DeepAmes: Deep learning-powered Ames test predictive models for regulatory application by harnessing a comprehensive and consistent dataset

Li, Ting; Liu, Zhichao; Thakkar, Shraddha; Roberts, Ruth; Tong, Weida
FDA/NCTR; Boehringer Ingelheim; FDA/CDER; ApconiX Ltd./University of Birmingham; FDA/NCTR

Abstract

Mutagenicity is an important endpoint to assess the safety of consumer products (e.g., drugs and environmental chemicals). It is often required by regulatory agencies, such as FDA, EPA, etc. Ames assay is a widely used method to assess mutagenicity of chemicals. However, Ames assay requires approximately 2 g of sample for a dose-finding study and main study, which poses a challenge for chemical compounds of small amount, such as impurity of a drug and food flavor. Therefore, in silico approaches such as Quantitative Structure-Activity Relationships (QSARs) have been widely used to predict Ames test results. International Council for Harmonization (ICH) also provides guidelines to conduct such in silico analysis for regulatory application. We report here a high performance and robust model, called DeepAmes which is developed with a novel deep learning approach, for its potential utility in regulatory science. Specifically, we developed DeepAmes with a large and consistent Ames dataset (>10,000 compounds). We compared DeepAmes with QSAR models from 5 standard Machine Learning (ML) methods. All the six models were evaluated and compared using a test set of 1,543 compounds. DeepAmes achieved the top performance in prediction. With respect to applicability domain, DeepAmes yielded the best and stable performance till the compounds were 30% outside of training domain. The regulatory application of a model is focused on context-of-use, where false negatives are of utmost concern since positive prediction could be eliminated by downstream methods such as experimental methods. Compared to 5 standard ML models, DeepAmes yielded the highest sensitivity of 0.47, which was a 19.5% improvement in comparison with the largest sensitivity of the other five ML methods on the test set. More importantly, we provided a revised version of DeepAmes with a high sensitivity of 0.87. This improvement in sensitivity is significant given the fact that there are only ~15% mutagenic compounds in the test set. In conclusion, DeepAmes has the potential for regulatory application of predicting the Ames test result.

Introduction

- It is crucial to assess the mutagenicity potency for chemicals, an important factor that triggers regulatory actions for both new and existing chemicals.
- The 21st century toxicology has been focused on alternative approaches, e.g.,
  - International Council for Harmonization (ICH) guidelines.
  - CDER assessment of drug impurity mutagenicity by QSAR Modeling
- Many QSAR models were developed for mutagenicity assessment.
  - Most of the models were developed with small data sets ranging from hundreds to thousands. Several models developed with large (---→) data sets (such as >=13k compounds) but under various mutagenicity testing guidelines.
  - Very few models discuss the applicability domain.
- In this study, we proposed a DeepAmes model
  - Trained with more than 10k compounds) with consistent test guidelines collected from the National Institute of Health Sciences of Japan (NIHS.
  - Having great potential in regulatory decision making with its defined applicability domain and clear context of use.

Materials and Methods

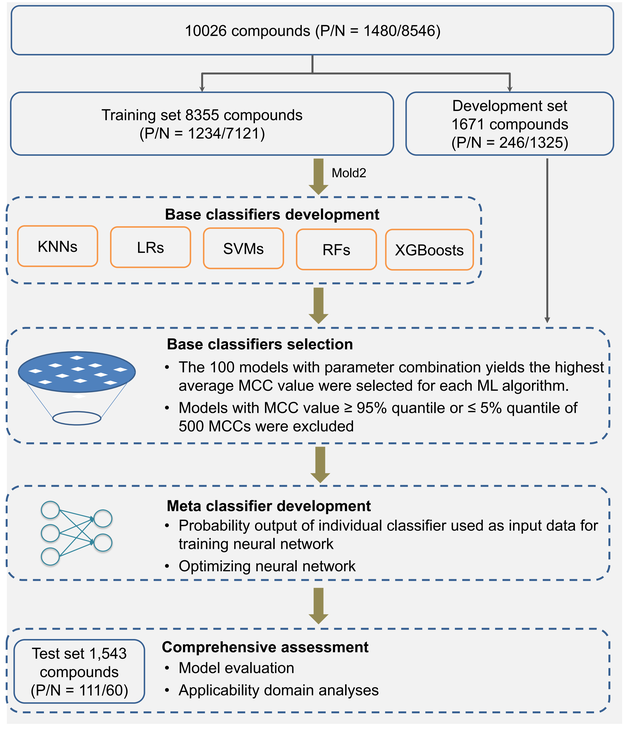

Figure 1. Figure 1. Overall workflow for the DeepAmes model including: **(1) Data preparation. (2) Base classifiers development:** Five algorithm was used to develop base classifiers. (3) **Base classifiers selection:** Select classifiers to generate model-level representation; **(4) Meta classifier development**: the probability prediction of the selected based classifiers was used to train the neural network. (5) **Comprehensive assessment and further improvement:** the test set was used to comprehensive evaluating of DeepAmes model.

