equally to this work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12539-020-00407-2.

✉ Hongsheng Liu
liuhongsheng@lnu.edu.cn

1 School of Life Science, Liaoning University, Shenyang 110036, China

2 Research Center for Computer Simulating and Information Processing of Bio-Macromolecules of Liaoning Province, Shenyang 110036, China

3 Engineering Laboratory for Molecular Simulation and Designing of Drug Molecules of Liaoning, Shenyang 110036, China

4 School of Information, Liaoning University, Shenyang 110036, China

5 Department of Biological Chemistry, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat-Ram, 91904 Jerusalem, Israel

Published online: 27 January 2021

9681 compounds, which was used to test these systems, and the accuracy of these models ranged from 66.4 to 75.4%. Subsequently, the automated detection of mutagens has become popular. Zhang et al. [11] developed LightGBM based on the gradient boosting algorithm. The AUC of the model achieved 0.836 in CV and 0.786 in external validation. Priyanka et al. combined some frequent molecular features and machine learning models to develop a tool that can predict various toxicity endpoints [12]. The accuracy of the mutagenicity prediction model on this platform is 84%. Yang et al. [13] developed admetSAR 2.0 by implementing more than 40 models for molecular property prediction. Among them, the accuracy of the mutagenicity prediction model is only 57.3%. The application of neural networks in compound property prediction has been widely used [14–16] and has yielded some satisfactory results. The performance of the deep neural network (DNN) developed by Dahl is better than that of standard random forest (RF) methods with the dataset obtained by Merck [17]. Mayr et al. [18] developed the multi-task DNN models and won the Tox21 Challenge. The feature used for training the model contains many different descriptors, including 3D and 2D descriptors, predefined toxicophores, and extended connectivity fingerprint descriptors (ECFPs), so the model can infer self-features during training. Koutsoukas et al. [19] found that the performance of DNN was significantly better than that of some frequently used machine learning models, such as support vector machine (SVM) and RF. These results demonstrate that, compared with the conventional machine learning methods, the DNN has obvious advantages.

For drug discovery applications, representing molecules by encoding their structural information is a major problem in cheminformatics and machine learning. Molecular representations are extremely diverse, and the most widely used representations are molecular fingerprints [20]. The molecular fingerprints can be regarded as pre-defined structural fragments [21], such as PubChem fingerprints containing 881 bits that represent the common substructures [22]. The most advanced fingerprints are ECFPs [23], which are a modification of the Morgan algorithm [24]. Recently, deep learning has become popular in various fields, and it has also affected computational chemistry, such as for computer-aided drug development and molecular property prediction [25–28]. Existing deep learning models for drug development usually take pre-defined molecular descriptors (mostly ECFPs) as input features. However, this approach is not compatible with representation learning. Recent deep learning focuses more on end-to-end methods. In these methods, the molecular representations are directly learned from strings or graphs [11, 29]. Bruna et al. [30] used spectral convolutional neural networks (CNNs) on molecular graphs to learn molecular features more effectively. Masci et al. proposed a convolutional network on non-Euclidean manifolds, which is an approach for combining local structure descriptors into global descriptors [31]. Duvenaud et al. [32] proposed an architecture based on CNN that can generalize fingerprints automatically. Goh et al. [33] developed Chemception, which can predict its chemical properties directly through 2D molecular images. Their work illustrated that this method can more accurately predict compound solubility and photovoltaic efficiency. Using automatic representations, we can develop mutagenicity prediction models without any handcrafted rules.

In this study, we built a novel mutagenicity prediction model with high predictive performance and interpretability based on a data-driven molecular representation trained through GCNN architectures. We also investigated the model's interpretability and explored the learned fragments in this study. We hope that this model could screen potential mutagens as soon as possible in the drug discovery process.

2 Materials and Methods

2.1 Dataset

In this study, we used two datasets, namely, the training set and the external validation set. The model was built through the training set, and the generalization performance of the model was evaluated through the external validation set. The training set was the Ames mutagenicity benchmark dataset developed by Hansen et al. [34]. The dataset was established in 2009, and recorded the experimental results of the Ames mutagenicity test of the compound in the Chemical Carcinogenesis Research Information System (CCRIS), the GeneTox, the VITIC, and three other studies [35–37]. In the training set, the chemicals are labeled as positive (active) and negative (inactive). To build predictive models more robustly, we filtered mixtures, polymers, compounds containing heavy metals, and compounds with less than 3 carbon atoms. Finally, we obtained 6307 compounds in the training dataset, including 3407 mutagens and 2900 nonmutagens, which were used to build the classification.

Additionally, we also constructed an external validation dataset, which was used to verify the generalization performance of the model in this research. This dataset gathered data from the CCRIS database, the National Toxicology Program (NTP) database, and Instituto Superiore di Sanita Salmonella Typhimurium (ISSSTY) database. To ensure the reliability of generalization performance verification, the external validation dataset needs to be independent of the training set completely. Then, we removed some compounds with the same filtering criteria as those in the training set. Finally, there were 1383 compounds in the external dataset, 703 of which were mutagens and 680 of which were nonmutagens. An overview of the datasets used in this study is

provided in Table 1. The compounds in the training set are shown in Table S1, and the external set of compounds is shown in Table S2.

