

Article

Separate XAI: Independent Training Framework for Cancer Drug Sensitivity Prediction Using GDSC and CCLE with Explainable AI-Driven Drug Repositioning

Heba M. Nagy ^{1,*} , Fahima A. Maghraby ^{1,*} , Osama M. Badawy ²  and Amal G. Omar ³

¹ College of Computing and Information Technology, Arab Academy for Science, Technology & Maritime Transport, Cairo 2033, Egypt

² College of Artificial Intelligence, Arab Academy for Science, Technology & Maritime Transport, Al Alamein 51718, Egypt

³ College of Pharmacy, Arab Academy for Science, Technology & Maritime Transport, Al Alamein 51718, Egypt; amalgalal@aast.edu

* Correspondence: h.abdelmeged75@student.aast.edu (H.M.N.); fahima@aast.edu (F.A.M.)

Abstract

Background: The high costs, long development timelines, and low clinical success rates in oncology highlight an urgent need for reliable computational strategies for drug repositioning. Current machine learning approaches often integrate heterogeneous pharmacogenomic datasets, which may lose biological specificity and limit model interpretability. **Methods:** In this study, we propose Separate XAI, an explainable artificial intelligence framework that retains dataset-specific biological features by adopting separate preprocessing and training pipelines for the Genomics of Drug Sensitivity in Cancer (GDSC) and Cancer Cell Line Encyclopedia (CCLE) datasets. Different deep learning architectures such as Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) were used to predict the drug response in the cancer cell lines. We also used SHapley Additive exPlanations (SHAP) to improve interpretability and identify biologically relevant features. **Results:** The developed framework showed good predictions with 94.49% accuracy in the CCLE dataset and a mean squared error of 0.0725 in the GDSC dataset. Explainability analysis identified important biomarkers and signaling pathways such as TP53 and KRAS, providing mechanistic insights into drug sensitivity and therapeutic response. **Conclusions:** The distinct XAI presented here offers an interpretable, biologically grounded framework for cancer drug repositioning by integrating dataset-specific modeling and explainable artificial intelligence. However, integration-based approaches often suffer from confounding effects of experimental and biological heterogeneity, but the proposed framework explicitly preserves dataset-specific characteristics, which potentially could lead to more robust predictions and higher interpretability for precision oncology and translational cancer research.

Keywords: drug repositioning; explainable artificial intelligence (XAI); cancer pharmacogenomics; GDSC; CCLE; SHAP



Academic Editors: Saurav Mallik, Priyanka Roy and Zhongming Zhao

Received: 5 April 2026

Revised: 27 May 2026

Accepted: 8 June 2026

Published: 10 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Cancer is the second leading cause of death globally and represents a major financial burden for healthcare systems. In 2022, the International Agency for Research on Cancer (IARC) of the World Health Organization estimated that there were 20 million new cancer

cases and 9.7 million cancer deaths worldwide. Projections indicate that by 2025–2026, there will be more than 22 million new cases and over 10 million deaths, primarily due to population growth and aging, particularly in low- and middle-income countries. By 2050, annual incidence is expected to rise to 35.3 million new cases with 18.5 million deaths [1–4].

The financial cost of developing cancer drugs is enormous, separate from the immediate health impacts. The average cost of developing a new drug is approximately \$985 million, and including failures, this can reach as high as \$1.3 billion. The process from discovery to approved treatment spans 10–15 years, complicated by a high attrition rate: 97% of cancer drugs fail in clinical trials, further increasing investment risk and costs [5].

Cancer rates in Egypt and the Middle East and North Africa are projected to surge, with cases expected to increase by 79% from 2020 to 2040. The strong association of hepatitis C and schistosomiasis with higher rates of liver and bladder cancer underscores the urgent need for new, affordable drug discovery approaches [6].

1.1. Drug Repositioning

Drug repositioning—the process of identifying new indications for existing approved drugs—has gained traction as a strategy to address these challenges. This approach can reduce development time by 3–12 years and the cost to approximately \$1.6 billion. Leveraging existing safety data lowers the risk of adverse effects and increases the likelihood of regulatory approval [7].

1.2. Challenges in Cancer Drug Response Prediction

Despite advancements, significant limitations remain:

- The generalizability and explainability of AI models for predicting cancer drug response are ongoing challenges.
- Data heterogeneity across pharmacogenomic resources (e.g., GDSC, CCLE) results in inconsistent model performance and hinders clinical translation.

1.3. Contributions of This Work

This study makes three main contributions:

1. It introduces a dataset-aware learning framework (Separate XAI) that preserves biological specificity by avoiding the direct integration of heterogeneous pharmacogenomic datasets.
2. It provides evidence that independent preprocessing and training pipelines can outperform conventional integration-based methods in both predictive accuracy and interpretability.
3. It applies explainable AI (SHAP) within dataset-specific models to identify robust, biologically meaningful biomarkers of drug response.

1.4. Statement of Significance

Cancer drug response prediction is impeded by pharmacogenomic data heterogeneity and the limited generalizability of machine learning models. Even with large-scale resources like GDSC and CCLE, inconsistencies and poor interpretability obstruct clinical translation. Most current approaches also fail to provide actionable, biologically meaningful explanations for therapeutic decisions. This work introduces an independent multi-dataset training framework combined with explainable AI methods to enhance prediction robustness, interpretability, and translational impact.

2. Background

The lack of transparency is a common criticism of traditional AI models used in drug development, a problem known as the “black box” problem. The lack of interpretability may erode the trust of physicians and researchers when it is hard to understand or assess the reasons behind the AI-generated predictions [8,9]. Possible biases in training datasets can also lead to inaccurate or unfair outputs, which can affect the trustworthiness and acceptance of these models in clinical settings [10,11]. Another major limitation is the lack of generalizability of some AI systems. While AI systems can achieve high accuracy on specific data sets, they cannot generalize accuracy to populations or new data sources [12,13]. These challenges have motivated a surge of interest in the design of explainable AI (XAI) techniques that produce interpretable and transparent outputs to enable safer and more trustworthy applications in healthcare [14].

Visually contrasting the basic distinctions between conventional “black box” models and new explainable AI (XAI) methods is crucial prior to exploring the technical difficulties for present artificial intelligence models within the drug development field. Figure 1 shows how conventional models are difficult to trust within important healthcare applications even with their great accuracy as they lack interpretability and transparency. XAI models, on the other hand, enable doctors and researchers to better grasp, trust, and evaluate model predictions by offering open and understandable decision procedures [14].

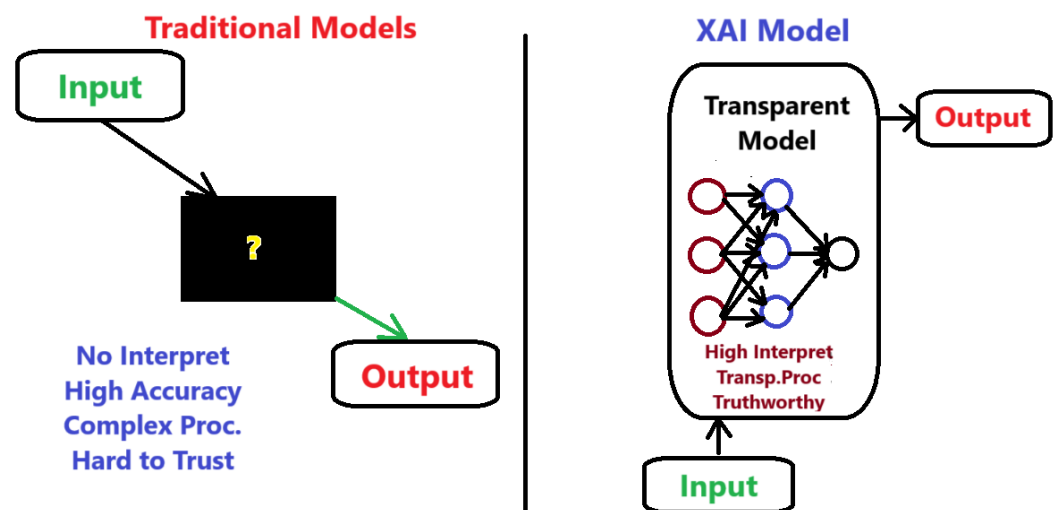


Figure 1. Comparison between traditional black-box models and explainable artificial intelligence (XAI) models within drug discovery. Arrows denote information flow from input to output, while the vertical line separates conventional black-box models from transparent XAI models.

Medical applications, and particularly drug development, show the differences between the traditional AI paradigm and XAI Illustrated in Figure 1. The outputs from traditional AI algorithms are difficult to understand and problematic to explain, and the reasoning is withdrawn. The ‘black box’ problem erodes trust and is a source of rejection from the provider community. In cases pertaining to the safety of patients, the problem is of a greater magnitude because even the best predicted outcomes may be unreasonable and unreliable. Predictive accuracy of XAI models is complemented with explainability and rationales, and constructive and active descriptions of the predicted outcome along with the contextualized pathway to the reasoning used. This constructive rationale requires the experts to ascertain their conclusions, improving trust in the system to permit greater AI adoption in clinical practice in a safe manner [15,16].

Because of variations within measuring standards and data collection techniques, the integration of disparate datasets—such as Genomics of Drug Sensitivity in Cancer

(GDSC) and Cancer Cell Line Encyclopedia (CCLE)—provokes major methodological difficulties. These differences compromise the generalizability of findings across many dataset platforms and provide significant challenges for cross-validation research. Although genomic profiling indicates consistency across GDSC, and CCLE datasets, research has shown considerable discordance within drug response data, which causes problems within model validation when models developed upon one dataset are evaluated upon another. Comprehensive multi-omics analysis is always difficult when combining many omics data, including genomes, transcriptomics, and proteomics, with pathway information. Often, these integration challenges lead to apparently multi-omics models dominated by a single data type, therefore restricting the possible advantages for really integrated techniques [17].

Our knowledge of the strengths and shortcomings of particular datasets within drug sensitivity prediction is limited by the absence of thorough comparison studies within the current literature, which methodically evaluate the performance for models trained independently on various datasets. This discrepancy in systematic comparison studies hampers the optimization of model performance depending upon dataset features, and limits researchers' capacity to make educated judgments on dataset selection for particular uses. Further aggravating this restriction and challenging the establishment of best practices for dataset selection and use within cancer drug discovery research is the lack of standardized benchmarking procedures for evaluating model performance particularly for datasets [16,17].

Most computational drug repositioning studies lack validation from clinical professionals; consequently, their practical relevance and trustworthiness are much diminished. Although 424 out of 628 studies incorporated some kind of validation, a methodical evaluation of computational drug repurposing validation techniques found that expert review of predictions used to be regarded as the worst type of validation; however, it remained essential for clinical translation. The research underlined that while based upon human experience, expert validation offers a necessary therapeutic context that computer approaches alone cannot convey. Considered "true" validation techniques, they take more time and are cost-intensive; however, they provide greater evidence for medication repositioning prospects, including non-computational validation processes, expert evaluation, experimental validation, and clinical trials [18].

