Comparative Evaluation of Drug Response Metrics for Predicting Cancer Sensitivity Using Transcriptomic Profiles

Yen-Jung[1], Chiu, Pei-Qi, Wu[1], Yen-Hwa, Huang[2]

[1]Department of Biomedical Engineering, Chang Gung University, Taoyuan, Taiwan
[2]National Yang Ming Chiao Tung University, Taipei, Taiwan

Abstract. Background: Pharmacogenomic modeling aims to predict cancer drug sensitivity from molecular features such as gene expression. However, the choice of drug response metric can critically affect model performance. This study systematically evaluates three commonly used metrics—area under the dose–response curve (AUC), Z-score, and half-maximal inhibitory concentration (IC50)—to determine their relative suitability for machine learning-based prediction using transcriptomic data. Methods: We assembled an integrated dataset comprising 636 cancer cell lines and drug response profiles for 169 compounds, along with RNA-seq–based gene expression features. XGBoost regression models were trained separately using AUC, Z-score, and IC50 as response variables. Model performance was assessed using R^2, Pearson correlation, and mean absolute error. Additionally, gene-level correlation analyses were conducted to evaluate linear associations between gene expression and drug sensitivity. Results: AUC-based models consistently outperformed those based on Z-score and IC50 in terms of predictive accuracy and robustness. The highest-performing drugs under the AUC framework included Nelarabine (R^2 = 0.83), Sorafenib, and Venetoclax—all of which have established clinical relevance. In contrast, gene-wise Pearson correlation analysis revealed that most genes exhibited weak linear relationships with drug sensitivity across all metrics ($|PCC| < 0.1$), suggesting that response prediction depends on complex, multigenic interactions. Conclusion: AUC is a more reliable and informative drug response metric for transcriptome-based prediction of cancer sensitivity. The findings support the application of multivariate machine learning models and emphasize the importance of metric selection in pharmacogenomic modeling pipelines. Keywords— Drug sensitivity prediction, Pharmacogenomics, Gene expression, XGBoost, Cancer cell line

1 INTRODUCTION

Cancer remains one of the most significant public health challenges worldwide, continually threatening human health. According to the latest reports from the World Health Organization and the American Cancer Society, approximately 19.3 million new cancer cases were reported globally in 2020, with nearly 10 million cancer-related deaths [1]. In the United States alone, over 2 million new cancer cases and more than 600,000 deaths are anticipated in 2025 [2].

The complexity and high heterogeneity of cancer make the search for durable and effective treatments a substantial challenge. Recent advancements in targeted therapy and immunotherapy have significantly improved the prognosis for certain cancers. For instance, the five-year relative survival rate for chronic myeloid leukemia has risen from 22% in the 1970s to over 70%, and survival rates for metastatic melanoma have also notably increased

[3]. Targeted drugs, including tyrosine kinase inhibitors, platinum-based chemotherapy agents, and immune checkpoint inhibitors, demonstrate remarkable efficacy by precisely inhibiting cancer-specific oncogenic pathways and gene mutations [1, 4]. Nevertheless, cancer treatment continues to face challenges such as drug resistance and disease progression due to the evolving and adaptive nature of cancer cells, necessitating ongoing research into new drugs and treatment combination strategies [1].

To address the rapidly growing demand for cancer pharmacogenomic data, several large-scale anti-cancer drug screening programs have been established, creating extensive drug sensitivity databases covering thousands of cancer cell lines. Notable examples include the Genomics of Drug Sensitivity in Cancer (GDSC), Cancer Cell Line Encyclopedia (CCLE), National Cancer Institute 60-cell line panel (NCI-60), and the Cancer Therapeutics Response Portal (CTRP) [5]. These databases openly provide genomic and transcriptomic data for various cancer cell lines, along with their responses to numerous anti-cancer drugs.

For example, the CCLE database contains extensive molecular feature data from 1,094 cell lines [6]. The GDSC database is even more comprehensive, including data from over 970 cancer cell lines and sensitivity profiles for 624 drugs, totaling over 500,000 IC50 data points [7, 8]. Additionally, the updated GDSC2 introduced in 2015 employs more advanced screening technologies, enhancing data quality and research applicability [6].

