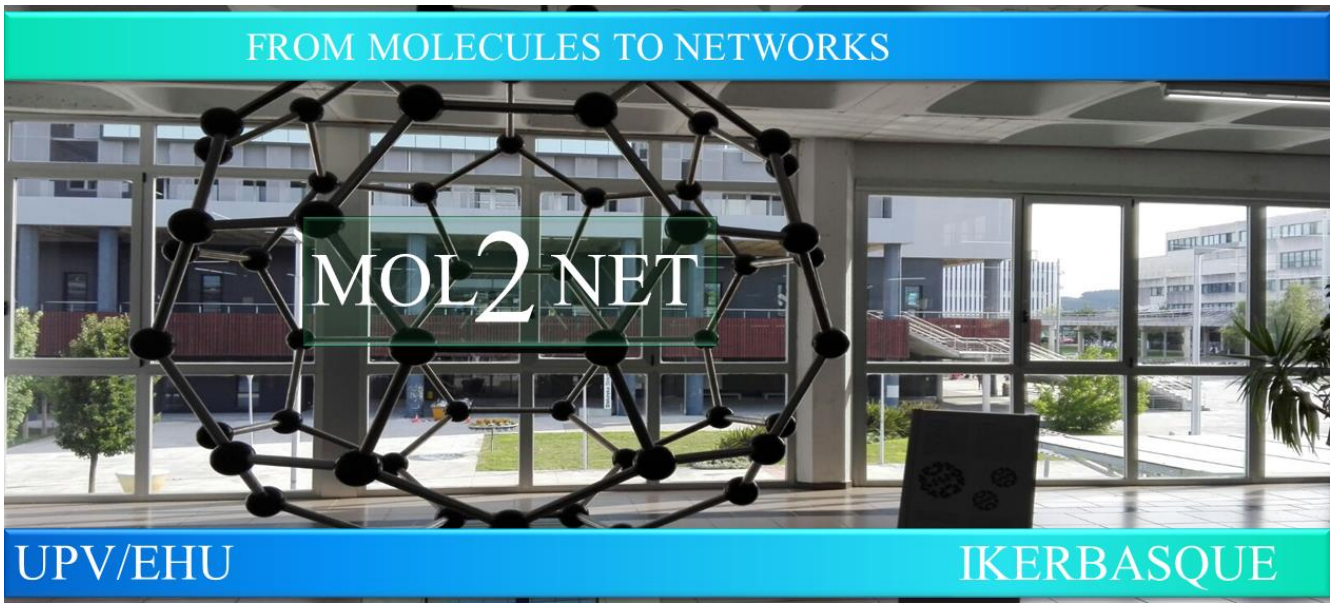
MOL2NET'23, Conference on Molecular, Biomedical & Computational Sciences and Engineering, 9th ed.




# Chemoinformatics and Anticancer Drug Prediction

*Maider Baltasar Marchueta, PhD student*

*Department of Organic and Inorganic Chemistry, Faculty of Science and Technology, University of Basque Country (UPV/EHU), Leioa, Biscay, Spain*
*IKERDATA S.L.*

**Abstract.**

Computational methods have gone through enormous improvements during the recent years. In this context, different enhancements in chemoinformatic and anticancer drug predictions will be listed as follow.

Computational methods have gone through enormous improvements during the recent years. They have been used as very helpful tools in many research fields including healthcare. In fact, the accessibility to healthcare data has become easier and it incited researchers to combine computational methods and healthcare results. Among various computational methods, machine-learning (ML) and deep-learning (DL) methods provide the most consistent and effectual responses that provide useful information about different diseases and drug discovery. In this context, different enhancements in chemoinformatic and anticancer drug predictions will be listed as follow.