2.2 Graph Convolutional Neural Networks

Considering the atoms in the compound as nodes and the bonds as edges, the molecule can be regarded as an undirected graph. The GCNN model merges information from distant atoms along the direction of the bond. The differentiable network layers use this information to generate identifiers for all substructures through adaptive learning, and the entire process is learnable. In this manner, useful representations can be extracted according to the present task.

This model first converts SMILES (Simplified Molecular Input Line Entry Specification) strings into molecular graphs with RDKit [38], which is an open-source cheminformatics software. SMILES is a specification describing molecular 3D structures with ASCII strings, which can encode molecules into human-readable strings. Then, a feature vector is assigned to each atom and the bonds attached to the atom. In this vector, features of the atom include the atom's element, the atom's degree, the number of hydrogen atoms connected to the atom, the aromaticity indicator, and the implicit valence, and the features of the bond include its type (single, double triple, or ring), and its conjugation. Considering the different contributions of neighboring atoms, the convolution operation can be performed at different levels, which is implemented through the graph convolutional modules consisting of graph convolutional layers, batch normalization layers, and graph pool layers. The convolved vectors are first aggregated through a fully connected dense layer and are then summed to form the final vector. The final vector is the neural fingerprint of the compounds, which is used to represent the molecules. Finally, the neural fingerprints are sent to a classification layer to obtain a score that provides a value or label indicating compound toxicity. The architecture is shown in Fig. 1. The code for this model is available at https://github.com/Liu-Lab-Lnu/MutagenPred_GCNNs.

With the exception of loss function and activity function, we optimized these hyper-parameters, which were FPL (length of fingerprint), FPD (depth of fingerprint), CKW (width of convolution kernel), HLS (number of the hidden units in the output layer), L2P (L2 penalty), IWS (scale of initial weight),

Table 1 Overview of the datasets used in this study

Dataset	Source	Compounds	Positive	Negative
Training	Hansen et al	6307	3407	2900
External	CCRIS, NTP, ISSSTY	1383	703	680

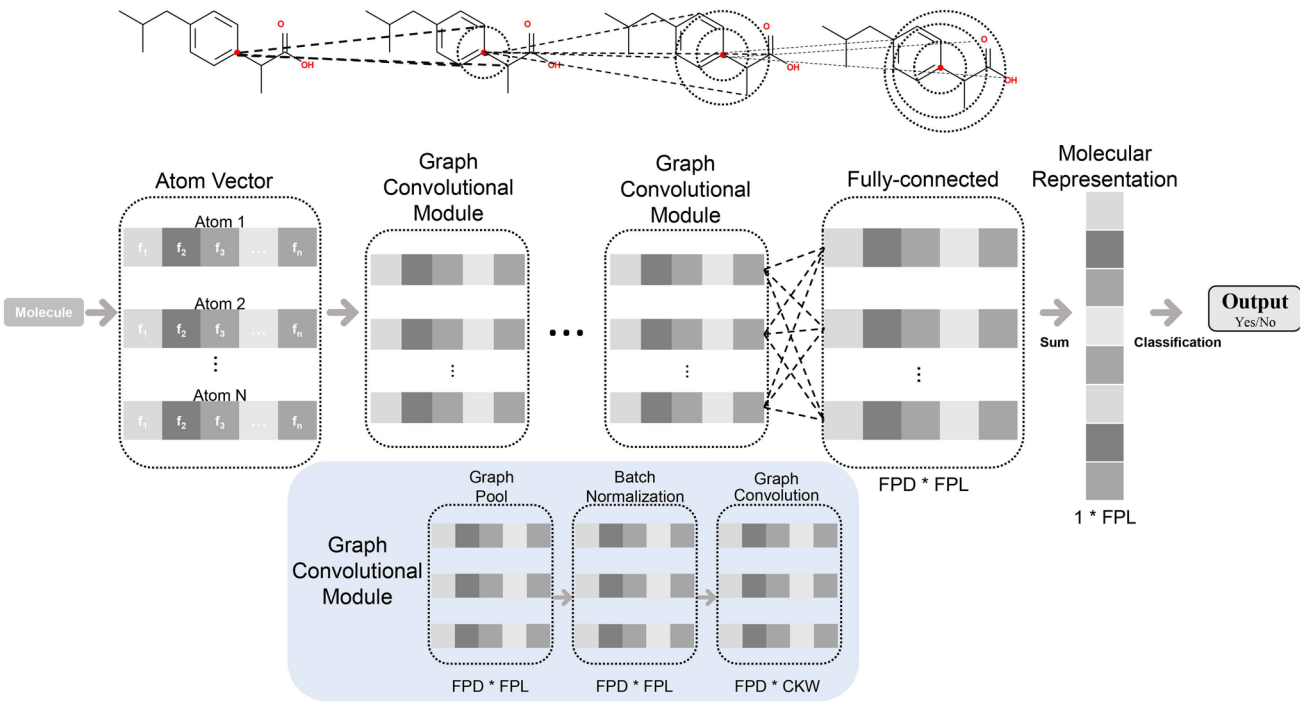

Fig. 1 The process of molecular encoding. First, the input layer assigns to each atom a vector of features, and the vectors then go through several graph convolutional modules consisting of a graph convolutional layer, a batch normalization layer, and a graph pool layer. The vectors generated from different convolution operations go through a fully connected dense layer and are then summed to form the final vector for the compound. Finally, the final vector was fed to the classification layers to obtain the predictive result

