Using Deep Neural Network to Predict Drug Sensitivity of Cancer Cell Lines

Yake Wang[1], Min Li[1(✉)], Ruiqing Zheng[1], Xinghua Shi[2], Yaohang Li[3], Fangxiang Wu[4], and Jianxin Wang[1]

[1] School of Information Science and Engineering, Central South University, Changsha, China
limin@mail.csu.edu.cn
[2] Department of Bioinformatics and Genomics, College of Computing and Informatics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[3] Department of Computer Science, Old Dominion University, Norfolk, VA 23529, USA
[4] Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, SK S7N5A9, Canada

Abstract. High-throughput screening technology has provided a large amount of drug sensitivity data for hundreds of compounds on cancer cell lines. In this study, we have developed a deep learning architecture based on these data to improve the performance of drug sensitivity prediction. We used a five-layer deep neural network, named as DeepPredictor, that integrated both genomic features of cell lines and chemical information of compounds to predict the half maximal inhibitory concentration on the Cancer Cell Line Encyclopedia (CCLE) dataset. We demonstrated the performance of our deep model using 10-fold cross-validations and leave-one-out strategies and showed that our model outperformed existing approaches.

Keywords: Cancer cell lines · Drug sensitivity · DeepPredictor
Deep learning · Predictive models

1 Introduction

Cultured Cancer cell lines are fundamental materials to study the molecular basis of drug activity [1] and to discover novel anticancer drugs in cancer biology research. Several large-scale high-throughput screening efforts have catalogued genomic information of a panel of in vitro cell lines as well as their drug sensitivity profiles against hundreds of compounds.

These datasets have been studied in many aspects, such as novel anticancer drug discovery [2], and anticancer drug repositioning [3]. In the context of drug discovery and repositioning, computational approaches have been widely used to predict drug sensitivity data over cancer cell lines. Though initial method Quantitative Structure-Activity Relationship (QSAR) was widely adopted, it could not generalize across cancer cell lines since the properties of cell lines were not considered. Menden et al. [4]

D.-S. Huang et al. (Eds.): ICIC 2018, LNCS 10955, pp. 223–226, 2018.
https://doi.org/10.1007/978-3-319-95933-7_27

made the first effort to integrate cell line genomic features and drug chemical features to model drug sensitivity data using a three-layer neural network and random forests (RF). Ammad-ud-din et al. [5] proposed an extended QSAR model via integrating drug properties matrix and cell line properties matrix to predict drug response data using Kernelized Bayesian Matrix Factorization (KBMF). [6] built a dual-layer integrated cell line-drug network to deduce drug response data. Cortes-Ciriano et al. [7] integrated the transcript profiles of genes and compounds' Morgan fingerprints to predict the drug responses using random forests (RF) and support vector machine (SVM).

In this study, we propose a novel deep learning architecture to predict drug sensitivity of cancer cell lines by integrating gene expression data of cell lines and compound fingerprints. The experimental results show that our approach outperforms the state-of-the-art approaches on drug sensitivity prediction.

2 Method

2.1 Dataset

We extracted 504 cell lines with response data against 24 drugs from CCLE dataset. To reduce the input feature dimension, we chose the top-1000 genes that display the highest variance across cell lines and normalized their expression data to zero mean and unit variance. We calculated compounds' Morgan fingerprints and used them as compounds' feature representations. The response data IC50 s were converted to $\ln IC50(\mu M)$ to compare with previous study [7]. The final data matrix contains 491 cell lines and 23 drugs with 96.25% completeness.

2.2 DeepPredictor

Deep learning has recently achieved remarkable success in many areas including computer vision [8], speech recognition [9], and natural language processing [10]. A feedforward deep network is one of the classic deep learning models. It contains several stacked layers with each neural unit connecting to all the units at next layer. Given enough neural units or deep layers, a feedforward deep network can become a universal approximator [11]. We can use a deep network to model drug sensitivity prediction considering the existence of large amount of drug pharmacological profiling data in a classification task. We train a feedforward deep neural network to model the continuous drug response data, which showed that this regression-based deep network outperformed start-of-the-art work in this field.

Specifically, our newly developed deep model, namely DeepPredictor, works as follows as illustrated in Fig. 1. We integrated the genomic information of cancel cell lines and chemical structures of compounds as the feature vectors, which were taken as the inputs of a feedforward neural network to predict the sensitivity data of corresponding cell line-compound pairs.