In this context, Davinder Paul Singh and Baijnath Kaushik [1] completed a systematic literature review for the prediction of anticancer drug response using various machine-learning and deep-learning techniques. They used PRISMA guidelines in order to obtain 105 papers that align with the concept of the review. The research papers were collected from Google Scholar, PubMed and Sciencedirect. They found that amidst the availability of various studies, there are certain challenges associated with each ML and DL method. Thus, future researchers can consider these limitations and challenges to develop a prominent anticancer drug response prediction method, and it would be greatly beneficial to the medical professionals in administering non-invasive treatment to the patients.

As a matter of fact, Lantian Yao *et al.* [2] developed a Deep Graphical Representation to accelerate the discovery of anticancer peptides. Cancer is one of the leading diseases around the world and peptide-based therapies have concerned significant attention in the last decade. Thus, developing a precise prediction of anticancer peptides (ACPs) leads to the discovery of novel cancer treatments. In this work, they proposed a novel machine learning framework (GRDF) that incorporated deep graphical representation and deep forest architecture for analyzing ACPs. GRDF extracts graphical features based on the physicochemical properties of peptides and integrates their evolutionary information along with binary profiles for constructing models. They also used deep forest algorithms that enables excellent performance on small datasets but without complicated tuning of hyperparameters. The experiment showed GRDF exhibited state-of-the-art performance on two elaborate datasets, set 1 and set 2, achieving 77.12% and 94.10% accuracy respectively. Moreover set 1 got 77.54% F1-score, whereas set 2 achieved 94.15% F1-score, exceeding existing ACP prediction methods. In addition, GRDF was well-interpretable, enabling researchers to better understand the features of peptide sequences. The promising results demonstrated that GRDF was remarkably effective in identifying ACPs. Therefore, the framework presented in this study could assist researchers in facilitating the discovery of anticancer peptides and contribute to developing novel cancer treatments.

In a similar concept, Tongtong Cui *et al.* [3] developed a gamma distribution based predicting model for breast cancer drug response based on multi-layer feature selection. They aimed to create a model to predict drug response based on data mining, which could actually provide clinical decision support for cancer patients. They planned a three-layer feature selection combined with the following features: a gamma distribution based Generalized Linear Model (GLM) and a two-layer feature selection combined with an Artificial Neural Networks (ANN). The two regression methods were applied to the Encyclopedia of Cancer Cell Lines (CCLE) and the Cancer Drug Sensitivity Genomics (GDSC) datasets. The method was able to reach better accuracy on anticancer drug response prediction contrasted with existing methods, by the usage of ten-fold cross-validation. It took $R^2$ value of 0.87 and RMSE value of 0.53. Through data validation, the significance of assessing the reliability of predictions by predicting

confidence intervals and its role in personalized medicine are illustrated. The correlation analysis of the genes selected from the three layers of features also shows the effectiveness of the proposed methods.

In conclusion, chemoinformatic methods are increasingly gaining importance in the field of healthcare. These various studies have demonstrated how these methods are proving to be valuable in obtaining a better understanding of anticancer drug predictions. This could potentially result in a higher adoption of these methods as tools in the near future, considering their potential to support scientists in making critical decisions.

## References

1. Singh DP, 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*. 2023 Jan; **101** (1): 175-194. doi: 10.1111/cbdd.14164. Epub 2022 Nov 10. PMID: 36303299.
2. Yao L, Li W, Zhang Y, Deng J, Pang Y, Huang Y, Chung CR, Yu J, Chiang YC, Lee TY. Accelerating the Discovery of Anticancer Peptides through Deep Forest Architecture with Deep Graphical Representation. *Int J Mol Sci*. 2023 Feb 21; **24** (5): 4328. doi: 10.3390/ijms24054328. PMID: 36901759; PMCID: PMC10001941.
3. Cui T, Wang Z, Gu H, Qin P, Wang J. Gamma distribution based predicting model for breast cancer drug response based on multi-layer feature selection. *Front Genet*. 2023 Feb 2; **14**: 1095976. doi: 10.3389/fgene.2023.1095976. PMID: 36816042; PMCID: PMC9932661.