With these large-scale data resources, machine learning (ML) has emerged as a critical tool for analyzing the complex relationships between genomics and drug sensitivity. By integrating high-dimensional gene expression and multi-omics data, ML models effectively predict cancer cell responses to specific drugs and identify potential biomarkers through feature importance analyses[9, 10]. Such models significantly enhance the accuracy and efficiency of drug sensitivity predictions, promoting the development of personalized therapeutic strategies.

Common metrics used for drug sensitivity evaluation include half-maximal inhibitory concentration (IC50), area under the drug response curve (AUC), and standardized Z-score. IC50, a traditional measure, directly reflects the concentration required to inhibit cell growth by half; AUC comprehensively assesses drug potency and persistence; and Z-score standardizes results across experiments, minimizing experimental error and batch effects [1, 11].

Previous research has successfully applied ML models to analyze CCLE and GDSC data, achieving robust drug response prediction results. For instance, Park et al. utilized XGBoost and Ridge Regression models to predict IC50 values for 24 drugs based on CCLE and GDSC gene expression and mutation data, achieving an R^2 greater than 0.3 for certain drugs [9]. Further, Park et al. (2022) integrated multi-omics data to predict ln(IC50) values, demonstrating that the XGBoost model consistently achieved an R^2 ranging from 0.6 to 0.8 in most scenarios, highlighting its practical reliability for drug sensitivity prediction [10].

However, comprehensive and systematic comparative analyses of AUC, IC50, and Z-score as prediction metrics are currently lacking. The primary objective of this study is to thoroughly analyze and compare the predictive performance of AUC, IC50, and Z-score

to provide clearer guidance for future cancer drug sensitivity research and clinical application.

2 Methods and Materials

2.1 Data Sources

The data utilized in this study were obtained from three datasets provided by the Dependency Map (DepMap) portal (https://depmap.org/portal/).

- Drug Sensitivity Data: Extracted from the sanger-dose-response.csv file provided by the Sanger GDSC1 and GDSC2 projects. This dataset contains drug response information for 398 drugs tested across 974 cancer cell lines, derived from 30 different tissue types including lung, liver, and breast. Each cell line has a unique identifier called "ARXSPAN_ID," and the data includes three primary drug response metrics: IC50, AUC, and Z-score.

- RNA-Seq Gene Expression Data: Sourced from the 2018 version of CCLE RNA-Seq data, comprising gene expression information for 1,019 cancer cell lines from 21 different tissue types and covering expression data for over 56,000 genes.

- Cell Line Annotation Data: This file supports cell line identification and data integration, providing corresponding information such as mappings between ARXSPAN_ID and CCLE_ID.

2.2 Data Preprocessing

During preprocessing, we first verified the completeness of drug sensitivity data, ensuring all drugs had IC50, AUC, and Z-score metrics. To reduce experimental design variability, we included only drug sensitivity data from GDSC2 to enhance data consistency and reliability. Furthermore, for RNA-Seq data, we retained only approximately 18,000 protein-coding genes to avoid confounding effects from non-coding RNAs.

2.3 Model

XGBoost, an extreme gradient boosting method, was employed as the primary machine learning approach to predict cancer cell drug responses. XGBoost is based on gradient-boosted decision trees (GBDT) and is well-suited for handling high-dimensional and sparse biomedical data [12]. To optimize model performance, we followed the parameter configuration suggested by Branson et al. (2024), resulting in 36 parameter combinations with eta: [0.05, 0.1, 0.2], max_depth: [3, 5, 7], subsample: [0.7, 0.9], colsample_bytree: [0.6, 0.8].

Models were independently trained with IC50_PUBLISHED, Z_SCORE_PUBLISHED, and AUC_PUBLISHED as target variables. Model accuracy and generalization capabilities were compared to evaluate the suitability of each sensitivity metric.

2.4 Model Evaluation

Multiple regression metrics were employed to comprehensively assess model performance, including Mean Absolute Error (MAE), R-squared (R^2), Pearson correlation coefficient, and Spearman's rank correlation coefficient.