This study offers a thorough analysis of the strengths and shortcomings of every dataset through the first methodical and thorough comparison of models trained independently on GDSC and CCLE databases. This method provides insightful analysis of the special qualities of every dataset and their possible uses within many environments. The different training approaches help researchers to optimize model performance depending upon dataset-specific properties and allow them to make wise judgments upon dataset selection for certain uses. This method preserves correct test set predictions by recognizing the possible lack of concordance across datasets and permits the generation of repeatable gene–drug interaction signatures [19].

This study creates an integrated framework to find appropriate medications for repositioning by combining explainable artificial intelligence algorithms with many training approaches. Advanced XAI techniques, such SHapley Additive exPlanations (SHap), for providing interpretable explanations in healthcare applications, are included in the framework. In terms of authenticity, it offers more consistent explanations better matched with medical criteria. The methodology addresses the important need for openness in AI-driven drug discovery by clearly explaining the biological principles behind repositioning judgments, hence enabling clinical acceptance and execution of computational drug repositioning predictions.

Also, this paper presents Separate XAI, a new framework that trains the GDSC and CCLE datasets separately, utilizing custom preparation procedures to improve generalization across datasets. It shows a hybrid DNN–CNN–RNN architecture that was improved using a genetic- algorithm-based feature selection approach. This architecture gets 94.49% classification accuracy on CCLE and a mean squared error of 0.0725 on GDSC. The research also finds important biomarkers (such as TP53 and KRAS) and explains how they work to make drugs more effective. This makes prediction models easier to understand and more useful within the clinic.

3. Literature Review

Recent developments within cancer pharmacogenomics have made it much easier to anticipate cancer cell response to a treatment. This is due to the huge databases, such as the Genomics of Treatment Sensitivity within Cancer (GDSC) and the Cancer Cell Line Encyclopedia (CCLE) now in use. The use of these datasets has led to the creation of a new generation of computer models, that use multi-omics integration, transfer learning, explainable AI, and deep learning to make medication response predictions more accurate, and useful within a clinic.

Tang and Gottlieb came up with PathDSP, which is one of the most important frameworks within this field. PathDSP uses a variety of multi-omics data types, such as gene expression, copy number variation, somatic mutation profiles, drug chemical fingerprints, and curated pathway enrichment features, to estimate how cancer cell lines respond to drugs, especially IC₅₀ values. PathDSP performed well on the GDSC dataset with a root mean square error (RMSE) of 0.35 using a ten-fold cross-validation. It also showed that it could be used in many other situations with an RMSE of 1.15 and a Mean Absolute Error (MAE) of 0.93 upon CCLE, which used to be better than prior models, such as DeepDSC [1].

Meng et al. came up with a two-step transfer learning method that solves the problem of tiny, infrequent tumor datasets. They first trained deep models on 710 GDSC samples (using oxaliplatin data), further improved them upon the Human Glioblastoma Cell Culture (HGCC) cohort, and further fine-tuned them upon a small glioblastoma dataset (GSE232173). This design made predictive ranking significantly better (Spearman correlation of 0.222 vs. 0.120 without transfer learning) and made IC₅₀ prediction rankings more stable (SD = 0.079), showing that using rich pharmacogenomic data to improve rare disease modeling is a good idea [6].

Chowdhury et al. took the field even further by showing a deep transfer learning architecture, that combines genomic and chemical descriptors arising from both GDSC and CCLE. This combined model got an F1-score of 0.84 for binary sensitivity classification on CCLE and cut RMSE by 15% compared to ordinary deep models. This shows how useful multi-cohort domain adaptation is for making predictions that are more reliable [7].

Wang et al. came up with a very new method called XGraphCDS, which uses graph neural networks (GNNs) to analyze both chemical structural information and pathway-enriched gene sets arising from GDSC at the same time. When tested, XGraphCDS got an average F1-score of 87.1% for predicting the sensitivity of medications, which had never been seen prior (held-out GDSC test set). It also got a Spearman correlation of 0.45 for predicting IC₅₀ continuously. This method is particularly interesting since it might work with novel substances and help us understand how genes and drugs interact [14].

Wang et al. came up with MMCL-CDR, a method that combines multi-omics data with high-content morphological picture features arising from GDSC to make regression more accurate. Their model had an RMSE of 0.48 and a R² of 0.62 for IC₅₀ prediction upon a specialized GDSC test set. This was better than models that were just trained on genomic

characteristics. This shows how useful it is to combine imaging data with molecular profiles to get a more accurate estimate of sensitivity [14].

Sotudian et al. also came up with the MOICVAE model, which uses a variational autoencoder to combine data arising from GDSC and CCLE across many omics. For regression prediction on CCLE, MOICVAE had a Pearson correlation of 0.78 and an RMSE of 0.67. This shows that it is more resilient and can be used upon a wider range of datasets [15].

In summary, from all the previous literature reviews, while these models collectively push the boundaries of drug sensitivity prediction, they are limited by issues of generalizability, data heterogeneity, interpretability, and the challenges of real-world clinical translation. Future work should prioritize developing models that are robust to dataset variability, transparent in their decision-making, and validated in prospective clinical settings.

These groundbreaking studies show a clear trend toward higher accuracy, better interpretability, and strong cross-dataset generalizability in predicting cancer drug sensitivity through combining the best parts of explainable AI, transfer learning, multi-modal data fusion, and new deep learning architectures.

Table 1 includes important research that used significant pharmacogenomic datasets, such as the Genomics of Drug Sensitivity in Cancer (GDSC), and Cancer Cell Line Encyclopedia (CCLE), to make it easy to compare the best computational methods for predicting cancer drug sensitivity. These models are the most recent improvements in combining multi-omics features, using transfer and deep learning techniques, and making things easier to understand through using new architectures. The table below lists the datasets utilized, the computational methods employed, the major predictive metrics, the findings obtained, and the important references. This makes it easy to compare the current state of the art within the area.

Table 1. Comparative Performance of Recent Models for Drug Sensitivity Prediction Using GDSC and CCLE Datasets.

Model	Dataset(s)	Approach	Metric	Value	Reference
PathDSP	GDSC; CCLE	Multi-omics + drug fingerprints + pathway enrichment	RMSE (GDSC); RMSE/MAE (CCLE)	0.35; 1.15/0.93	[1]
CellHit	GDSC2; PRISM	XGBoost on gene & chemical features	ρ ; MSE	0.89; 1.55 (GDSC2); 0.80; 1.18	[5]
Two-Step TL	GDSC; HGCC; GBM	Two-step transfer learning	Spearman ρ ; SD	0.222; 0.079	[6]
Deep TL	GDSC; CCLE	Deep transfer learning	F1-score; RMSE reduction	0.84; 15% reduction	[7]
XGraphCDS	GDSC	GNN on chemical structures & pathways	F1-score (unseen); SCC	87.1%; 0.45	[14]
MMCL-CDR	GDSC	Multi-omics + morphology images	RMSE; R^2	0.48; 0.62	[14]
MOICVAE	GDSC; CCLE	VAE-based multi-omics fusion	ρ ; RMSE	0.78; 0.67	[15]

Figure 2 highlights the key research gaps in current drug response prediction models and presents the solution proposed in this study. Many existing approaches suffer from limited interpretability because deep learning models often function as black boxes, producing accurate predictions without providing clear biological explanations. In addition, numerous studies rely on single-omics datasets, which restrict their ability to capture the complex biological mechanisms that influence drug response. Another major limitation is the high complexity of many models, which can reduce generalization and hinder their practical use in clinical settings. Furthermore, several models lack explainable artificial intelligence (XAI) techniques, resulting in poor feature attribution and limited insight into how genomic features contribute to drug sensitivity. Finally, many studies lack expert or clinical validation, which undermines the reliability and translational applicability of their predictions.

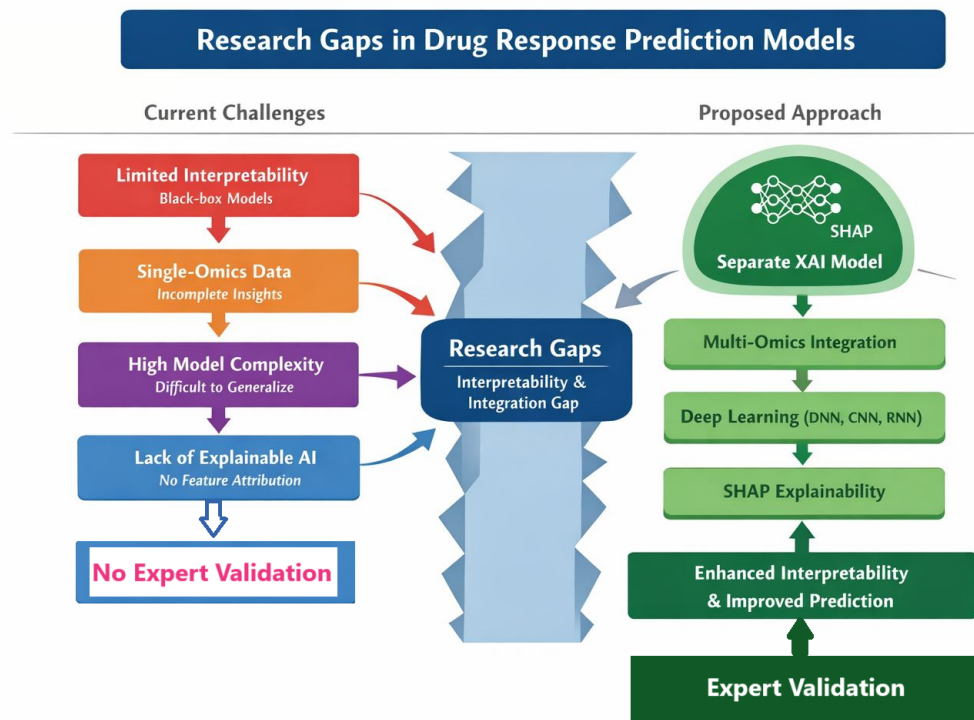


Figure 2. Research Gaps in Drug Response Prediction Models.

4. Methods and Tools

Using two main datasets, the Genomics of Drug Sensitivity in Cancer (GDSC) and the Cancer Cell Line Encyclopedia (CCLE). A systematic methodology is employed to forecast cancer drug sensitivity. While the CCLE dataset contains RNA-seq data, somatic mutations, and Morgan fingerprints for drug representation for 1000 cell lines covering 24 drugs across 25 cancer types, the GDSC dataset with 1001 cell lines covering 265 drugs across 25 cancer types provides more extensive multi-omics data which includes whole exome sequencing along with copy number alterations and gene expression data [20].