Results and Discussion

1. DeepAmes outperformed other models on both MCC and sensitivity.

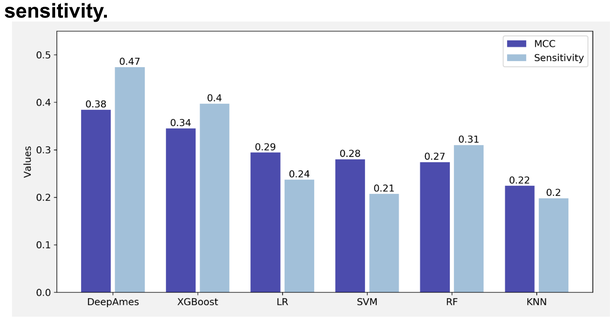

Figure 2. MCC and sensitivity of the six models on the test set.

2. DeepAmes yielded the highest MCC in both within and outside of the domain. The Euclidean distances of all the pairs of closest neighbors was first calculated for all the training compounds with the median value served as the threshold of applicability domain. There were 781 compounds inside the training domain while 762 compounds are outside of the domain. All the models performed better for the compounds that were within the training domain compared to these outsides of the domain.

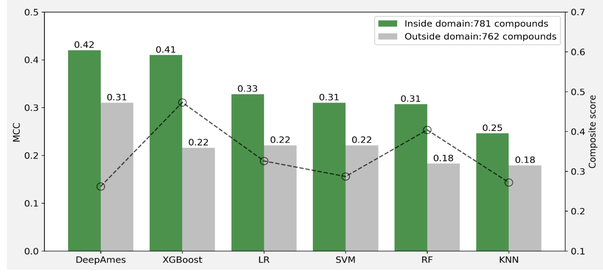

Figure 3. The MCC comparison between within and outside of the training domain on the test set.

3. DeepAmes hold the highest MCC till 30% beyond the training domain. We conducted a comparative analysis of applicability domain for six models by evaluating their performance on the compounds away from the training domain in every 5% incremental degree. As presented in Figure 2B, the MCC measures of all six models were generally decreased as the compounds furthering away from the training domain. The biggest drop was found on the compounds beyond 45% away from the training domain. DeepAmes hold the highest MCC till 30% beyond the training domain, while both KNN and SVM models hold a bit longer in applicability domain (35% and 40%, respectively).

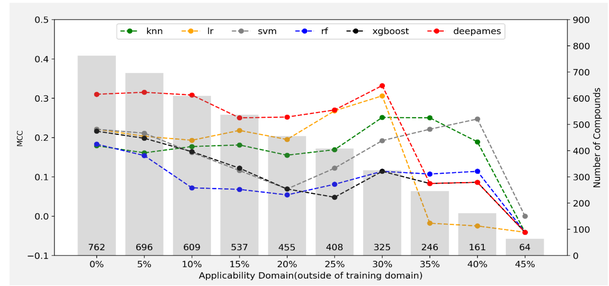

Figure 4: The MCC distribution on the compounds away from the training domain in every 5% incremental degree.

4. High sensitivity DeepAmes models with different penalized weight. To further improve sensitivity with aiming at regulatory application, we revised the DeepAmes model using the same framework of original approach. Specifically, we penalized with a higher "weight" on a wrong prediction for the positive class compared to the negative one. Since the number of negative compounds is six times of the positive compounds, we varied the weight by multiplying the prediction value with a number between seven to 18.

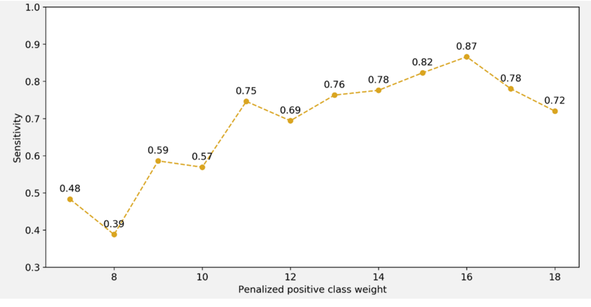

Figure 5: The plot of Sensitivity of DeepAmes against the positive class penalized weight on the test set.

Conclusion

- DeepAmes with model-Level representation outperformed molecule representation-based conventional machine learning models.
- With its defined applicability domain and clear context of use, DeepAmes has great potential in regulatory decision making.