Conclusion

This study introduced a novel framework XGDP to predict response levels of anti-cancer drugs and discover underlying mechanism of action of drugs. To enhance the predictive power of GNN models, first we adapted the Morgan algorithm that is used for computing ECFPs to form our node features. Same procedures as Morgan algorithm were followed to identify the substructures of the molecule but the feature vector of each atom was assigned as the membership of the identified structures. Then we incorporated the type of chemical bonds as the edge features. These strategies enabled us to depict the molecule in a more meticulous manner and was testified to improve the GNN's prediction in terms of RMSE and PCC. Furthermore, we also attempted to explore relational GNN in the drug response prediction task, which describes edges as different relations and develops distinct message passing patterns for them. It was shown that RGCN outperformed GCN without edge features. However, due to the limited GPU resources, we were not able to train the RGAT model with an optimal batch size. This part of experiments is left for future investigations.

Moreover, we leveraged state-of-the-art attribution approaches in deep learning, GNNExplainer and Integrated Gradients, to explain our developed model. The explanations were visualized as saliency maps of both molecules and genes. Remarkably, those saliency maps could be supported by the SAR studies of the drugs. Consequently, we claim that our model is able to capture the significant functional groups of drugs and their potential targeted genes, and thus reveal the comprehensive mechanism of action of drugs. In the future, we intend to extend this study to a multi-omics level. Although genes contain the most vital information of the cause of disease, they do not directly interact with drugs in most cases. Therefore, protein and metabolites data should be considered. In addition, gene mutation and DNA methylation data may have a more direct reflection on the somatic abnormality, which are also expected to be explored in future works.

Data Availability

The drug response data can be downloaded from GDSC. And the gene expression data can be downloaded from CCLE under mRNA expression. Our implementation is released on Github (https://github.com/SCSE-Biomedical-Computing-Group/XGDP). Data preprocessing can be referred to our codes.

Received: 16 June 2024; Accepted: 11 December 2024
Published online: 02 January 2025

References

1. Singh, D. P. & Kaushik, B. A systematic literature review for the prediction of anticancer drug response using various machine learning and deep learning techniques. *Chem. Biol. Drug Des.* (2022).
2. Rafique, R., Islam, S. R. & Kazi, J. U. Machine learning in the prediction of cancer therapy. *Comput. Struct. Biotechnol. J.* **19**, 4003–4017 (2021).
3. Firoozbakht, F., Yousefi, B. & Schwikowski, B. An overview of machine learning methods for monotherapy drug response prediction. *Brief. Bioinform.* **23**, bbab408 (2022).
4. Baptista, D., Ferreira, P. G. & Rocha, M. Deep learning for drug response prediction in cancer. *Brief. Bioinform.* **22**, 360–379 (2021).
5. Partin, A. et al. Deep learning methods for drug response prediction in cancer: Predominant and emerging trends. arXiv preprint arXiv:2211.10442 (2022).
6. Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. Opportunities and challenges in phenotypic drug discovery: An industry perspective. *Nat. Rev. Drug Discov.* **16**, 531–543 (2017).
7. An, X., Chen, X., Yi, D., Li, H. & Guan, Y. Representation of molecules for drug response prediction. *Brief. Bioinform.* **23**, bbab393 (2022).
8. Heller, S. R., McNaught, A., Pletnev, I., Stein, S. & Tchekhovskoi, D. InChI, the IUPAC international chemical identifier. *J. Cheminform.* **7**, 1–34 (2015).
9. Weininger, D. Smiles. A chemical language and information system 1 introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).
10. Suphavilai, C., Bertrand, D. & Nagarajan, N. Predicting cancer drug response using a recommender system. *Bioinformatics* **34**, 3907–3914 (2018).
11. Liu, P., Li, H., Li, S. & Leung, K.-S. Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. *BMC Bioinform.* **20**, 1–14 (2019).
12. Chang, Y. et al. Cancer drug response profile scan (CDRscan): A deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci. Rep.* **8**, 1–11 (2018).
13. Durant, J. L., Leland, B. A., Henry, D. R. & Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **42**, 1273–1280 (2002).
14. Reutlinger, M. et al. Chemically advanced template search (cats) for scaffold-hopping and prospective target prediction for 'orphan'molecules. *Mol. Inf.* **32**, 133 (2013).
15. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
16. Li, M. et al. DeepDsc: A deep learning method to predict drug sensitivity of cancer cell lines. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **18**, 575–582 (2019).
17. Shao, J. et al. S2dv: Converting smiles to a drug vector for predicting the activity of anti-HBV small molecules. *Brief. Bioinform.* **23**, 593 (2022).
18. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013).
19. Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E. & Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* **55**, 263–274 (2015).
20. Sun, M. et al. Graph convolutional networks for computational drug development and discovery. *Brief. Bioinform.* **21**, 919–935 (2020).
21. Hu, L. et al. Dual-channel hypergraph convolutional network for predicting herb-disease associations. *Brief. Bioinform.* **25**, bbae067 (2024).
22. Zhao, B.-W. et al. A geometric deep learning framework for drug repositioning over heterogeneous information networks. *Brief. Bioinform.* **23**, bbac384 (2022).
23. Korolev, V., Mitrofanov, A., Korotcov, A. & Tkachenko, V. Graph convolutional neural networks as "general-purpose" property predictors: The universality and limits of applicability. *J. Chem. Inf. Model.* **60**, 22–28 (2019).