and LRS (learning rate step). We used the Tree of Parzen Estimators algorithm (TPE), which is a Bayesian optimization method, to search for the hyper-parameters in the GCNNs, and the process was implemented by hyper-opt [39], which is a Python library for serial and parallel optimization over search spaces. We used the TPE to generate 500 sets of hyper-parameters for the classification models, which means that we selected the best hyper-parameters from 500 sets, and all these sets were evaluated with fivefold CV. The optimized hyper-parameters and their ranges are shown in Table S3. The entire process of hyper-parameter optimization is also presented in Table S4.

In the optimization process, we had fixed the loss function and the active function to indicate mean square error (MSE) and rectified linear unit (ReLU), respectively. After the optimization, the following hyper-parameters were finally used in this study: FPL of 64, FPD of 4, CKW of 99, HLS of 240, L2P of e^{-1}, IWS of e^{-2}, and LRS of e^{-8}.

2.3 Performance Evaluation

There may be a risk of overfitting when different sets of hyper-parameters are evaluated. To avoid this problem, we adopted fivefold CV to evaluate the performance of these models. Specifically, the training set was randomly split into five parts of equal size, four parts of which were used to train the model, and the remaining part of which was used to evaluate the model. Thus, we obtained five models and indicators. Furthermore, to avoid the influence of random factors on the results, we repeated this process 10 times. Therefore, we obtained 50 performance indicators for each set of hyper-parameters. These 50 indicators were averaged to evaluate the performance of each model. The final models were also validated with an external validation dataset to evaluate the generalization performance of the models.

The model was evaluated with the following four indicators: Q, SPC, SEN, and AUC. Q is an indicator that measures the overall prediction accuracy of models. SPC measures the predictive accuracy for negative data, and SEN measures the predictive accuracy for positive data. These three indicators can be calculated from the numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The indicators were calculated as follows:

$$Q = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \times 100\%$$

$$\mathrm{SPC} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \times 100\%$$

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\%.$$

Compared with Q, AUC has become a more widely used performance evaluation metric for the binary classification model, because it is less affected by the threshold.

3 Results and Discussion

In this study, we built a novel mutagenic classification model using molecular representations generated by a GCNN. The training dataset for building the model was the benchmark dataset, which contained 6307 compounds. In the process of building the model, we searched for hyper-parameters with the hyper-opt in the fivefold CV.

3.1 Performance of the Models

In the process of searching for hyper-parameters, 500 models were generated with different topological networks, and in these models, different hyper-parameters caused large differences in AUC values in CV (0.8009–0.8807). The top 10 models with the highest AUC values in CV were selected (Table S5).

After completing the selection of hyper-parameters, all 10 sets evaluated by fivefold CV were selected and stored for external predictions with 10 repeats of well-trained weights. The test results in the fivefold CV are shown in Table 2. Of these values, the AUC, Q, SEN, and SPC in the test set were between 0.8759 and 0.8782, 80.37% and 80.98%, 82.28% and 83.27%, and 77.89% and 78.83%, respectively, indicating good predictive power. Furthermore, CM_4 (classification model 4) had the highest AUC (0.8782), Q (80.98%), and SEN (83.27%), indicating that CM_4 had the best predictive ability.

The external validation set contained 1383 compounds, which were used to evaluate the generalization performance of the model. Since these compounds were not used for model construction, these compounds can be used to evaluate the model's ability to predict the mutagenicity of

Table 2 Performance of the top 10 models in the CV

Model	AUC	Q (%)	SEN (%)	SPC (%)
CM_1	0.8768	80.68	83.09	77.89
CM_2	0.8762	80.37	82.36	78.09
CM_3	0.8774	80.67	82.28	78.83
CM_4	0.8782	80.98	83.27	78.34
CM_5	0.8774	80.79	82.91	78.35
CM_6	0.8768	80.73	82.42	78.78
CM_7	0.8761	80.56	82.77	78.00
CM_8	0.8772	80.65	82.96	77.98
CM_9	0.8759	80.56	82.80	77.98
CM_10	0.8774	80.67	82.28	78.83

new compounds. As shown in Table 3, the AUC, Q, SEN, and SPC of these top 10 models for external validation are between 0.8248 (CM_9) and 0.8382 (CM_1), 75.25% (CM_4) and 76.63%(CM_1), 75.78% (CM_3, CM_4, and CM_10) and 78.92% (CM_7), and 72.65% (CM_9) and 76.32% (CM_8), indicating that the GCNN models can label the new mutagens. The ROC curves of the 10 models in the training process (Fig. 2a) and the external validation process (Fig. 2b) also demonstrate that the performance of these models in external validation was lower than that in the CV. These results indicate that our model is slightly overfitting, which is a problem that needs to be addressed in our future work.