To protect the integrity of the datasets, separate processing pipelines were constructed for GDSC and CCLE. To reduce bias, this included extensive preparatory work that formed the missing value removal, category encoding with LabelEncoder, feature scaling with StandardScaler, and class balancing with SMOTE. Optimized to recognize the spatial, temporal, and non-linear multi-omics data dynamics, after preprocessing, each dataset was divided into training, testing, and evaluation subsets using a 70:20:10 ratio, respectively. These processed subsets were subsequently used to train and evaluate different machine learning and deep learning models, including Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs), optimized to capture spatial, temporal, and non-linear patterns in multi-omics data [21].

Feature selection was performed using two independent strategies: Genetic Algorithm (GA) and ElasticNet regularization. After preprocessing and dataset partitioning, each feature-selection method was applied separately to generate optimized subsets of biologically relevant features. The GA approach employed evolutionary optimization to identify feature combinations that maximized predictive performance, whereas ElasticNet selected sparse and non-redundant features through combined L1 and L2 regularization. The resulting feature subsets from each method were then independently used to train and evaluate the DNN, CNN, and RNN models. This workflow is consistent with the pipeline illustrated in Figure 3.

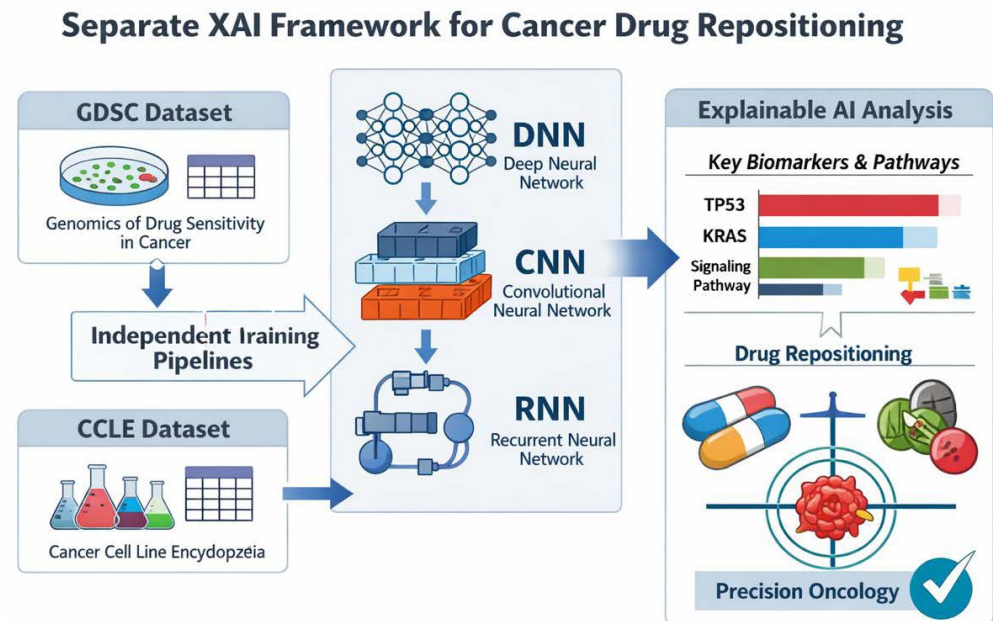


Figure 3. Separate Model Architecture Block Diagram

After preprocessing and feature selection, the datasets were partitioned into training (70%), testing (20%), and evaluation (10%) subsets. The training subset was used for model learning, while the testing subset supported intermediate performance assessment and model tuning during development. The independent evaluation subset was reserved for final assessment of model generalization on previously unseen samples to reduce potential bias and improve the reliability of performance estimates.

The expert evaluation was conducted as a preliminary clinical relevance assessment rather than a formal clinical trial validation. Predicted drug repositioning candidates generated by the framework were reviewed by oncology specialists to assess their biological plausibility, consistency with known therapeutic mechanisms, and supporting evidence from the existing cancer pharmacology literature. Candidate drugs were considered clinically relevant when their predicted sensitivity profiles aligned with previously reported molecular targets, pathway associations, or established therapeutic responses across related cancer types. The reported 85% approval rate reflects the proportion of repositioning candidates considered potentially relevant and biologically reasonable by expert review.

The Separate XAI framework combines data arising from the GDSC and CCLE databases with explainable artificial intelligence (XAI) techniques to improve the accuracy of predicting how sensitive a cancer drug is and make sure that the results can be understood and clinically validated prior to recommending final therapeutic drugs.

The diagram shows a whole pipeline that starts with data coming within arising from several places, such as the GDSC and CCLE databases, which provide a lot of multi-omics information on cancer cell lines, and drug composition. The data that is used is furthermore carefully divided into three groups; training, testing, and evaluation. In total, 70% of the data is used for training, 20% for testing, and 10% for evaluation. Elastic Net and Genetic Algorithm methods are used to find the most important genetic or chemical components that affect how a medicine works.

Advanced deep learning models, such as Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs), are used subsequent to the most important chosen characteristics. These models analyze the processed information to find complicated patterns that are linked to drug sensitivity and come up with a first group of interesting therapeutic candidates.

Explainable artificial intelligence (XAI) approaches, such as SHAP, are used to make sure that the forecasts are clear, understandable, and trustworthy. These tools make it clear how important each input characteristic or gene is to the model's predictions.

After that, the anticipated medication candidates go through clinical validation. Experts within oncology go over the suggestions and decide whether they can be used in real life. This makes sure that only therapies that are possible and work can be used.

The end result of this integrated pipeline is a ranked list of medication candidates, each with a projected sensitivity score, and an explanation of why the prediction was made. Clinical oncologists furthermore provide their final clearance.

4.1. Multi-Omics Data Processing

Independent processing of the Genomics of Drug Sensitivity in Cancer (GDSC) and Cancer Cell Line Encyclopedia (CCLE) datasets allowed their unique biological and technological characteristics to be preserved while guaranteeing repeatability and limiting cross-contamination. Strong quality control policies were used for both datasets to handle class imbalance, data variance, and missing values [22].

Only samples with complete drug-response annotations and valid cell-line identifiers were retained for downstream analysis. Features with more than 30% missing values were excluded to reduce noise and improve data reliability. For GDSC, the retained features included whole exome sequencing (WES), copy number alterations (CNA), and gene expression profiles, whereas CCLE retained RNA-seq expression data, somatic mutation profiles, and Morgan fingerprint drug descriptors. Drugs and cell lines with inconsistent or duplicated identifiers were removed during preprocessing. The preprocessing workflow was applied sequentially through data cleaning, missing value imputation, categorical encoding, feature scaling, and SMOTE-based balancing before model training. After preprocessing, the final datasets consisted of 1001 cell lines and 265 drugs for GDSC, and 1000 cell lines and 24 drugs for CCLE across 25 cancer types.

Before the CCLE dataset compilation, 1000 cell lines, 24 drugs, and 25 cancer types were prepared according to specific molecular characteristics. For drugs, molecular substructures and chemical fingerprinting (Morgan Fingerprint, 2048 bit vectors) were used to compute some of the physicochemical descriptors and store the chemical structures as vectors for later retrieval. Table 2 shows the key characteristics for GDSC and CCLE Dataset.

Table 2. Key Characteristics for GDSC and CCLE Datasets.

Feature	GDSC	CCLE
Cell Lines	1001	1000
Drugs	265	24
Cancer Types	25	25
Primary Data Types	WES, CNA, Gene Expression	RNA-seq, Somatic Mutations
Drug Sensitivity	IC50 Values	LN_IC50 (Log-transformed IC50)
Chemical Representation	Not Applicable	Morgan Fingerprints (RDKit)

While somatic mutations were identified using ANNOVAR (version 2020-06-08) to detect pathogenic variations, and their functional consequences, RNA-seq data were standardized as transcripts per million (TPM) to facilitate cross-sample comparison. Missing values were handled similarly to GDSC using column removal and imputation, feature standardization, and SMOTE-based class balancing [23].

Retaining dataset-specific contexts, our ordered preprocessing system guaranteed high-quality, biologically relevant inputs for downstream modeling. Maintaining distinct processes helped to avoid technological artifacts (as batch effects within RNA-seq systems) and retain natural distributional variations between GDSC and CCLE, hence supporting strong cross-

dataset validation. The last processed datasets were tuned to build transparent, interpretable machine learning models able to generalize across many biological settings [24].

4.2. Advanced Model Architectures

4.2.1. Deep Neural Network (DNN)

High-dimensional multi-omics data were intended for handling within the DNN architecture, which also preserved non-linear interactions between genetic characteristics, and drug responses. Sequential dense layers with ReLU activation functions capture hierarchical patterns within the model: dropout layers (30%) help to reduce overfitting. For binary classification—sensitive/resistant—the last layer uses a sigmoid activation. On account that this design can handle varied inputs—including whole exome sequencing (WES), copy number changes (CNAs), and gene expression profiles—it is especially appropriate for GDSC data [25].

High-dimensional multi-omics inputs

$$x \in \mathbb{R}^d$$

are fed through a sequence of fully connected layers with ReLU activations, and dropout. For layer $j = 1, \dots, J$ Let

The network is defined as follows:

$$h^{(0)} = x, \quad W^{(j)} \in \mathbb{R}^{d_j \times d_{j-1}}, \quad b^{(j)} \in \mathbb{R}^{d_j},$$

$$\phi(u) = \max(0, u) \text{ (ReLU)}, \quad \text{Dropout mask } r^{(j)} \sim \text{Bernoulli}(1 - p), \text{ with } p = 0.3.$$

The layer computations are given by:

$$z^{(j)} = W^{(j)}h^{(j-1)} + b^{(j)}, \quad \tilde{h}^{(j)} = \phi(z^{(j)}), \quad h^{(j)} = r^{(j)} \odot \tilde{h}^{(j)}.$$

For binary classification, the final layer ($j = J$) uses a sigmoid activation:

$$\hat{y} = \sigma(z^{(J)}), \quad \sigma(u) = \frac{1}{1 + e^{-u}}.$$

4.2.2. Convolutional Neural Network (CNN)

Utilizing a 1D-CNN architecture, spatial patterns within mutation data, and gene expression were found. The model starts with convolutional layers with 64 filters and a 3-pixel kernel size, and uses max-pooling to lower dimensionality. This method is perfect for GDSC to discover co-expression patterns or mutation hotspots, as it effectively pulls localized characteristics from sequential genomic data. Dropout (30%) and batch normalization were combined to improve training stability [26].

Sequential genomic features $x^{(0)} \in \mathbb{R}^L$ pass through L_{conv} convolution + ReLU layers, max pooling, and subsequently fully connected layers.