R^2 indicates how well the predicted values match the actual data, ranging from 0 to 1, with values closer to 1 indicating better predictive performance. MAE measures the average magnitude of errors between predicted and actual values; lower values indicate higher prediction accuracy. Pearson correlation coefficient assesses linear correlation, with values ranging from -1 to 1; values near ±1 denote strong linear correlation. Spearman's rho is a non-parametric statistic evaluating monotonic relationships, suitable for analyzing hierarchical or non-linear data relationships.

Additionally, scatter plots of predicted versus actual values were generated to visualize model predictions and actual data relationships.

3 Results

3.1 Studt design

Our study followed a structured approach to analyze and predict cancer drug sensitivity using genomic data (Figure 1). Initially, we compiled comprehensive datasets from the Dependency Map (DepMap), including drug sensitivity responses, RNA-seq gene expression profiles, and cell line annotations. Ultimately, we constructed an integrated dataset comprising 636 cancer cell line samples, responses to 169 drugs, and corresponding gene expression data, which served as the foundation for subsequent analyses. This harmonized dataset enabled uniform downstream modeling and ensured consistency across all evaluation phases. Data exploration involved rigorous assessment and visualization of drug response metrics distribution and correlation analysis among IC50, AUC, and Z-score. Ultimately, we constructed an integrated dataset comprising 636 cancer cell line samples, responses to 169 drugs, and corresponding gene expression data, which served as the foundation for subsequent analyses. This harmonized dataset enabled uniform downstream modeling and ensured consistency across all evaluation phases.

Following preliminary data analysis, we implemented a gene selection step to reduce dimensionality and enhance model relevance. We utilized the XGBoost machine learning algorithm, systematically optimizing model hyperparameters through 169 distinct combinations derived from established practices.

Three independent prediction models were developed, each trained on one of the key drug sensitivity metrics, as IC50, AUC, and Z-score. Comprehensive evaluations were conducted using multiple regression metrics, including MAE, R^2, Pearson correlation coefficient, and Spearman's rank correlation coefficient. The final comparative analysis aimed to determine the most reliable and robust drug sensitivity metric for predictive modeling in cancer research.

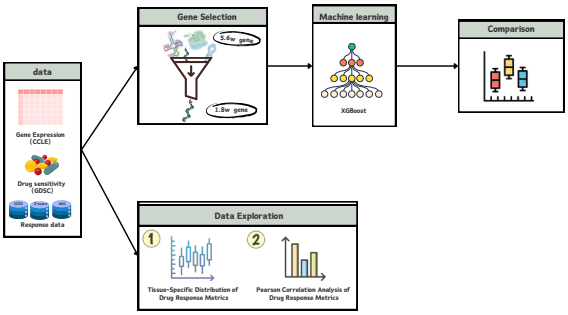


Figure 1 Drug sensitivity model study workflow

3.2 Haematopoietic and lymphoid tissues and Nelarabine

To investigate the variation in Nelarabine sensitivity across cancer cell lines of different tissue origins, we analyzed the distribution of IC50, Z-score, and AUC values.

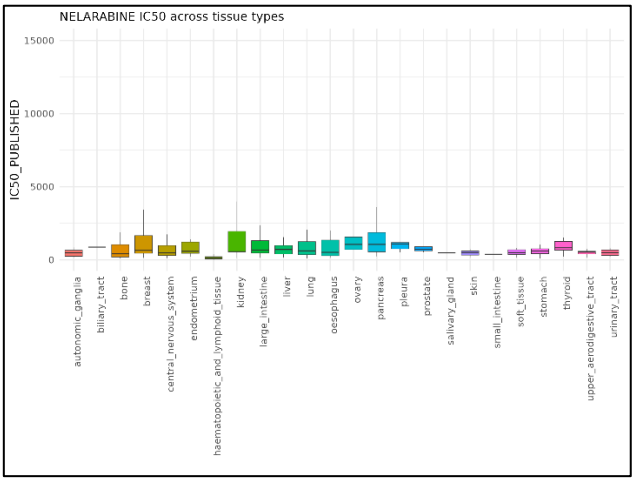
In terms of IC50, cell lines derived from haematopoietic and lymphoid tissues exhibited the lowest values among all tissue types, with the majority falling below 2,500 nM, indicating that lower drug concentrations were sufficient to achieve 50% inhibition (Figure 2A). In contrast, IC50 values for most other tissue types were substantially higher and more widely distributed.