Table 3 Performance of the top 10 models in external validation

Model	AUC	Q (%)	SEN (%)	SPC (%)
CM_1	0.8382	76.63	78.77	74.41
CM_2	0.8345	75.54	78.21	72.79
CM_3	0.8321	75.69	75.78	75.59
CM_4	0.8274	75.25	75.78	74.71
CM_5	0.8364	75.69	76.07	75.29
CM_6	0.8339	76.19	77.78	74.56
CM_7	0.8315	75.83	78.92	72.65
CM_8	0.8344	76.48	76.64	76.32
CM_9	0.8248	75.62	77.07	74.12
CM_10	0.8321	75.69	75.78	75.59

3.2 Comparison with Previous Methods

In the present study, we compared the results of an objective performance assessment of some existing models with our models for mutagenicity prediction. Four of them are commercial models, which are DFW, LSMA, CASE Ultra, and Toxtree. The comparison result with commercial models is summarized in Fig. 3a, which shows that the model in this study achieved the highest accuracy of 80.98%. Toxtree has the highest SEN (85.2%), but the SEN of our model differs only slightly (1.9%) from that value. Compared with Toxtree, our model has a higher Q and SPC, and these results mean that our model has a better predictive ability and can also identify non-mutagens more accurately. Additionally, we also compared the performance of some machine learning prediction models with ours. The comparison results are shown in Fig. 3b. AdmetSAR and ProTox-II are freely available web servers, which are based on the machine learning models for the prediction of various toxicity endpoints. Xu et al. [40] combined five different molecular fingerprints (CDK fingerprints, Estate fingerprints, MACCS keys, PubChem fingerprints, and Substructure fingerprints) and five different machine learning algorithms (SVM, C4.5 decision tree, artificial neural network, k-nearest neighbor, and naïve Bayes). In Xu's paper, the best model was the SVM trained on PubChem fingerprints, so we adapted this model in Fig. 3c. Furthermore, none of these models are trained on the benchmark dataset. SVM-ECFP and SVM-FP7 are trained on ECFP fingerprints and fingerprints extracted from GCNNs, which are implemented by us. The SVM is trained on 10 deep fingerprints, which are extracted from CM 1 to CM 10 from Table 2. Notably, all performances in Fig. 3 are

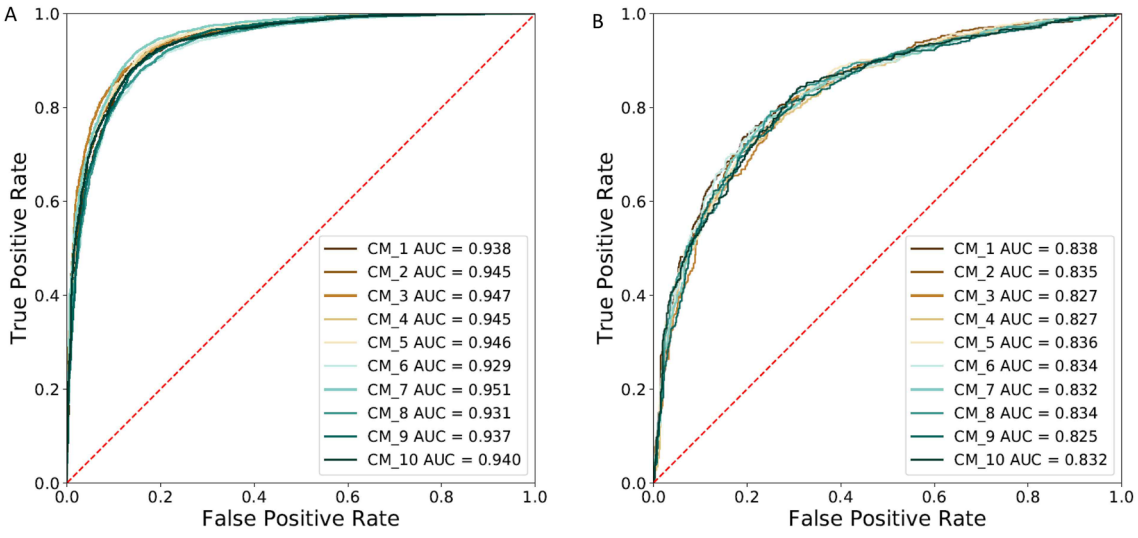

Fig. 2 The ROC curve of the top 10 models. **a** The ROC curve during the training process of these 10 models. **b** The ROC curve during the external validation process of these 10 models