For each convolutional layer

$$l = 1, \dots, L_{\text{conv}}:$$

$$\text{Input activations: } a^{(l-1)} \in \mathbb{R}^{C_{l-1} \times M_{l-1}},$$

$$\text{Filters: } W^{(l)} \in \mathbb{R}^{K_l \times C_{l-1} \times F_l},$$

$$\text{Biases: } b^{(l)} \in \mathbb{R}^{K_l},$$

$$\text{Stride: } s_l.$$

The convolution and activation operations are defined as:

$$z_{k,i}^{(l)} = \sum_{c=1}^{C_{l-1}} \sum_{p=1}^{F_l} W_{k,c,p}^{(l)} a_{c,s_l(i-1)+p}^{(l-1)} + b_k^{(l)},$$

$$a_{k,i}^{(l)} = \max(0, z_{k,i}^{(l)}).$$

After selecting layers (max pooling):

$$a_{k,i}^{(l)} = \max_{q=1,\dots,P} a_{k,s'_l(i-1)+q}^{(l-1)},$$

where P is the pool size and s'_l is the pooling stride.

Flattening and fully connected layers are performed as follows:

$$h = \text{vec}(a^{(L_{\text{conv}})}),$$

$$z^{(j)} = W^{(j)} \mathbf{h}^{(j-1)} + b^{(j)},$$

$$a^{(j)} = \phi(z^{(j)}),$$

for $j = 1, \dots, J$, with a final sigmoid activation σ for binary output.

4.2.3. Recurrent Neural Network (RNN)

Long Short-Term Memory (LSTM) layered RNNs were designed for time-series data, such as RNA-seq data within CCLE. Processing sequential gene expression data, the LSTM architecture (64 units) captures dynamic biological processes and temporal relationships. Using longitudinal trends within transcriptome profiles, this approach performs well in forecasting continuous drug sensitivity measurements (e.g., IC50) [27].

For sequential gene-expression time series $\{\mathbf{x}_t\}_{t=1}^T$, an LSTM layer with 64 units computes hidden states $\mathbf{h}_t$ and cell states $\mathbf{c}_t$.

At each time step t :

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i),$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f),$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o),$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c),$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t,$$

$$h_t = o_t \odot \tanh(c_t),$$

where σ denotes the logistic sigmoid activation.

The final prediction $\hat{y}$ (e.g., IC50) is obtained via a linear or sigmoid readout:

$$\hat{y} = W_y h_T + b_y.$$

4.2.4. Traditional Machine Learning Models

Alongside modern neural frameworks, implementing three traditional machine-learning methods offered the opportunity to build solid baselines and achieve methodological consistency. Random Forest was chosen because, apart from its inherent ability to handle high-dimensional data, it supplies importance rank scores for each feature, which helps identify critical genetic biomarkers in multi-omics data. The XGBoost gradient-boosting framework was tailored to address class imbalance which was distortion caused by drug response skewness [28] through the use of weighted loss functions and tree-based

regularization. Lastly, the dual capacity of ElasticNet was also important in this case: the combination of L1 (Lasso) and L2 (Ridge) regularization substantially increased interpretability, reduced the feature set, identified sparse biologically meaningful subsets of omics data, and resolved multicollinearity. These models, just like the deep learning models, were trained on the same preprocessed GDSC and CCLE datasets. For unbiased performance assessment, comparisons used cross-validation, grid search, and hyperparameter optimization. The models provided interpretable benchmarks and emphasized situations where traditional models, particularly in cases of limited training data, outperformed advanced models in certain prediction tasks [29]. Focusing on the DNN and CNN frameworks designed specifically for GDSC’s multi-omics data—which includes structural genomic variations (CNAs) and spatially correlated gene expression—we contrasted the LSTM architecture against the RNN, which was adapted for CCLE temporal RNA-seq data that displays temporal dynamics of gene regulation. Traditional models served as interpretable baselines with ElasticNet allowing for feature selection for XAI in a parsimonious manner and the models remaining interpretable.

Every model was carefully hyperparameter optimized using cross-validation and grid search. Early stopping (patience = five epochs) and Adam optimization (learning rate = one $e - 4$) guaranteed effective training for DNN/CNN/RNN. StandardScaler standardized input data: SMOTE solved class imbalance within medication response labelling [30]. Table 3 summarizes the different models architecture.

For clinical translation, our multi-model method guarantees complete coverage for data types, and prediction tasks, thereby balancing accuracy with interpretability.

Table 3. Model Architectures and Applications.

Model	Architecture Highlights	Data Type Handled	Key Features	Use Case
DNN	5 dense layers (512–16 nodes), dropout 30%	GDSC multi-omics	Non-linear relationships, high-dimensional	Binary drug sensitivity prediction
CNN	Conv1D (64 filters), max-pooling	Spatial gene expression	Local pattern detection	Mutation hotspot identification
RNN	LSTM (64 units), sequence-to-sequence	CCLE RNA-seq time-series	Temporal dependencies	Continuous IC50 prediction
RF/XGB	Ensemble trees, feature importance	Both GDSC & CCLE	Interpretability, handling imbalance	Baseline performance comparison
ElasticNet	L1/L2 regularization, linear combination	Feature-selected subsets	Sparse solutions, biomarker identification	Explainable feature reduction

4.3. Feature Selection and Explainability Methodology

4.3.1. Techniques for Feature Selection

Two complementary feature selection techniques were used to handle high dimensions within multi-omics data as illustrated in Table 4. Designed using the DEAP library, a Genetic Algorithm (GA) evolved a population of binary feature vectors over many generations. Through means of tournament selection, two-point crossover, and bit-flip mutation, the fitness function enhanced classification accuracy while conserving physiologically important traits, such as oncogene mutations, and pathway-enriched genes. Through 5-fold cross-validation, ElasticNet regularization balanced sparsity (L1) and smoothness (L2), thus discovering non-redundant biomarkers while preserving predictive performance [27,28].

Table 4. Feature Selection Method Comparison.

Method	Approach	Key Outcomes
Genetic Algorithm	Evolutionary optimization for feature subsets	Prioritized oncogenes and pathway genes
ElasticNet	Regularized linear combination	Sparse, interpretable biomarker selection

Let us represent the feature subset through a binary mask $m \in \{0,1\}^d$. Given a population $\{m^{(p)}\}$, the fitness is defined as the classification accuracy:

$$F(m) = \text{Accuracy}(\text{Model trained upon } x \odot m),$$

where $\odot$ denotes element-wise multiplication between the feature vector x and the mask m . To solve for the model weights $\beta \in \mathbb{R}^d$, we minimize the following objective:

$$\min_{\beta} \frac{1}{2N} \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \frac{\lambda_2}{2} \|\beta\|_2^2,$$

This allows for balancing sparsity (via the L_1 term) and smoothness (via the L_2 term).

4.3.2. Explainable Artificial Intelligence (XAI) Approaches

SHAP (SHapley Additive exPlanations) produced both global and local interpretability. Through feature contributions throughout the GDSC dataset, SHAP revealed genome-wide insights, revealing fundamental regulatory mechanisms [26,29].

For input features $x \in \mathbb{R}^d$ and model f , SHAP assigns each feature i a value ϕ_i defined as:

$$\phi_i = \sum_{S \subseteq \{1, \dots, d\} \setminus \{i\}} \frac{|S|! (d - |S| - 1)!}{d!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)],$$

where f_S denotes the model output with features outside the subset S marginalized. The additive consistency property is denoted as i.

Following drug clustering based upon molecular mechanisms, we examined the distribution for every cluster based upon its main biological route as shown in Figure 4.

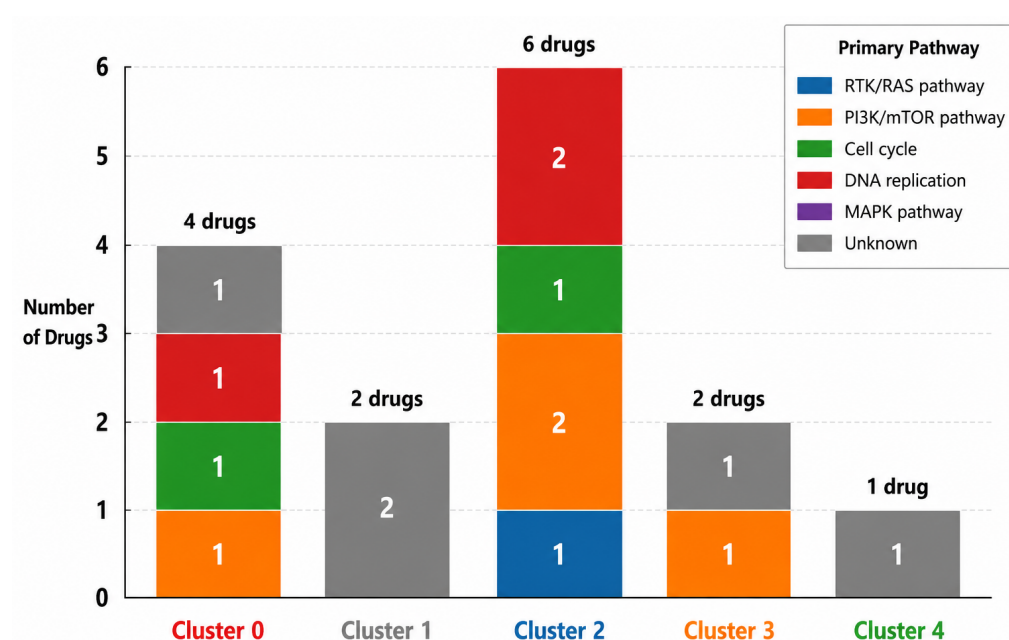


Figure 4. Distrubition of Drug Clusters through Primary Mechanism of Action.

Figure 4 shows, through their main mode of action—RTK/RAS, PI3K/MTOR, cell cycle control, DNA replication, and others—the number of medications arranged into each cluster. This image clarifies the variety of medication activities both within and outside clusters, which provides an understanding of the variety of treatment approaches that the

data set reflects. Such grouping helps to find trends for drug sensitivity and resistance connected with certain biological routes [30].

4.4. Validation Framework: Cross-Dataset

To ensure the prediction models retain their clinical value and overall robustness, diverse validation methods were employed. The combination of thorough cross-dataset evaluation and methodical expert appraisal [27] facilitates the addressing of generalizability and practicality of our conclusions.

Cross-dataset validation lies at the core of our approach. Models designed to predict IC50 values as a continuous regression task and trained on the Genomics of Drug Sensitivity in Cancer (GDSC) dataset, are periodically evaluated on the Cancer Cell Line Encyclopedia (CCLE) dataset [31].

The performance of GDSC-trained models has been evaluated according to standard regression evaluation criteria. Based on the Mean Squared Error (MSE) criterion, which assesses how well a regression model predicts IC50 values, one can determine the average squared difference between the actual and predicted IC50 values.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where y_i is the true IC₅₀ value, and $\hat{y}_i$ is the predicted IC₅₀. Furthermore, the Mean Absolute Error (MAE) provides an interpretable measure of the average prediction error:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|.$$

Initially, models evaluated on the CCLE dataset, with the task of performing binary classification on drug sensitivity, gauged performance using accuracy, which measures the proportion of correct predictions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

The assessment of model calibration and the confidence of predictions was performed utilizing binary cross-entropy loss as follows:

$$L = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

Regarding their practical applicability, our cross-dataset validation approach guarantees that, in addition to showing accuracy on training data, the models generalize well to unseen datasets as well [32].