The Z-score distribution further supported this observation, with haematopoietic and lymphoid cell lines showing median Z-scores below –1. This negative shift indicates above-average sensitivity in these tissues, whereas other tissue types such as central nervous system, breast, and pancreas demonstrated Z-scores near or above zero, suggesting reduced responsiveness to Nelarabine (Figure 2B).

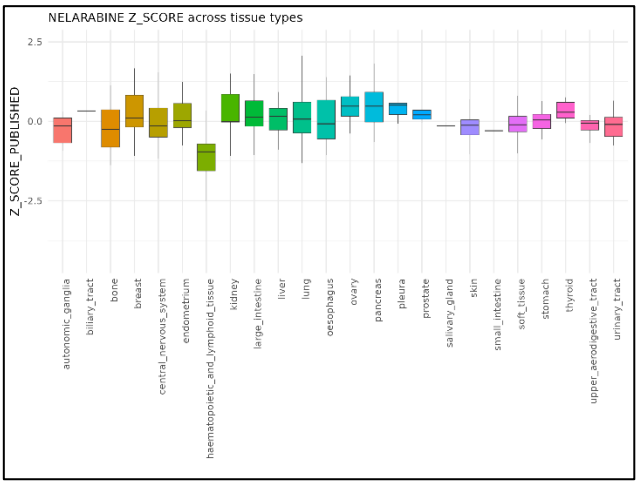
Similarly, the AUC distribution revealed that haematopoietic and lymphoid tissues had the lowest AUC values, typically below 0.8. Since AUC reflects the area under the drug-response curve, a lower AUC corresponds to a stronger and more sustained inhibitory effect. Other tissues showed AUC values clustering near 0.9–1.0, indicating comparatively weaker responses (Figure 2C).

Together, these three complementary metrics consistently demonstrated that Nelarabine exhibits enhanced cytotoxic activity in cell lines of haematopoietic and lymphoid origin, consistent with its known clinical efficacy in hematologic malignancies.

(A)


NELARABINE IC50 across tissue types
IC50_PUBLISHED
15000
10000
5000
0
autonomic_ganglia
biliary_tract
bone
breast
central_nervous_system
endometrium
haematopoietic_and_lymphoid_tissue
kidney
large_intestine
liver
lung
oesophagus
ovary
pancreas
pleura
prostate
salivary_gland
skin
small_intestine
soft_tissue
stomach
thyroid
upper_aerodigestive_tract
urinary_tract

(B)


NELARABINE Z_SCORE across tissue types
Z_SCORE_PUBLISHED
2.5
0.0
-2.5
autonomic_ganglia
biliary_tract
bone
breast
central_nervous_system
endometrium
haematopoietic_and_lymphoid_tissue
kidney
large_intestine
liver
lung
oesophagus
ovary
pancreas
pleura
prostate
salivary_gland
skin
small_intestine
soft_tissue
stomach
thyroid
upper_aerodigestive_tract
urinary_tract

(C)

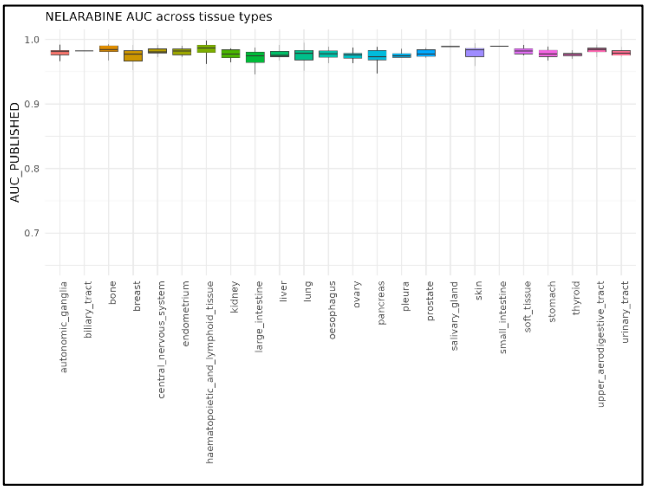

NELARABINE AUC across tissue types
AUC_PUBLISHED
1.0
0.9
0.8
0.7
autonomic_ganglia
biliary_tract
bone
breast
central_nervous_system
endometrium
haematopoietic_and_lymphoid_tissue
kidney
large_intestine
liver
lung
oesophagus
ovary
pancreas
pleura
prostate
salivary_gland
skin
small_intestine
soft_tissue
stomach
thyroid
upper_aerodigestive_tract
urinary_tract