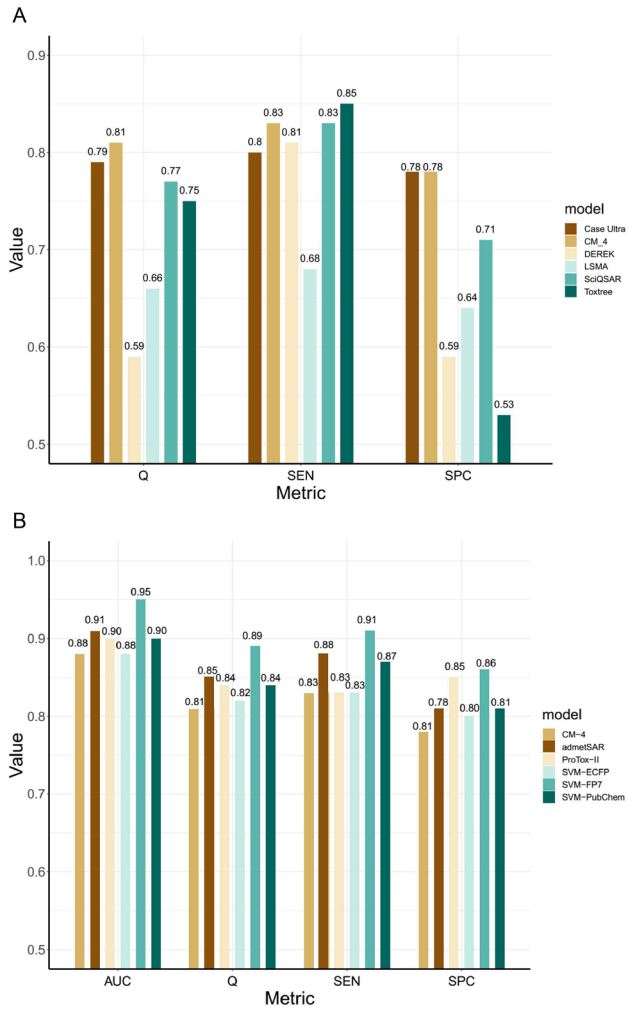

Fig. 3 Performance of the model in this study (CV) and some existing models. **a** The performance of the model in this study and some commercial models. **b** The performance of the models in this study and some machine learning models

the results of CV. From Fig. 3b, it is obvious that the SVM trained on fingerprints generated by GCNNs is better than others in CV, which means that the fingerprints generated by GCNNs perform better than PubChem and ECFP in CV. This phenomenon suggests that the GCNNs extracted useful substructures from the training data.

3.3 Construction of Conventional Classification Models with Extracted Deep Fingerprints

In this part, we evaluated whether the conventional machine learning models could make accurate predictions based on the fingerprints generated by GCNNs. We extracted the deep fingerprints generated by the GCNNs and developed some conventional machine learning models to suggest that the features learned by our models were effective. In this study, we used SVM, RF, KNN, and XGBoost as examples. To study this effect, we extracted deep fingerprints from the well-trained GCNNs, and transferred the entire data set into a matrix to featurize and vectorize the compounds. The matrix for the dataset, 6307 (number of compounds) * 64 (number of features), was regarded as inputs for SVM, RF, KNN, and XGBoost. The conventional machine learning models were implemented by the scikit-learn package [41]. The parameters of conventional machine learning methods were searched through a grid search implemented by the scikit-learn package, and the search process was also evaluated by the fivefold CV. The hyper-parameters and the ranges needed to be tuned are shown in Table S6, and the values are also shown. The models were also evaluated by fivefold CV and external validation. Finally, we summarize the parameter search results and the validation results of these conventional machine learning models. In the main manuscript, we take the SVM as an example, which is shown in Table 4, and the other models are shown in Table S7 to Table S9.

The AUC values in CV and external validation were between 0.9223 and 0.9502 and between 0.8106 and 0.8265, respectively. SVM_FP7 had the best AUC, 0.9502, in fivefold CV, but SVM_FP1 had the best AUC of 0.8265 in the external set. In both fivefold CV and external validation, the AUC values were greater than 0.8, which means that the SVM developed with deep fingerprints can classify the mutagens correctly. This means that the deep fingerprints generated by our models were effective. Moreover, we developed these traditional machine learning models via ECFPs, which were calculated with RDKit. The hyper-parameters of these models have also been tuned specifically. Compared with the same model prediction based on ECFP, however, we found that the performance of deep fingerprints was lower in external validation. The phenomenon cannot prove that the deep fingerprints are invalid, but illustrates that the generalization ability of deep fingerprints is weaker than that of ECFPs.

3.4 Analysis of Structural Alerts of Mutagens

In this section, the exploration of the fingerprints generated by the GCNNs further verified the effectiveness of GCNNs and showed the potential of GCNNs to predict mutagens and reveal the toxicophores. We verified the toxicophores found by GCNNs with confirmed toxicophores. We predicted the molecules in the external validation dataset with different machine learning models. In the external validation set, 390 molecules were correctly predicted by all models. There are 29 molecules, and only GCNN provided the correct prediction results. After the prediction, we clustered the molecules and compared them with the structure alerts (SAs) in ToxAlerts (TAs) [42] manually. Kazius et al. [35] discovered eight different substructure representations. At

Table 4 Performances and parameters of SVM with deep fingerprints and ECFP

Model	*C*	Gamma	AUC-CV[a]	AUC-EXT[b]	Q-CV[a]	Q-EXT[b]
SVM-FP1	100	0.01	0.9394	0.8294	87.09	76.63
SVM-FP2	1000	0.001	0.9475	0.8272	88.11	75.69
SVM-FP3	10	0.01	0.9484	0.8230	87.82	74.38
SVM-FP4	10	0.01	0.9453	0.8242	87.60	75.54
SVM-FP5	10	0.01	0.9467	0.8285	87.78	75.47
SVM-FP6	10	0.01	0.9271	0.8325	85.89	76.41
SVM-FP7	10	0.01	0.9526	0.8247	88.74	75.83
SVM-FP8	10	0.01	0.9274	0.8282	85.67	76.12
SVM-FP9	100	0.01	0.9392	0.8226	87.00	75.76
SVM-FP10	10	0.01	0.9405	0.8235	87.03	75.40
SVM-ECFP	1	0.1	0.8766	0.8659	81.75	79.88