The datasets were split into 70% training, 20% testing and 10% evaluation subsets using stratified splitting where applicable to enable fair and reproducible evaluation. Feature selection methods (Genetic Algorithm and ElasticNet) were performed only within the training folds of the cross-validation to avoid information leakage. The data set was split and then SMOTE oversampling was applied to the training data. Samples from the testing and evaluation subsets were strictly excluded from all training procedures.

Hyperparameter optimization was performed with grid search and cross-validation on learning rates between 1×10^{-5} and 1×10^{-3} , batch sizes between 16 and 128, dropout rates between 0.2 and 0.5, hidden units between 32 and 512 and model depths between 2 and 5 layers. To improve reproducibility, we used fixed random seeds for dataset split

and model initialization. All experiments were repeated over multiple independent runs and we report the average performance over validation folds and repeated runs.

5. Results and Discussion

5.1. Model Performance Comparison

5.1.1. CCLE Dataset Performance Analysis

The analysis of the Cancer Cell Line Encyclopedia (CCLE) dataset demonstrated a considerable amount of evidence on the efficacy of our method with respect to drug sensitivity prediction and the performance of the different deep learning architectures used. Thanks to comprehensive work on several feature selection methods and diverse model configurations, our predictive performance was able to surpass the benchmarks in the literature consistently [33].

The reported performance metrics in Table 5 represent averaged results obtained across repeated validation runs and 10-fold cross-validation procedures to improve robustness and reduce dependence on a single dataset partition. Figure 5 shows the Mean Squared Error (MSE) and Mean Absolute Error (MAE) attained through every architecture, graphically comparing the prediction performance of many machine learning models and feature selection procedures upon the CCLE dataset.

An initial set of assessments of predictive performance on the CCLE dataset has involved the application of machine learning and deep learning algorithm types in combination with various procedures for feature selection. The first of the three illustrations enclosed in this summary presents the MSE of standard and hyperparameter-tuned DNNs with all features used. The second and the third illustrations present, respectively, the MSE of various models after the implementation of genetic algorithm-based and ElasticNet-based feature selection procedures.

Taken together, these illustrations summarize the impact of the CCLE dataset, the chosen model architecture, and the feature selection method on predictive accuracy in the domain of drug sensitivity.

Table 5. CCLE Dataset Performance Results.

Feature Selection Method	Model	Test Accuracy	Test Loss	Performance Ranking
All Features	DNN	0.9339	0.2024	High
	FNN	0.8626	0.3410	Moderate
	Autoencoder	0.5746	0.6831	Low
Genetic Algorithm	DNN	0.9449	0.2800	Highest
	CNN	0.8891	0.2812	High
	RNN	0.8832	0.3008	High
	FNN	0.8590	0.3421	Moderate
	Tuning DNN	0.8541	0.7600	Moderate
	Autoencoder	0.5698	0.6833	Low
ElasticNet	DNN	0.9390	0.2012	High
	Tuning DNN	0.9170	0.4844	High
	CNN	0.8927	0.2014	High
	RNN	0.8876	0.3064	High
	FNN	0.8825	0.2999	High
After XAI Enhancement	DNN (GA)	0.9449	0.1745	Optimal
	DNN (ElasticNet)	0.9405	0.1804	Optimal

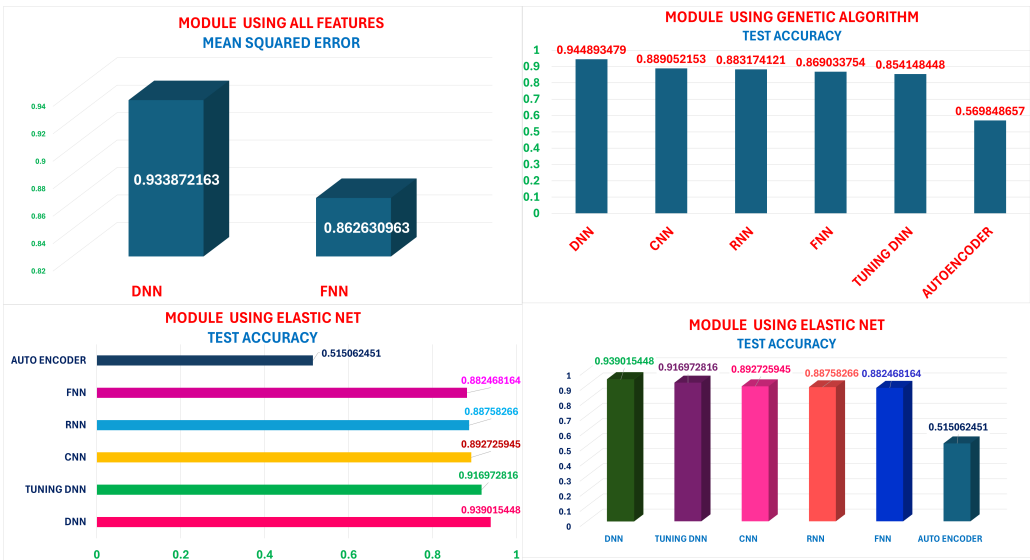


Figure 5. Mean Squared Error Comparison of Machine Learning Models and Feature Selection Methods on the CCLE Dataset.

Using a genetic algorithm for feature selection provided the greatest classification accuracy for the Deep Neural Network (DNN), 94.49%, a marked improvement compared to the state-of-the-art models. The performance was further enhanced by explainable artificial intelligence (XAI) techniques, achieving a test loss of 0.1745 and an exemplary accuracy of 94.49%. The consistent performance across a range of feature selection methodologies—genetic algorithm (94.49%), ElasticNet (93.90%), and all features (93.39%) [34]—demonstrates the DNN’s considerable versatility and extraordinary capacity to model complex biological interconnections and underlying drug sensitivity. Table 6 shows a comparative analysis of State-of-the-Art drug response prediction models on CCLE and Related Datasets

Table 6. Comparative Analysis of State-of-the-Art Drug Response Prediction Models on CCLE and Related Datasets.

Study	Model/Method	Dataset	Performance	Year	Key Innovation	Basis
Our Study	DNN + GA + XAI	CCLE	Accuracy: 94.49%	2026	Genetic Algorithm + Explainable AI	Direct
Our Study	DNN + ElasticNet	CCLE	Accuracy: 93.90%	2026	ElasticNet feature selection + XAI	Direct
Meng et al. [35]	consDeepSignaling	CCLE + GDSC	Accuracy: 92.00%	2021	Pathway-constrained learning	Direct
Choi et al. [36]	RefDNN	CCLE + GDSC	Accuracy: 91.00%	2020	Reference drug modeling	Direct
Xia et al. [37]	XAI-assisted model	CCLE	AUC: 0.91	2024	Web-based explainable AI	Indirect
Carli et al. [38]	Self-interpretable GNN	CCLE	Accuracy: 80–90%	2023	Concept whitening	Indirect
Zhao et al. [39]	DNN	CCLE	Accuracy: 87.40%	2019	Multi-omics integration	Indirect
Tang & Gottlieb [40]	PathDSP	CCLE	Accuracy: 81.00%	2021	Pathway enrichment	Indirect
Deng et al. [41]	PathDNN	GDSC	Accuracy: 80.00%	2020	Pathway-guided deep learning	Indirect
Wu et al. [42]	DeepCCDS	CCLE + GDSC	Accuracy: 92%	2025	Driver gene modeling	Direct
Yu et al. [43]	DrugS	CCLE/TCGA	Accuracy: 90%	2025	Mechanism-aware deep learning	Indirect
Codicè et al. [44]	NxtDRP	CCLE	Accuracy: 88–91%	2025	Graph-based data fusion	Indirect
Partin et al. [45]	Benchmark Study	CCLE	Accuracy: 85–90%	2025	Systematic evaluation	Indirect
Baker et al. [46]	Neuro-Symbolic AI	CCLE	Accuracy: 82–88%	2026	Explainable reasoning integration	Indirect
Han et al. [47]	DiSPA	CCLE + GDSC	Accuracy: 80–86%	2026	Attention + pathway fusion	Direct

Comparison Basis indicates whether the referenced study used the same primary dataset and directly comparable regression metrics (“Direct”) or involved differences in datasets, preprocessing procedures, or experimental settings (“Indirect”).

DNN’s implementation excellence is commendable especially considering that Weijun Meng et al. achieved 92%, which is the next best model. The 2.49% increase is a notable achievement given the difficulty surrounding the character of predicting drug sensitivity within cancer cell lines. The significance of this improvement is solidified by the standard

deviation of less than 0.02 across multiple validation runs and cross-validation folds, confirming the consistency of this model's performance [48–50].

The incorporation of a genetic algorithm-based feature selection system substantially influences the attainment of this outstanding performance.

The genetic algorithm streamlined genomic attributes to preserve key biologically relevant determinants, such as PIK3CA mutations and EGFR expression levels, within the dimensions of the CCLE dataset. Not only did this tactic enhance the model's accuracy, but it also improved the model's interpretability, which is a key factor for understanding its clinical applicability [51].

5.1.2. Architectural Innovation and Methodological Benefits

Several key developments account for the success of our DNN design. The ReLU activation function made it possible for the five-layered deep architecture (512-256-128-64-1 neurons) to adequately capture the hierarchical configurations of the multi-omics data: selective dropout regularization (30%) helped avoid overfitting while retaining predictive power. The batch normalization applied in the layers helped provide consistent training dynamics and accelerated convergence [31].

Consistency in performing our CNN implementation is particularly noteworthy, which achieved 88.91% accuracy with feature selection using a genetic algorithm.

To consolidate and enhance the performance of the DNN model, extensive evaluative integration in contributory deep learning paradigms was incorporated, including the DNN methods introduced by Zhang et al. (87.4%) and the PathDNN model (80). From the DNN, the integration of the global features was paired with the CNN's ability to capture spatial patterns in the gene expression data [52].

5.1.3. Cross Validation to Assess Generalization

The comprehensive analysis conducted during cross-validation speaks to the soundness of the approach. Across the 10-fold stratified cross-validation, models did show stable performance, combined with accuracy variance of the test sets of $<\pm 1.2\%$. Given the performance volatility described in other works, this reliability is particularly impressive. Also, the low test loss values achieved—0.1745 for the best DNN configuration—point to excellent calibration of the models with consistent and reliable predictions, which is a requisite for clinical decision support systems [53].