24. Xiong, Z. et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J. Med. Chem.* **63**, 8749–8760 (2019).
25. Nguyen, T., Nguyen, G. T., Nguyen, T. & Le, D.-H. Graph convolutional networks for drug response prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **19**, 146–154 (2021).
26. Liu, Q., Hu, Z., Jiang, R. & Zhou, M. DeepCDR: A hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics* **36**, i911–i918 (2020).
27. Zhu, Y. et al. TGSA: Protein–Protein association-based twin graph neural networks for drug response prediction with similarity augmentation. *Bioinformatics* **38**, 461–468 (2022).
28. Ma, T. et al. DualGCN: A dual graph convolutional network model to predict cancer drug response. *BMC Bioinform.* **23**, 129 (2022).
29. Zuo, Z. et al. SWnet: A deep learning model for drug response prediction from cancer genomic signatures and compound chemical structures. *BMC Bioinform.* **22**, 1–16 (2021).
30. Chen, D. et al. Algebraic graph-assisted bidirectional transformers for molecular property prediction. *Nat. Commun.* **12**, 1–9 (2021).
31. Bishop, C. M., Svensén, M. & Williams, C. K. GTM: The generative topographic mapping. *Neural Comput.* **10**, 215–234 (1998).
32. Yoshimori, A. Prediction of molecular properties using molecular topographic map. *Molecules* **26**, 4475 (2021).
33. Ying, Z., Bourgeois, D., You, J., Zitnik, M. & Leskovec, J. Gnnexplainer: Generating explanations for graph neural networks. *Advances in Neural Information Processing Systems* **32** (2019).
34. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *International Conference on Machine Learning*, 3319–3328 (PMLR, 2017).
35. Yang, W. et al. Genomics of drug sensitivity in cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2012).
36. Cancer Cell Line Encyclopedia Consortium; Genomics of Drug Sensitivity in Cancer Consortium. Pharmacogenomic agreement between two cancer cell line data sets. *Nature* **528**, 84–87 (2015).
37. Wang, Y. et al. PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* **37**, W623–W633 (2009).
38. RDKit: Open-source cheminformatics. http://www.rdkit.org. [Online; Accessed 11-Apr.-2013].
39. Duan, Q. et al. L1000cds2: Lincs 1000 characteristic direction signatures search engine. *NPJ Syst. Biol. Appl.* **2**, 1–12 (2016).
40. Ramsundar, B. et al. *Deep Learning for the Life Sciences* (O'Reilly Media, 2019). https://www.amazon.com/Deep-Learning-Life-Sciences-Microscopy/dp/1492039837.
41. Vaswani, A. et al. Attention is all you need. *Advances in Neural Information Processing Systems* **30** (2017).
42. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016).
43. Veličković, P. et al. Graph attention networks. arXiv preprint arXiv:1710.10903 (2017).
44. Schlichtkrull, M. et al. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, 593–607 (Springer, 2018).
45. Busbridge, D., Sherburn, D., Cavallo, P. & Hammerla, N. Y. Relational graph attention networks. arXiv preprint arXiv:1904.05811 (2019).
46. Brody, S., Alon, U. & Yahav, E. How attentive are graph attention networks? arXiv preprint arXiv:2105.14491 (2021).
47. Kokhlikyan, N. *et al.* Captum: A unified and generic model interpretability library for pytorch. arXiv preprint arXiv:2009.07896 (2020).
48. Shrikumar, A., Greenside, P., Shcherbina, A. & Kundaje, A. Not just a black box: Learning important features through propagating activation differences. arXiv preprint arXiv:1605.01713 (2016).
49. Bach, S. et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS One* **10**, e0130140 (2015).
50. Shrikumar, A., Greenside, P. & Kundaje, A. Learning important features through propagating activation differences. In *International Conference on Machine Learning*, 3145–3153 (PMLR, 2017).
51. Yuan, H., Yu, H., Wang, J., Li, K. & Ji, S. On explainability of graph neural networks via subgraph explorations. In *International Conference on Machine Learning*, 12241–12252 (PMLR, 2021).
52. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. In Wallach, H. et al. (eds.) *Advances in Neural Information Processing Systems 32*, 8024–8035 (Curran Associates, Inc., 2019).
53. Fey, M. & Lenssen, J. E. Fast graph representation learning with pytorch geometric. arXiv preprint arXiv:1903.02428 (2019).
54. Berman, H. M. et al. The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
55. Lu, J., Xia, S., Lu, J. & Zhang, Y. Dataset construction to explore chemical space with 3d geometry and deep learning. *J. Chem. Inf. Model.* **61**, 1095–1104 (2021).
56. Solca, F. et al. Target binding properties and cellular activity of afatinib (BIBW 2992), an irreversible ErbB family blocker. *J. Pharmacol. Exp. Ther.* **343**, 342–350 (2012).
57. Canning, P. et al. Structural mechanisms determining inhibition of the collagen receptor ddr1 by selective and multi-targeted type ii kinase inhibitors. *J. Mol. Biol.* **426**, 2457–2470 (2014).
58. Abouantoun, T. J., Castellino, R. C. & MacDonald, T. J. Sunitinib induces PTEN expression and inhibits PDGFR signaling and migration of medulloblastoma cells. *J. Neuro-oncol.* **101**, 215–226 (2011).
59. Freshour, S. L. et al. Integration of the drug–gene interaction database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res.* **49**, D1144–D1151 (2021).
60. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
61. Fang, Z., Liu, X. & Peltz, G. GSEApy: A comprehensive package for performing gene set enrichment analysis in python. *Bioinformatics* **39**, btac757 (2023).

Acknowledgements

This research was supported by AcRF Tier-1 grant RG14/23 of Ministry of Education, Singapore.

Author contributions

C.W., A.K.G., and J.C.R. conceived the experiment(s), C.W. and A.K.G. conducted the experiment(s), C.W., J.C.R., and A.K.G. analysed the results. C.W. and J.C.R. wrote and reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.C.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.