[a]Performance in cross-validation

[b]Performance in external validation

least 70 mutagens were detected in their dataset with these eight substructure representations, with an accuracy of over 70%. These eight substructures are the general toxicophores. The prediction results demonstrated that the GCNN model could identify the general toxicophores recorded in the TAs, such as aromatic nitro (TA321), three-membered heterocycles (TA323), azo-type (TA326), and polycyclic aromatic system (TA328), as shown in Fig. 4a. Among these SAs, the TA321 is a symbolic toxicophore for mutagenicity, and the mutagenic mechanism of compounds containing TA321 has also been explained and verified. The TA323, TA326, and TA328 are similar to existing toxicophores [43, 44]. Additionally, the TAs have also reported 19 specific toxicophores derived from the general toxicophores but that are more complex [35]. The GCNNs can identify not only the general SAs but also the specific SAs, such as quinones (TA369), nitrogen and sulfur mustard (TA344), and alkyl esters of phosphonic or sulphonic acids (TA426). The molecules with these SAs are shown in Fig. 4b. Chesis et al. detected the mutagenicity of TA369 in the TA104 Salmonella test strain [45]. TA344 possessed significant intrinsic reactivity, and it was proven to be a specific toxicophore [35]. Furthermore, Ashby et al. classified TA426 as positive by structural criteria [46, 47]. As shown in Fig. 4c, some molecules without recorded SAs are labeled correctly, which means that our GCNNs may identify some SAs that have not

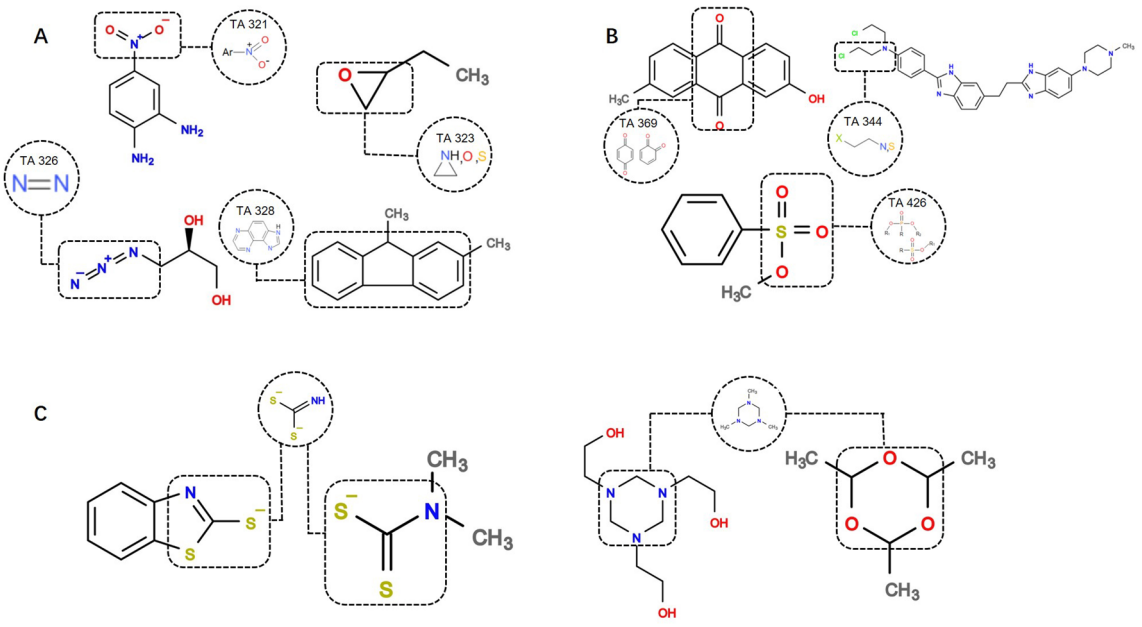

Fig. 4 The SAs of mutagens in the external validation dataset. **a** The mutagens are predicted correctly by the models. **b** The mutagens were labeled correctly by the GCNNs and with SAs recorded in TAs. **c** The mutagens labeled correctly by the GCNNs without SAs recorded in TAs

been associated with mutagenicity. This phenomenon demonstrates that our GCNNs may help to identify mutagenicity SAs without in vivo experiments. Considering the sensitivity of our models to the existing SAs, our model can provide the basis for some new, unknown SAs.

4 Conclusion

This study employed a GCNN-based architecture to represent the molecular structures and applied it to the task of compound mutagenicity prediction. Compared with reported classification models, the CM_4 model exhibited outstanding performance (AUC of 0.8782, Q of 80.98%, SEN of 83.27%, and SPC of 78.34% in CV).