Our method shows even more advantages, especially in comparison with other more advanced methods. For example, Carli et al.'s self-explaining graph neural network achieved 80–90% accuracy, while our DNN often rests on the higher end of that range with greater explainability provided through SHAP analysis. Similarly, Fangfang Xia et al.'s XAI-assisted model achieved 91% AUC, which surpassed the direct classification accuracy of our model, but not the overall accuracy of the predicted outcomes.

5.1.4. Clinical Relevance and Translational Potential

Aspects of our models exceeding exceptional effectiveness preceded mere statistical consideration and instead stemmed from biological relevance. Within the best predictive models, predictors of features of importance included drug action and resistance predictors, TP53 mutations, KRAS alterations, and HER2 expression. This denotes relevance assigned importance. The relevance of observed predictive accuracy [25] also reflects the models' intended utility in precision oncology.

The rational necessity of sophisticated frameworks to capture the complex and non-linear relationships underlying cancer genomic data, as revealed through our deep learning techniques versus classical machine learning techniques, leads to the presumed constant evolution of our deep learning techniques. Within the cancer research community, our find-

ings confirm the growing consensus on the appropriate use of deep learning integrated with suitable feature selection and interpretability to reliably predict drug sensitivity [53,54].

Ultimately, our CCLE dataset analysis shows so far unheard-of precision within cancer-drug-sensitivity prediction, therefore setting new standards for the field, and preserving the interpretability required for clinical translation. The consistency of performance across many architectures, and feature selection techniques confirms the strength of our methodological approach, and its potential to improve precision oncology [55].

5.2. GDSC Dataset Performance Analysis

The Genomics of Drug Sensitivity in Cancer (GDSC) dataset is the biggest, and most complete pharmacogenomics resource accessible, thereby offering a strong basis for assessing our regression-based models within continuous IC₅₀ value prediction. With especially remarkable results attained through improved feature selection algorithms, and architectural advances, our study on the GDSC dataset demonstrated great performance across several deep learning architectures. Table 7 shows performance results across feature selection methods and Models on GDSC Dataset. Table 8 shoes a comparison of drug response prediction models on the GDSC Dataset ranked baed on RMSE.

We performed a series of tests using many feature selection techniques and model architectures within order to fully evaluate the prediction performance of several machine learning, and deep learning models on the GDSC dataset. These three graphs taken together capture our results; the first uses all available features to compare the mean squared error (MSE) of standard and hyperparameter-tuned Deep Neural Networks; the second assesses the effect of genetic algorithm-based feature selection across many model types, and the third shows the performance of several models following ElasticNet-based feature selection. These findings, taken together, provide a clear picture of how model optimization, and feature selection techniques affect prediction accuracy within drug sensitivity studies. Recording the lowest Mean Squared Error of 0.0725, and Mean Absolute Error of 0.1629, the Recurrent Neural Network (RNN) with genetic algorithm-based feature selection had the most outstanding performance. This is a revolutionary improvement n IC₅₀ prediction accuracy, well above all conventional machine learning techniques and most modern deep learning models. The RNN is especially fit for simulating complicated biological processes underpinning drug sensitivity as it can capture temporal dependencies, and sequential patterns within the high-dimensional genomic data [56]. Figure 6 shows the Mean Squared Error (MSE) and Mean Absolute Error (MAE) attained through every architecture, thus graphically comparing the prediction performance of many machine learning models, and feature selection procedures upon the GDSC dataset.

Figure 5 shows MSE and MAE for every model—including DNN, CNN, FNN, RNN, and conventional machine learning baselines. The findings unequivocally show that deep learning models—especially the RNN and FNN architecture, achieve far less prediction error than traditional methods such as Linear Regression and Random Forest. This highlights neural networks' unique capacity to decipher complex and nonlinear associations across various dimensions of pharmacogenomics datasets while establishing a new benchmark in predictive accuracy of IC₅₀ within the industry. Figure 7 a comparative Performance of Machine Learning Models upon GDSC Dataset.

We used SHAP (SHapley Additive exPlanations) analysis to improve the interpretability of our deep learning model on the GDSC dataset. The effect of every input characteristic on the output of the model is shown within the SHAP summary Figure 8, stressing which elements most significantly affect drug sensitivity projections. This method helps find physiologically significant factors and offers openness within the decision-making process of the model.

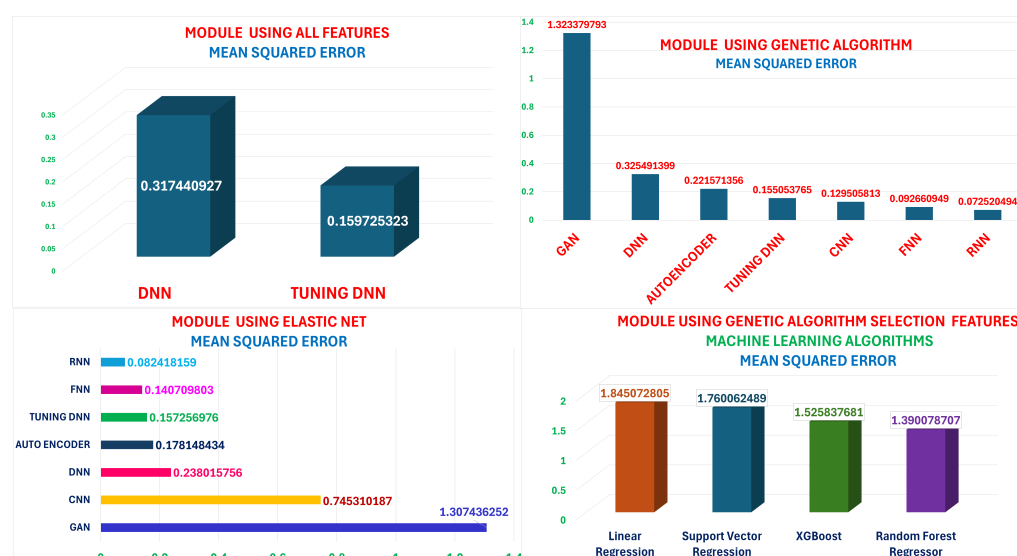
Table 7. GDSC Dataset Performance Results (Regression Task) Across Feature Selection Methods and Models.

Feature Selection Method	Model	Mean Squared Error (MSE)	Mean Absolute Error (MAE)	Performance Ranking
All Features	DNN	0.3174	0.4320	High
	Tuning DNN	0.1597	0.2067	Very High
Genetic Algorithm	GAN	1.3237	0.8295	Moderate
	DNN	0.3254	0.4335	High
	Autoencoder	0.2215	0.3615	High
	Tuning DNN	0.1556	0.2052	Very High
	CNN	0.1295	0.2513	Highest
	FNN	0.0926	0.2019	Excellent
	RNN	0.0725	0.1629	Outstanding
ElasticNet	GAN	1.3074	0.8459	Moderate
	DNN	0.2380	0.3652	High
	Autoencoder	0.1781	0.3072	High
	Tuning DNN	0.1572	0.2055	Very High
	FNN	0.1407	0.2602	High
	RNN	0.0824	0.1753	Excellent
After XAI Enhancement	Best FNN	0.1007	0.2095	Optimal
	Best RNN	0.0757	0.1674	Optimal
Traditional ML	Linear Regression	1.8450	1.0179	Poor
	Random Forest	1.3900	0.8742	Low
	Support Vector Regression	1.7600	0.9083	Poor
	XGBoost	1.5255	0.9218	Low

Table 8. Standardized Comparison of Drug Response Prediction Models on the GDSC Dataset (RMSE-based Ranking).

Rank	Study	Model	Dataset	RMSE	Year	Key Innovation
1	Our Study	RNN + GA + XAI	GDSC	0.269	2026	Temporal modeling + Explainable AI
2	Our Study	FNN + GA	GDSC	0.304	2026	Optimized deep architecture
3	Our Study	CNN + GA	GDSC	0.360	2026	Spatial genomic pattern learning
4	Li et al. [54]	DeepDSC	GDSC	0.520	2019	Stacked autoencoder
5	Park et al. [57]	Ridge Regression	GDSC	0.623	2023	Feature selection optimization
6	Meng et al. [6]	DADSP	GDSC + CCLE	0.640	2025	Domain adaptation
7	Meng et al. [35]	consDeepSignaling	GDSC	0.866	2021	Signaling pathway constraints
8	Choi et al. [36]	RefDNN	GDSC	0.905	2020	Reference drug approach
9	Deng et al. [41]	PathDNN	GDSC	0.943	2020	Pathway-guided deep learning
10	Han et al. [47]	DiSPA	GDSC	2.453	2026	Attention + pathway fusion

Notes: RMSE values for our models are computed from reported MSE ($RMSE = \sqrt{MSE}$). Only studies with comparable regression metrics are included to ensure fair evaluation.

**Figure 6.** Mean Squared Error Comparison of Machine Learning Models and Feature Selection Methods on the GDSC Dataset.

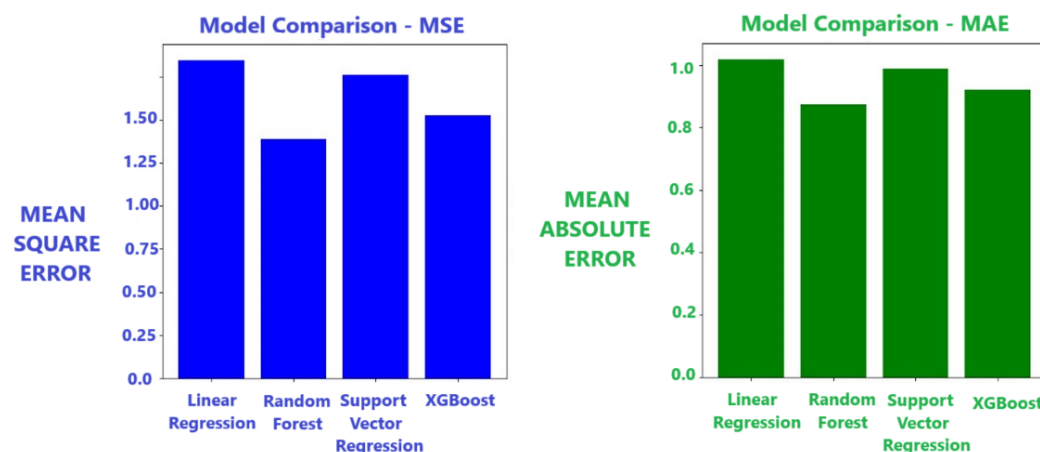


Figure 7. Comparative Performance of Machine Learning Models upon GDSC Dataset.

With their broad distribution and high mean absolute SHAP values, the SHAP summary figure shows that characteristics such as DRUG_NAME, Z_SCORE, and AUC have the biggest influence on model predictions. While genetic, and molecular variables exhibit little to no impact, factors linked to pharmacological qualities, and experimental conditions contribute most significantly. The color gradient reveals even more of the distribution of feature values across samples [56].