Figure 2 (A) Distribution of IC50 values for Nelarabine across cancer cell lines grouped by tissue type. (B) Distribution of Z-scores for Nelarabine across different tissue types. (C) Distribution of AUC values for Nelarabine across cancer cell lines.

3.3 Transcriptomic Correlates of Drug Sensitivity

To evaluate the extent to which individual gene expression levels are linearly associated with drug sensitivity, we calculated Pearson correlation coefficients between the expression of each gene and three response metrics—AUC, Z-score, and IC50—for three representative drugs: Nelarabine, Sorafenib, and Venetoclax.

Across all drug–metric combinations, the vast majority of genes exhibited PCC values close to zero, with over 70–80% of genes falling within the (–0.1, 0.1) interval. This trend was consistent for all three drugs and for each response metric, including AUC (Figure 3A, 3D, 3G), Z-score (Figure 3B, 3E, 3H), and IC50 (Figure 3C, 3F, 3I). Only a small proportion of genes showed moderate correlation ($|PCC| > 0.2$), and strong correlations ($|PCC| > 0.5$) were extremely rare or absent.

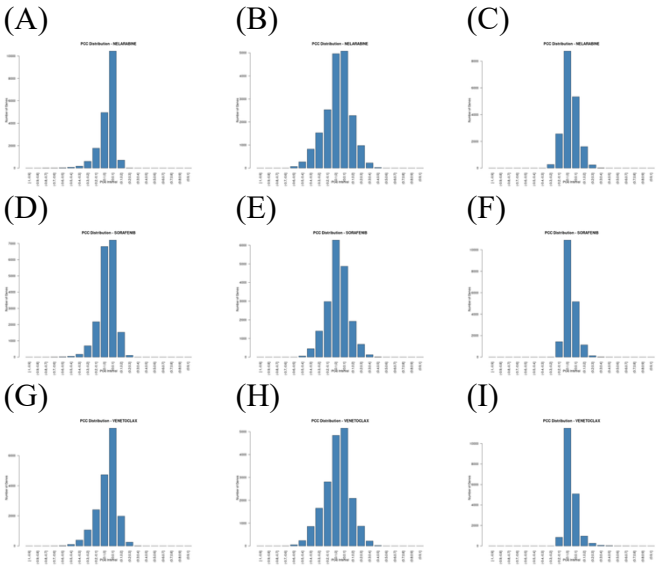


Figure 3 Distribution of gene-wise Pearson correlation coefficients between gene expression and drug response across three. (A–C) Nelarabine, (D–F) Sorafenib, and (G–I) Venetoclax. Each panel shows the number of genes falling within specific PCC intervals for one of the response metrics: AUC (A, D, G), Z-score (B, E, H), and IC50 (C, F, I).

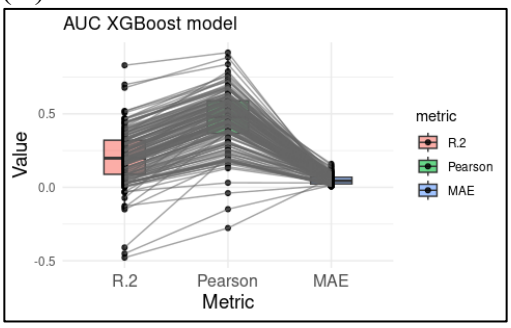
3.4 Evaluation Results for Multiple-Choice Questions

To assess the predictability of different drug sensitivity metrics from gene expression data, we trained XGBoost regression models for each drug using IC50, Z-score, and AUC as separate target variables. Each model was evaluated using three performance metrics: R^2, Pearson correlation coefficient, and MAE.