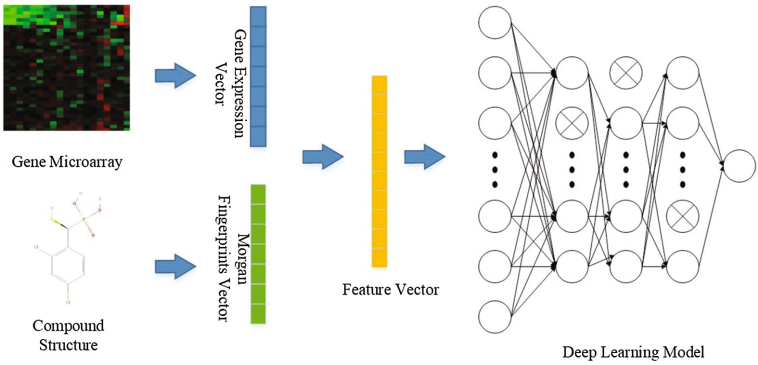

Fig. 1. The overall workflow of our DeepPredictor model.

3 Results

We performed 10-fold cross-validations to obtain the predictive performance measurement of our deep learning models. And we also left cell lines from each tissue out and each compound out to test the ability of the models trained on the rest data to predict sensitivity data on novel cell lines and compounds. The final RMSE and R^2 values were averaged.

We compared DeepPredictor with the state-of-the-art study [7], RF, on CCLE dataset. As summarized in Table 1, cross-validation results showed that DeepPredictor outperforms RF with lower RMSE value of 0.51 and higher R^2 value of 0.75. Leave-one-tissue-out and leave-one-compound-out experiments also showed that DeepPredictor outperformed RF on the CCLE dataset in predicting sensitivity data involving novel cell lines or compounds.

Table 1. Results on CCLE dataset.

Metrics	CV		LOTO		LOCO	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
RF	1.02 ± 0.05	0.74 ± 0.03	0.97 ± 0.26	0.75 ± 0.12	1.62 ± 1.32	0.18 ± 0.15
DeepPredictor	0.51 ± 0.06	0.75 ± 0.04	0.60 ± 0.22	0.68 ± 0.13	1.36 ± 1.56	0.04 ± 0.08

4 Conclusion

In this paper, we proposed DeepPredictor, a deep learning model, to predict drug sensitivity of cancer cell lines. The model was trained on CCLE dataset. And the results showed DeepPredictor outperformed state of the art models with the lowest prediction errors and high coefficient determination. The cross-validation results show that DeepPredictor has the best interpolation ability to fill in missing drug sensitivity values, and the LOTO and LOCO results show that DeepPredictor has the lowest extrapolation

error. We thus believe that DeepPredictor can benefit clinical cancer therapy and future study on drug sensitivity prediction.

Acknowledgement. We would like to thank Isidro Cortés Ciriano at Department of Biomedical Informatics, Harvard Medical School for providing the research data and discussing with us during research. This work is supported by the National Science Fund for Excellent Young Scholars under Grant No. 61622213, the National Natural Science Foundation of China under grant No. 61772552, and the Fundamental Research Funds for the Central Universities of Central South University under grant No. 2018zzts560 and No. 2018zzts028.

References

1. Weinstein, J.N.: Drug discovery: cell lines battle cancer. Nature **483**(7391), 544–545 (2012)
2. Wilding, J.L., Bodmer, W.F.: Cancer cell lines for drug discovery and development. Cancer Res. **74**(9), 2377–2384 (2014)
3. Jin, G., Wong, S.T.: Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. Drug Discov. Today **19**(5), 637–644 (2014)
4. Menden, M.P., Iorio, F., Garnett, M., et al.: Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. PLoS ONE **8**(4), e61318 (2013)
5. Ammad-ud-din, M., Georgii, E., Gonen, M., et al.: Integrative and personalized QSAR analysis in cancer by kernelized Bayesian matrix factorization. J. Chem. Inf. Model. **54**(8), 2347–2359 (2014)
6. Zhang, N., Wang, H., Fang, Y., et al.: Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model. PLoS Comput. Biol. **11**(9), e1004498 (2015)
7. Cortes-Ciriano, I., van Westen, G.J., Bouvier, G., et al.: Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. Bioinformatics **32**(1), 85–95 (2016)
8. Krizhevsky, A., Sutskever I., Hinton G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
9. Hinton, G., Deng, L., Yu, D., et al.: Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Sig. Process. Mag. **29**(6), 82–97 (2012)
10. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems, pp. 3104–3112 (2014)
11. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Netw. **2**(5), 359–366 (1989)