

CHAPTER – 9

CONCLUSION AND FUTURE WORK

CHAPTER 9

CONCLUSION AND FUTURE WORK

9.1 Conclusion

The "Leveraging ANN for Targeted Drug Sensitivity Prediction on GDSC Data" project focused on creating a robust and user-friendly system for predicting cancer drug sensitivity based on genomic data. Leveraging the Genomics of Drug Sensitivity in Cancer (GDSC) dataset, this system utilized key genomic features such as gene mutations, tissue descriptors, and copy number alterations (CNAs) to predict IC50 values, a crucial metric in assessing drug efficacy.

The ANN model demonstrated excellent performance with a high R-squared (R^2) value and low Mean Squared Error (MSE), confirming its ability to generalize well across diverse genomic data and accurately predict drug sensitivity. The data preprocessing pipeline, including normalization, feature scaling, and one-hot encoding, was crucial in preparing the data for effective model training, ensuring the model could handle genomic features efficiently. Hyperparameter tuning and early stopping techniques were applied to optimize performance and prevent overfitting, enhancing the model's stability and generalizability.

A significant contribution of this project is the creation of an intuitive Flask-based web interface that allows users clinicians and researchers alike to easily input genomic data and receive real-time IC50 predictions. The interface streamlines the process by using dropdowns and input fields for data entry, ensuring accessibility and ease of use. The system is capable of comparing the predicted IC50 values for multiple drugs, assisting users in selecting the best treatment options based on drug efficacy.

In conclusion, this project demonstrates the practical application of ANN models in personalized cancer treatment by predicting drug responses based on genomic data. By integrating accurate predictions with a user-friendly interface, this system makes precision oncology more accessible and applicable for both clinical and research settings.

9.2 Future Work

While the current version of the system delivers accurate drug response predictions, several areas for improvement and expansion remain:

1. **Incorporating Multi-Omics Data:** The current model relies on genomic features from the GDSC dataset. Future versions could integrate additional multi-omics data, such as proteomics and metabolomics, to provide a more comprehensive view of tumor biology. This could improve prediction accuracy by capturing complex interactions within different layers of biological data.
2. **Expanding Data Sources:** To improve the generalizability of the model, future work could incorporate data from additional sources, such as the Cancer Therapeutics Response Portal (CTRP) and Patient-Derived Xenografts (PDX). Combining data from these datasets will introduce greater diversity, allowing the model to better handle different cancer types and subtypes, thereby improving the reliability of IC50 predictions.
3. **Implementing Advanced Feature Selection and Dimensionality Reduction:** Although the current feature selection process yields good results, further optimization using advanced techniques like autoencoders or Principal Component Analysis (PCA) could be beneficial. These methods could reduce the dimensionality of the input features, potentially improving computational efficiency while maintaining or even enhancing model performance.
4. **Enhancing Model Interpretability:** To increase clinician trust in the model's predictions, future work could incorporate explainable AI methods such as Shapley values or Layer-wise Relevance Propagation (LRP). These techniques would provide insights into how specific genomic features influence IC50 predictions, helping clinicians make more informed treatment decisions.
5. **Optimizing Web Application for Cloud Deployment:** Currently, the application is deployed locally. Transitioning to cloud-based deployment would allow for greater scalability, enabling wider access to the model across different locations. Cloud deployment would also support continuous integration and real-time updates, ensuring the system remains current and capable of handling an increasing number of users.
6. **Incorporating Personalized Drug Response Reports:** While the current interface provides valuable IC50 predictions, future versions could include features that generate personalized drug response reports for patients. This could enhance the real-world applicability of the system, making it easier for clinicians to deliver tailored treatment recommendations based on the model's predictions.

7. **Improving Model Training and Validation:** Future work could refine the model training process by incorporating techniques like K-fold cross-validation and regularization methods. These would help ensure that the model performs consistently across different data segments and prevents overfitting, further improving its reliability and robustness.

By addressing these areas, the project can continue to evolve, increasing its complexity, accuracy, and user accessibility. The integration of multi-omics data, the expansion of data sources, and the enhancement of model interpretability will make the system even more powerful, supporting the development of personalized cancer treatments. The future work outlined here underscores the potential of this predictive model to drive forward personalized oncology and improve clinical decision-making.

REFERENCES

- [1] Katyna Sada Del Real and Angel Rubio. (2023). "Discovering the Mechanism of Action of Drugs with a Sparse Explainable Network", *Journal of Biomedical Informatics*, 128, 104019.
- [2] Alexander Partin, Thomas S. Brettin, Yitan Zhu, Oleksandr Narykov, Austin Clyde, Jamie Overbeek, and Rick L. Stevens. (2023). "Deep Learning Methods for Drug Response Prediction in Cancer: Predominant and Emerging Trends", *Nature Reviews Drug Discovery*, 22(4), 317-334.
- [3] Bara A. Badwan, Gerry Liaropoulos, Efthymios Kyrodimos, Dimitrios Skaltsas, Aristotelis Tsirigos, and Vassilis G. Gorgoulis. (2023). "Machine Learning Approaches to Predict Drug Efficacy and Toxicity in Oncology", *Computational and Structural Biotechnology Journal*, 21, 3187-3202.
- [4] Brent M. Kuenzi, Jisoo Park, Samson H. Fong, Kyle S. Sanchez, John Lee, Jason F. Kreisberg, Jianzhu Ma, and Trey Ideker. (2020). "Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells", *Cell Reports*, 32(6), 108053.