In addition to the predictive capability, this paper focused on the interpretability of these models. In this study, deep fingerprints extracted from the classifiers were able to support conventional machine learning classification models very well. General SVM, RF, KNN, and XGBoost models using these deep fingerprints had outstanding AUC (0.9526, 0.9268, 0.9426, and 0.9472, respectively) and Q (88.74%, 85.90%, 87.30%, and 87.98%, respectively) values in CV. In addition, we also investigated the activation fragments. Compared with the toxic fragments reported in other studies, we found that these deep fingerprints are very similar to them. Considering the success of the GCNNs in molecular mutagenicity prediction, it can be foreseen that GCNNs are also very promising in other toxicity prediction tasks.

References

1. Parasuraman S (2011) Toxicological screening. J Pharmacol Pharmacother 2(2):74–79. https://doi.org/10.4103/0976-500X.81895
2. Segall MD, Chris B (2014) Addressing toxicity risk when designing and selecting compounds in early drug discovery. Drug Discov Today 19(5):688–693. https://doi.org/10.1016/j.drudis.2014.01.006
3. Ames BN, Lee FD, Durston WE (1973) An improved bacterial test system for the detection and classification of mutagens and carcinogens. PNAS 70(6):1903–1903. https://doi.org/10.1073/pnas.70.3.782
4. Hillebrecht A, Muster W, Brigo A, Kansy M, Weiser T, Singer T (2011) Comparative evaluation of in silico systems for ames test mutagenicity prediction: scope and limitations. Chem Res Toxicol 24(6):843–854. https://doi.org/10.1021/tx2000398
5. Lhasa Ltd. L, UK DEREK for Windows. http://www.lhasalimited.org
6. Leadscope Inc. C, OH. Leadscope Model Applier. http://www.leadscope.com
7. MultiCASE Inc. B, OH. MultiCASE. http://www.multicase.com
8. Saiakhov RD, Chakravarti S, Fuller MA, Klopman G (2011) Case ultra: an expert system for computational toxicology with a novel approach for improving risk assessment of chemicals. Toxicol Lett. https://doi.org/10.1016/j.toxlet.2011.05.355
9. Saiakhov R, Chakravarti S, Klopman G (2013) Effectiveness of CASE ultra expert system in evaluating adverse effects of drugs. Mol Inform 32(1):87–97. https://doi.org/10.1002/minf.201200081
10. Benigni R, Bossa C, Tcheremenskaia O (2013) Nongenotoxic carcinogenicity of chemicals: mechanisms of action and early recognition through a new set of structural alerts. Chem Rev 113(5):2940–2957. https://doi.org/10.1021/cr300206t
11. Zhang J, Mucs D, Norinder U, Svensson F (2019) LightGBM: an effective and scalable algorithm for prediction of chemical toxicity-application to the Tox21 and mutagenicity datasets. J Chem Inf Model 59(10):4150–4158. https://doi.org/10.1021/acs.jcim.9b00633
12. Priyanka B, Eckert AO, Schrey AK, Robert P (2018) ProTox-II: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Res 46(W1):W257–W263. https://doi.org/10.1093/nar/gky318
13. Hongbin Y, Chaofeng L, Lixia S, Jie L, Yingchun C (2019) admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. Bioinformatics 35(6):1067–1069. https://doi.org/10.1093/bioinformatics/bty707
14. Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. Mol Pharm 13(7):2524–2530. https://doi.org/10.1021/acs.molpharmaceut.6b00248
15. Pan Z, Yu W, Yi X, Khan A, Yuan F, Zheng Y (2019) Recent progress on generative adversarial networks (GANs): a survey. IEEE Access 7:36322–36333. https://doi.org/10.1109/ACCESS.2019.2905015
16. Khan A, Sohail A, Zahoora U, Qureshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 53:5455–5516. https://doi.org/10.1007/s10462-020-09825-6
17. Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V (2015) Deep neural nets as a method for quantitative structure–activity relationships. J Chem Inf Model 55(2):263–274. https://doi.org/10.1021/ci500747n
18. Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) DeepTox: toxicity prediction using deep learning. Front environ sci 3:80. https://doi.org/10.3389/fenvs.2015.00080
19. Koutsoukas A, Monaghan KJ, Li X, Huan J (2017) Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. J Cheminform 9(1):42. https://doi.org/10.1186/s13321-017-0226-y
20. Todeschini R, Consonni V, Mannhold R, Kubinyi H, Folkers G (2009) Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices references. Methods and principles in medicinal chemistry. Wiley, Hoboken. https://doi.org/10.1002/9783527628766
21. Shen J, Cheng F, Xu Y, Li W, Tang Y (2010) Estimation of ADME properties with substructure pattern recognition. J Chem Inf Model 50(6):1034–1041. https://doi.org/10.1021/ci100104j
22. Cw YAP (2010) Software news and update 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
23. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. J Chem Inf Model 50(5):742–754. https://doi.org/10.1021/ci100050t
24. Morgan H (1965) The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service. J Chem Doc 5(2):107–113. https://doi.org/10.1021/c160017a018