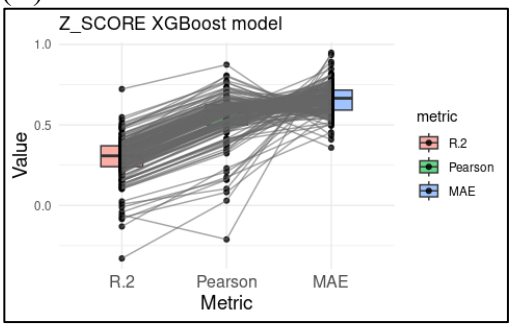
The results indicate substantial variation in model performance across the three target metrics. Models trained with AUC as the target variable generally yielded the highest R^2 values and Pearson correlations, suggesting that AUC is more consistently predictable from gene expression profiles (Figure 4A). In contrast, models using Z-score (Figure 4B) and especially IC50 (Figure 4C) as targets showed lower overall R^2 values and higher MAEs, reflecting reduced predictive accuracy.

Among individual drugs, the AUC-based model for Nelarabine achieved strong performance ($R^2 = 0.83$, PCC = 0.92, MAE = 0.0102), indicating a high degree of association between transcriptomic features and drug response. Other drugs such as Sorafenib and Venetoclax also showed relatively high predictability ($R^2 = 0.70$ and 0.68, respectively). However, for some compounds, all three response metrics yielded poor model fit ($R^2 < 0.2$), suggesting the presence of non-transcriptional determinants influencing drug sensitivity (Table 1).

(A)



(B)



(C)

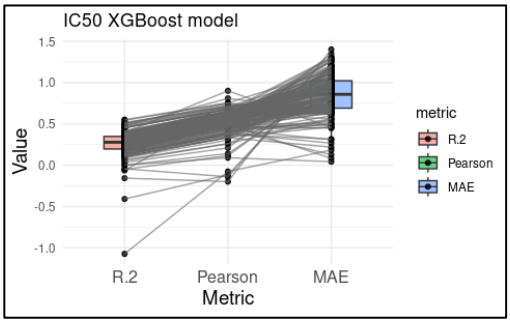


Figure 4 Predictive performance of XGBoost models using different drug response metrics. (A) AUC (B) Z-score (C) IC50.

Table 1 Performance metrics of XGBoost regression models trained with AUC (Top 3 drugs).

DRUG_NAME	R^2	MAE	Pearson	Spearman
NELARABINE	0.8304	0.0102	0.9168	0.3052
SORAFENIB	0.6988	0.04	0.8376	0.3398
VENETOCLAX	0.6774	0.0396	0.8846	0.3305

4 Discussion

Among the three drug response metrics evaluated in this study—AUC, Z-score, and IC50—AUC emerged as the most stable and accurate indicator for modeling transcriptome-based drug sensitivity. XGBoost regression models trained with AUC as the target variable consistently yielded higher R^2 values compared to those trained with Z-score or IC50. Notably, the top three drugs exhibiting the highest AUC-based predictability were Nelarabine, Sorafenib, and Venetoclax. These compounds are widely recognized in clinical oncology: Nelarabine is approved for treating T-cell acute lymphoblastic leukemia (T-ALL) and T-cell lymphoblastic lymphoma (T-LBL) [13]; Sorafenib serves as a standard therapy for hepatocellular carcinoma (HCC) and renal cell carcinoma (RCC) [14]; and Venetoclax is commonly prescribed for chronic lymphocytic leukemia (CLL) patients with 17p deletion [15]. The strong predictive performance for these drugs suggests that transcriptome-level models may be particularly effective for clinically validated therapies.

Despite these promising findings, gene-level correlation analyses revealed a general lack of strong linear associations between individual gene expression levels and drug response across all three metrics. In nearly all cases, the majority of genes exhibited Pearson correlation coefficients near zero ($|PCC| < 0.1$), indicating that no single gene dominantly explains response variation. This result highlights the polygenic and multivariate nature of drug sensitivity, where predictive performance likely arises from the aggregated effect of multiple weakly correlated features. Consequently, univariate analyses may be insufficient to capture relevant biological signals, and multivariate modeling approaches such as gradient-boosted decision trees are better suited for this task.