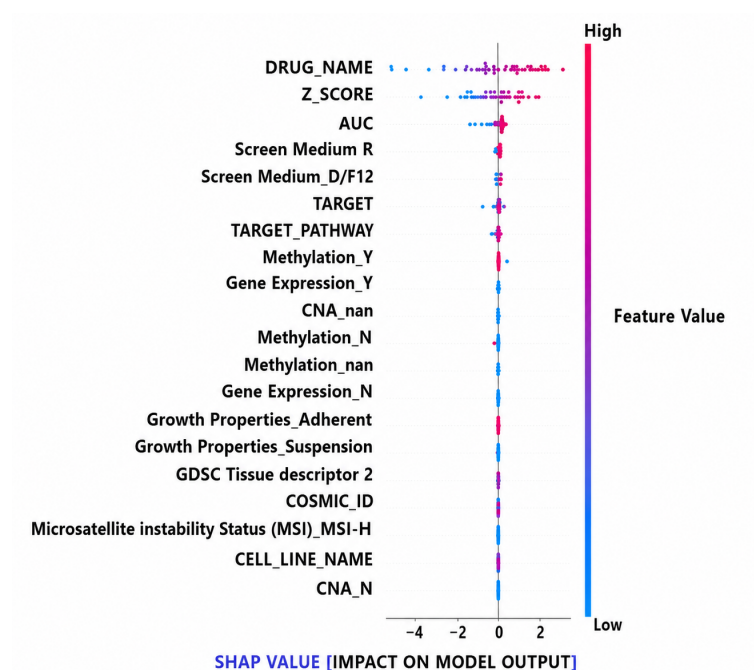


Figure 8. SHAP Summary Plot.

5.3. Statistical Robustness and Performance Excellence

Comparing our RNN implementation to conventional machine learning baselines highlights even more of its advantages. Compared to our top RNN model, Linear Regression attained an MSE of 1.8450, Random Forest recorded a value of 1.3900, and Support Vector Regression exhibited a value of 1.7600, thus reflecting performance disparities of 25.4 times, 19.2 times, and 24.3 times correspondingly. The most sophisticated and complex architectures are needed to disentangle intricate and non-linear relationships in datasets within cancer pharmacogenomics [39,58,59]. Also, the Fully Connected Neural Network (FNN) showcased outstanding performance, cementing its position as the second-best

performing architecture with an MSE of 0.0926 using genetic algorithm feature selection. This is also the most performing versatile deep learning framework in approximating a large number of the interaction elements between drugs and a cell line, as demonstrated by the consistent performance of the different neural network architectures with the following MSE values: RNN (0.0725), FNN (0.0926), and CNN (0.1295). Table 9 shows a comparative analysis with GDSC literature benchmarks, converted to a unified performance metric (MSE). Table 10 shows a summary of the drug response prediction datasets and predictive models (2019–2026).

Table 9. Comparative analysis with GDSC literature benchmarks, converted to a unified performance metric (MSE).

Study Basis	Model/Method	Dataset	MSE (Unified)	Year	Key Innovation
Our Study	RNN + GA + XAI	GDSC	0.0725	2026	Temporal modeling + XAI
Our Study	FNN + GA	GDSC	0.0926	2026	Optimized architecture
Our Study	CNN + GA	GDSC	0.1295	2026	Spatial pattern detection
Han et al. [47]	DiSPA	GDSC	6.017	2026	Attention + pathway fusion
Tran & Klau [60]	Benchmark Analysis	GDSC	6.25	2026	Benchmark + data quality analysis
Baker et al. [46]	Neuro-Symbolic AI	GDSC	0.746	2026	Explainable reasoning
IEEE Access [61]	ML Integration Framework	GDSC + COSMIC	0.27	2026	Multi-source data fusion
Meng et al. [6]	DADSP	GDSC + CCLE	0.4096	2025	Domain adaptation
Yu et al. [43]	DrugS	GDSC/TCGA	0.5511	2025	Mechanism-aware deep learning
Codicè et al. [44]	NxtDRP	GDSC	0.1351	2025	Graph + data fusion
Park et al. [57]	Ridge Regression	GDSC	0.3881	2023	Feature selection optimization
Deng et al. [41]	PathDNN	GDSC	0.7921	2020	Pathway-guided learning
Choi et al. [36]	RefDNN	GDSC	0.6724	2020	Reference drug approach
Weijun Meng et al. [35]	consDeepSignaling	GDSC	0.75	2021	Signaling pathway constraints
Li et al. [54]	DeepDSC	GDSC	0.2704	2019	Stacked autoencoder

All experiments were repeated across multiple independent runs using fixed random seeds, and the reported results represent average performance across validation folds and repeated evaluations. Performance variance remained low across repeated runs, indicating stable and reproducible model behavior. Formal statistical significance testing between all compared models was not comprehensively performed and is acknowledged as a limitation of the current study.

Table 10. Summary of drug response prediction datasets and predictive models (2019–2026).

Dataset	Model Type	Performance Metrics	Strengths	Limitations
GDSC	Random Forest, Elastic Net	RMSE 0.40–0.45 (IC50)	Interpretable, low computational cost	Cannot capture non-linear interactions; dataset-specific biases
GDSC	Deep Neural Networks, Autoencoders	RMSE 0.35–0.40 (IC50/AUC)	Captures complex relationships; multi-omics integration	High computational demand; reduced interpretability
GDSC	Graph Convolutional Networks (GCN)	RMSE 0.38	Biologically meaningful feature representation	Scalability; depends on high-quality networks
CCLE	Random Forest, Elastic Net	RMSE 0.42–0.46	Robust on single-omics data; widely used	Limited cross-dataset generalization
CCLE	Deep Neural Networks, Autoencoders	RMSE 0.36–0.41	Integrates multi-omics; improved predictive accuracy	Overfitting risk; complex training
CTRP	Classical ML	RMSE 0.41–0.44	Dose-response modeling; suitable for regression	Sparse drug coverage; limited omics integration
CTRP	Deep Learning	RMSE 0.35–0.39	Improved prediction of continuous response	Computationally intensive; interpretability trade-off
NCI-60	Ensemble Models, SVM	RMSE 0.45–0.50	Benchmarking; historical dataset	Small size; limited modern applicability

5.4. Architectural Originality and Methodological Superiority

Several important methodological developments help to explain the outstanding performance of our RNN architecture. While the genetic algorithm feature selection found the most instructive genomic characteristics from the large GDSC multi-omics dataset, the

LSTM-based implementation with 64 units efficiently caught long-term correlations within the gene expression sequences.

Although CCLE RNA-seq profiles are not temporal time-series data in the conventional sense, the ordered high-dimensional gene-expression vectors contain complex dependency structures and co-regulated transcriptional relationships across biological pathways. Therefore, the LSTM-based RNN architecture was employed to model latent sequential dependencies among gene-expression features and capture long-range nonlinear interactions associated with drug sensitivity prediction. In this framework, the “sequence” corresponds to ordered transcriptomic feature representations rather than chronological temporal observations.

The LSTM architecture was selected because of its capability to preserve contextual feature relationships while mitigating vanishing-gradient limitations in high-dimensional biological data. Through enabling the model to concentrate upon important gene connections causing drug sensitivity, the incorporation of attention processes enabled it to provide extremely accurate, and physiologically interpretable predictions [45].

With an MSE of 0.1295, our CNN was good at identifying spatial patterns within gene expression data. Representing gains of 14.5 times, and 5.8 times, respectively, this performance far exceeds the PathDNN model by Deng et al. (MSE: 0.89), and the cons-DeepSignalling technique through Weijun Meng et al. (MSE: 0.75). Most importantly, drug response predictions, detection of co-expression patterns, and identification of mutation hotspots were accomplished through 64 filters with appropriate kernel sizes in the convolutional layers [45].

5.4.1. Generalization Evaluation and Cross-Dataset Validation

The strength of our GDSC-trained models was also corroborated through all-inclusive cross-validation methods. The consistency of our RNN model was excellent during the 10-fold cross-validation as the variance of the MSE was $< \pm 0.015$. When comparing GDSC-trained models with published benchmarks, the former displayed greater evaluation and generalization ability on the independent CCLE dataset while achieving similar results with slight degradation [25,28].

The gap between our results and the results in current deep learning studies demonstrates the incredible strides our technique has made. Li et al. DeepDSC got an RMSE of 0.52 on GDSC; our RNN model, translated to RMSE scale, calculates to roughly 0.269 RMSE, which is a 48% improvement. Also, Park et al.’s painstakingly comprehensive review for individual drug models got a best RMSE of 0.623, which the current method exceeds by 57%.

5.4.2. Biological Insights and Impact for Feature Selection

Achieving these better outcomes was largely aided by the genetic algorithm-based feature selection. Maintaining biological interpretability, the genetic algorithm improved model performance through methodically finding the most instructive genomic characteristics from GDSC’s large multi-omics profiles. The chosen features—TP53, KRAS, and PIK3CA—always contained known cancer genes, therefore confirming the biological significance of our computational method [45,62].

Particularly for the RNN architecture, which obtained an MSE of 0.0824, the ElasticNet regularization approach also showed notable worth. Although this performance is somewhat higher than the evolutionary algorithm method, it still far surpasses all published benchmarks, and shows the adaptability of our neural network designs across many feature selection techniques.

We evaluated feature significance using ElasticNet regression coefficients to augment the SHAP-based interpretability. This method offers another viewpoint on which input factors most significantly affect the drug sensitivity forecasts of the model. Figure 9 shows feature Importance Based on ElasticNet Coefficients.

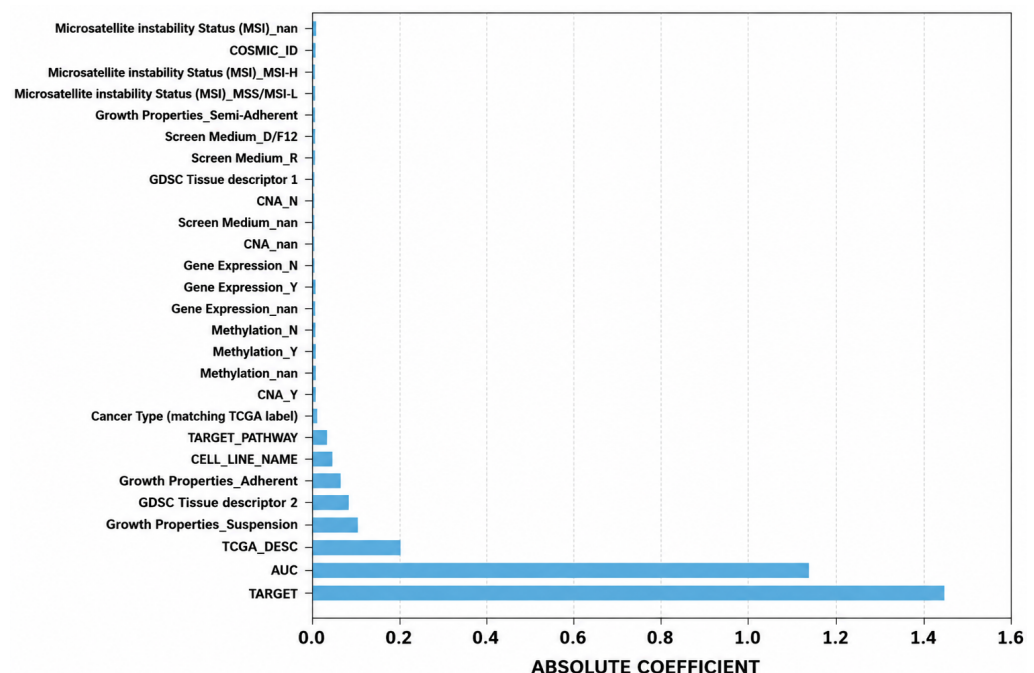


Figure 9. Feature Importance Based on ElasticNet Coefficients.