25. Li H, Liang Y, Xu Q (2009) Support vector machines and its applications in chemistry. Chemom Intell Lab Syst 95(2):188–198. https://doi.org/10.1016/j.chemolab.2008.10.007
26. Hautier G, Fischer CC, Jain A, Mueller T, Ceder G (2010) Finding nature's missing ternary oxide compounds using machine learning and density functional theory. Chem Mater 22(12):3762–3767. https://doi.org/10.1021/cm100795d
27. Müller K-R, Rätsch G, Sonnenburg S, Mika S, Grimm M, Heinrich N (2005) Classifying 'drug-likeness' with kernel-based learning methods. J Chem Inf Model 45(2):249–253. https://doi.org/10.1021/ci049737o
28. Bartók AP, Gillan MJ, Manby FR, Csányi G (2013) Machine-learning approach for one-and two-body corrections to density functional theory: applications to molecular and condensed water. Phys Rev B 88(5):054104. https://doi.org/10.1103/PhysRevB.88.054104
29. Preuer K, Klambauer G, Rippmann F, Hochreiter S, Unterthiner T (2019) Interpretable deep learning in drug discovery. Explainable AI: interpreting, explaining and visualizing deep learning. Springer, Berlin. https://doi.org/10.1007/978-3-030-28954-6_18
30. Bruna J, Zaremba W, Szlam A, LeCun Y (2013) Spectral networks and locally connected networks on graphs. https://arxiv.org/abs/1312.6203
31. Masci J, Boscaini D, Bronstein MM, Vandergheynst P (2015) Geodesic convolutional neural networks on Riemannian manifolds. In: 2015 IEEE international conference on computer vision workshop (ICCVW), Santiago, pp 832–840. https://doi.org/10.1109/ICCVW.2015.112
32. Duvenaud DK, Maclaurin D, Iparraguirre J, Bombarell R, Hirzel T, Aspuru-Guzik A, Adams RP (2015) Convolutional networks on graphs for learning molecular fingerprints. In: Advances in neural information processing systems, vol 28. Curran Associates, Inc. http://papers.nips.cc/paper/5954-convolutional-networks-on-graphs-for-learning-molecular-fingerprints.pdf
33. Goh GB, Siegel C, Vishnu A, Hodas NO, Baker N (2017) Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR models. https://arxiv.org/abs/1706.06689
34. Katja H, Sebastian M, Timon S, Andreas S, Antonius TL, Thomas SH, Nikolaus H, Klaus-Robert M (2009) Benchmark data set for in silico prediction of Ames mutagenicity. J Chem Inf Model 49(9):2077–2081. https://doi.org/10.1021/ci900161g
35. Kazius J, McGuire R, Bursi R (2005) Derivation and validation of toxicophores for mutagenicity prediction. J Med Chem 48(1):312–320. https://doi.org/10.1021/jm040835a
36. Helma C, Cramer T, Kramer S, De Raedt L (2004) Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. J Chem Inf Comput Sci 44(4):1402–1411. https://doi.org/10.1021/ci034254q
37. Feng J, Lurati L, Ouyang H, Robinson T, Wang Y, Yuan S, Young SS (2003) Predictive toxicology: benchmarking molecular descriptors and statistical methods. J Chem Inf Comput Sci 43(5):1463–1470. https://doi.org/10.1021/ci034032s
38. Landrum G (2016) RDKit: open-source cheminformatics software. https://www.rdkit.org/
39. Bergstra J, Yamins D, Cox D (2013) Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In: Sanjoy D, David M (eds) Proceedings of the 30th international conference on machine learning, vol 1. PMLR, pp 115–123. https://doi.org/10.5555/3042817.3042832
40. Xu C, Cheng F, Chen L, Du Z, Li W, Liu G, Lee PW, Tang Y (2012) In silico prediction of chemical ames mutagenicity. J Chem Inf Model 52(11):2840–2847. https://doi.org/10.1021/ci300400a
41. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: machine learning in python. J Mach Learn Res. https://doi.org/10.1145/2786984.2786995
42. Sushko I, Salmina E, Potemkin VA, Poda G, Tetko IV (2012) ToxAlerts: a web server of structural alerts for toxic chemicals and compounds with potential adverse reactions. J Chem Inf Model 52(8):2310–2316. https://doi.org/10.1021/ci300245q
43. Fishbein L (2011) Potential industrial carcinogens and mutagens. Elsevier, Amsterdam. https://www.elsevier.com/books/potential-industrial-carcinogens-and-mutagens/fishbein/978-0-444-41777-0
44. Klopman G, Frierson MR, Rosenkranz HS (1990) The structural basis of the mutagenicity of chemicals in Salmonella typhimurium: the Gene-Tox Data Base. Mutat Res Fundam Mol Mech Mutagen 228(1):1–50. https://doi.org/10.1016/0027-5107(90)90013-T
45. Chesis L, Smith MT (1984) Mutagenicity of quinones: pathways of metabolic activation and detoxification. PNAS 81(6):1696–1700. https://doi.org/10.1073/pnas.81.6.1696
46. Ashby J, Tennant RW (1988) Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP 204(1):17–115. https://doi.org/10.1016/0165-1218(88)90114-0
47. Benigni R, Bossa C (2008) Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. Mutat Res Rev Mutat Res 659(3):248–261. https://doi.org/10.1016/j.mrrev.2008.05.003