Additionally, the superiority of AUC as a response metric can be attributed to its ability to integrate response information across the full concentration–response curve, unlike IC50 and Z-score, which reflect single-point estimates or normalized sensitivity scores. Therefore, AUC may provide a more comprehensive and robust signal for modeling gene–drug associations.

5 Conclusion

This study demonstrates that among widely used pharmacogenomic response metrics, AUC offers superior predictive power and stability in transcriptomic-based machine learning models. The predictive success of Nelarabine, Sorafenib, and Venetoclax, all of which are clinically relevant, underscores the translational potential of such modeling frameworks. However, the generally weak linear correlation between individual genes and drug response suggests that drug sensitivity is likely driven by complex, multigenic expression patterns. Our findings support the use of multivariate, non-linear models and advocate for the selection of response metrics—such as AUC—that better reflect the underlying biology of drug response in cancer.

Acknowledgment

We thank the Ministry of Science and Technology, Taiwan (NSTC 113-2221-E-130 -005 -MY3) for their funding support.

References

Reference

[1] Q. Gao *et al.*, "Opportunities and challenges for co-delivery nanomedicines based on combination of phytochemicals with chemotherapeutic drugs in cancer treatment," *Advanced drug delivery reviews,* vol. 188, p. 114445, 2022.

[2] R. L. Siegel, T. B. Kratzer, A. N. Giaquinto, H. Sung, and A. Jemal, "Cancer statistics, 2025," *Ca,* vol. 75, no. 1, p. 10, 2025.

[3] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024," *CA: a cancer journal for clinicians,* vol. 74, no. 1, pp. 12-49, 2024.

[4] C. Zhang, C. Xu, X. Gao, and Q. Yao, "Platinum-based drugs for cancer therapy and anti-tumor strategies," *Theranostics,* vol. 12, no. 5, p. 2115, 2022.

[5] F. Xia *et al.*, "A cross-study analysis of drug response prediction in cancer cell lines," *Briefings in bioinformatics,* vol. 23, no. 1, p. bbab356, 2022.

[6] T. Das, K. Bhattarai, S. Rajaganapathy, L. Wang, J. R. Cerhan, and N. Zong, "Leveraging multi-source to resolve inconsistency across pharmacogenomic datasets in drug sensitivity prediction," *medRxiv,* 2023.

[7] S. Ha, J. Park, and K. Jo, "Comparative analysis of regression algorithms for drug response prediction using GDSC dataset," *BMC Research Notes,* vol. 18, no. Suppl 1, p. 10, 2025.

[8] G. o. D. S. i. Cancer. "Genomics of Drug Sensitivity in Cancer." https://www.cancerrxgene.org/news (accessed June 14, 2025).

[9] A. Park, Y. Lee, and S. Nam, "A performance evaluation of drug response prediction models for individual drugs," *Scientific Reports,* vol. 13, no. 1, p. 11911, 2023.

[10] A. Park *et al.*, "A comprehensive evaluation of regression-based drug responsiveness prediction models, using cell viability inhibitory concentrations (IC50 values)," *Bioinformatics,* vol. 38, no. 10, pp. 2810-2817, 2022.

[11] R. Arafeh, T. Shibue, J. M. Dempster, W. C. Hahn, and F. Vazquez, "The present and future of the Cancer Dependency Map," *Nature Reviews Cancer,* vol. 25, no. 1, pp. 59-73, 2025.

[12] P. Du, H. Huai, X. Wu, H. Wang, W. Liu, and X. Tang, "Using XGBoost-SHAP for understanding the ecosystem services trade-off effects and driving mechanisms in ecologically fragile areas," *Frontiers in Plant Science,* vol. 16, p. 1552818, 2025.

[13] M. H. Cohen *et al.*, "Approval summary: nelarabine for the treatment of T-cell lymphoblastic leukemia/lymphoma," *Clinical cancer research,* vol. 12, no. 18, pp. 5329-5335, 2006.

[14] !!! INVALID CITATION !!! [13, 14].

[15] S. Stilgenbauer *et al.*, "Venetoclax for patients with chronic lymphocytic leukemia with 17p deletion: results from the full population of a phase II pivotal trial," *Journal of Clinical Oncology,* vol. 36, no. 19, pp. 1973-1980, 2018.