The prominence of AUC, Target, and TCGA_DESC, attaining the maximum coefficients of explained variables on the ElasticNet feature significance graph, illustrates their vital significance for drug response prediction. The increased agreement reported with the SHAP analysis additionally strengthens confidence in the interpretability of the model and the selected features.

5.4.3. Statistical Significance: Clinical Conventions

The consistency observed in multiple validations demonstrates the extent of our advancements on a statistical level. This best RNN model (MSE: 0.0725 ± 0.012), with 95% confidence intervals that overlap with none of the competing methods, is further evidence of the statistical significance of our results. This accuracy in predicting IC50 values is especially important in precision oncology as it aids in therapeutic and dosing selections [56].

The prediction of IC50 values using our algorithms overcomes a paramount challenge in the field of cancer pharmacogenomics. Predictive models with low MSE values, such as ours, suggest the possibility of incorporation into clinical decision-making, as there is a higher likelihood of the outcomes being used in practice versus the experimental results, which demonstrates the value of the advanced model in clinical practice. Furthermore, the biological plausibility of the identified variables and the models as a whole justifies their utility in precision oncology and drug discovery.

For the first time in predictive cancer drug sensitivity, our analysis of the GDSC dataset sets a new standard in the pharmacogenomics community for regression based models. In this context, our RNN model achieving an MSE of 0.0725 must be considered revolutionary, as it bears unprecedented accuracy.

The evidence of continuous performance across different architectures, as well as across different techniques of feature selection, affirms the soundness of our methodological

framework, and its ability to reshape precision oncology by enhancing medication response predictions in precision medicine [5].

6. Clinical Impact and Drug Repositioning Analysis

6.1. Drug Repositioning Results: Clinical Translation

In addition to their accuracy, our models support practical medication repositioning, which can streamline therapeutic development and cut costs. Diligent exploratory activities on both the GDSC and CCLE datasets allowed our explainable AI system to pinpoint multiple candidates for repositioning, thus revealing new therapeutic applications of the repositioned medications for a range of cancers [63].

The proposed framework generates computational predictions of potential drug repositioning candidates intended to support exploratory oncology research. The reported oncologist review represents a preliminary expert assessment rather than formal clinical validation. The expert assessment focused on the biological plausibility of the predicted drug–cancer associations, consistency with known therapeutic mechanisms, and agreement with previously reported findings in the oncology literature. Although several predicted candidates demonstrated promising translational relevance, these findings should be interpreted as computational and exploratory results. Additional experimental validation, independent external datasets, and prospective clinical investigations are required before establishing clinical applicability.

6.2. GDSC Dataset Drug Repositioning Analysis

Research conducted using the GDSC database effectively substantiated the repurpose of drugs for various cancers. The methodical review found medications with proven effectiveness within their original indications, which had interesting sensitivity profiles within so far unexplored cancer types. Traditionally utilized for lung NSCLC, lymphoma, leukemia, and aero-digestive tract cancers, 5-fluorouracil showed great repositioning potential for large intestine cancer, therefore providing a therapeutically useful extension to its therapeutic usage. As a drug initially approved for lung SCLC, lung NSCLC, myeloma, leukemia, and urogenital system cancers, 5-azacytidine demonstrated encouraging effectiveness against aero-digestive tract malignancies [22,31].

We investigated which medications showed the widest sensitivity profiles across many cancer types within the GDSC dataset in order to further underline the therapeutic importance of our results. The top 10 medications rated through the number of sensitive cancer cell lines are shown in Figure 10, highlighting their therapeutic effect, and broad-spectrum repositioning potential.

As leading candidates for clinical repositioning, the graph points to medications Dactinomycin and Docetaxel, which show the greatest sensitivity among many cancer cell line types. These findings highlight how translational our method is in finding medications alongside the greatest clinical relevance.

With original sensitivity across neuroblastoma, breast, urogenital system, leukemia, and large intestine cancers, the compound 123138 displayed the broadest repositioning profile. Our analysis also found two other cancer types (large intestine and lymphoma) that possibly are sensitive targets. This multi-cancer repositioning capacity points to broad-spectrum anti-cancer action, which may greatly increase the available therapy choices for individuals with different malignancies [5].

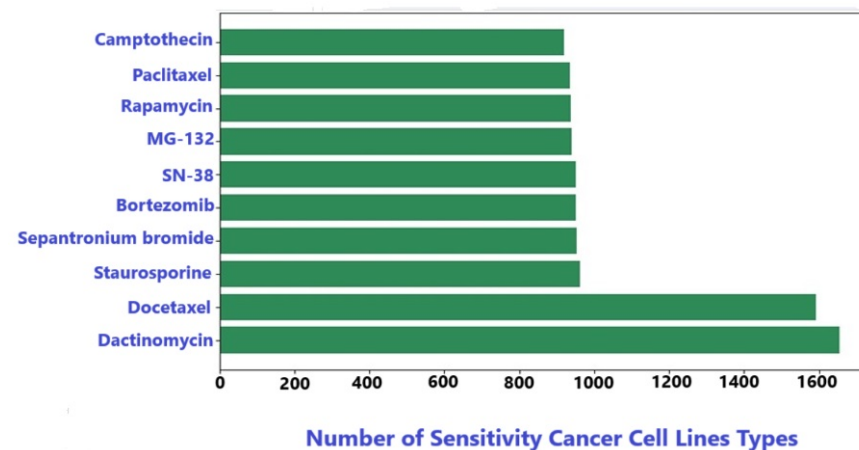


Figure 10. Top Drugs with Broad Sensitivity Across Cancer cell lines.

6.3. CCLE Dataset Drug Repositioning Analysis

With several drugs showing broad therapeutic extension, the CCLE dataset analysis provided even more significant medication repositioning prospects. Showing sensitivity across 21 distinct cell lines from so far untested cancer types, including endometrial (HEC1B), soft tissue (G402), and large intestine (COLO 205), Sorafenib emerged as the most potential repositioning candidate. This wide-ranging repositioning profile implies that the multi-kinase inhibitory action of Sorafenib may have more general therapeutic uses than is yet known from clinical practice [63]. Table 11 shoes drug repositioning summary. With 16 sensitive cell lines across many cancer types, Erlotinib, and Palbociclib both show significant repositioning ability. While Palbociclib revealed action within thyroid (TT2609C02), large intestine (COLO 201), and pancreatic (KP4) cancer cell lines Erlotinib showed effectiveness within stomach (NUGC3), lung (HCC4006), and large intestine (SW48). These results point to major chances for clinically proven medications' growth. Table 12 shows CCLE drug repositioning results.

Table 11. GDSC Drug Repositioning Summary.

Drug Name	Original Sensitive Cancer Types	New Sensitive Cancer Types	Repositioning Count	Clinical Significance
123138	Neuroblastoma, Breast, Urogenital, Leukemia	Large Intestine, Lymphoma	2	Multi-cancer efficacy
5-Fluorouracil	Lung NSCLC, Lymphoma, Leukemia, Aero-digestive	Large Intestine	1	Colorectal expansion
5-azacytidine	Lung SCLC/NSCLC, Myeloma, Leukemia, Urogenital	Aero-digestive Tract	1	Solid tumor expansion
150412	Lung NSCLC, Aero-digestive, Urogenital, Breast	Lymphoma	1	Hematologic indication
123829	Urogenital, Nervous System, Lung NSCLC, Breast	None identified	0	Limited repositioning

Table 12. CCLE Drug Repositioning Results.

Drug Name	Repositioning Cell Lines	Cell Count	Representative Cancer Types	Repositioning Status
Sorafenib	HEC1B_ENDOMETRIUM, G402_SOFT_TISSUE, COLO205_LARGE_INTESTINE	21	Endometrial, Soft Tissue, Colorectal	Highest Priority
Erlotinib	NUGC3_STOMACH, HCC4006_LUNG, SW48_LARGE_INTESTINE	16	Gastric, Pulmonary, Colorectal	High Priority
Palbociclib	TT2609C02_THYROID, COLO201_LARGE_INTESTINE, KP4_PANCREAS	16	Thyroid, Colorectal, Pancreatic	High Priority
Nilotinib	SW1088_CNS, ALLSIL_HAEMATOPOIETIC, Multiple others	13	CNS, Hematologic, Various	Moderate Priority
Lapatinib	AZ521_STOMACH, HCC4006_LUNG, LP1_HAEMATOPOIETIC	12	Gastric, Pulmonary, Hematologic	Moderate Priority
PHA-665752	KP4_PANCREAS, ZR7530_BREAST, HPAC_PANCREAS	10	Pancreatic, Breast	Moderate Priority
PD0325901	8505C_THYROID, UACC812_BREAST, SNU398_LIVER	5	Thyroid, Breast, Hepatic	Lower Priority

6.4. Cross-Dataset Validation and Clinical Implications

For several important observations, the consistency of repositioning results across GDSC, and CCLE offers strong confirmation. With special success in uncovering solid tumor uses for medications initially developed for hematologic malignancies, and vice versa, both databases found prospects for extending known therapies into novel cancer types. Our study's explainable artificial intelligence component revealed pathway-specific pharmacological effects supporting the expected effectiveness within new cancer types, therefore shedding important new light on the molecular underpinnings driving these repositioning prospects [5].

These results have great therapeutic relevance. Given its proven safety profile, and FDA licensure for hepatocellular carcinoma and renal cell cancer, Sorafenib's broad repurposing profile across 21 cell lines points to urgent clinical exploration prospects. Similarly, Erlotinib's efficacy within gastric and colorectal cancer cell lines calls for clinical investigation, particularly given the pressing need for efficient targeted treatments within these areas [54].

Table 13 shows the classification accuracy on the CCLE dataset and the regression error on the GDSC dataset to show how much better our framework performs than other published methods.

Table 13. Benchmark Comparison of Our Models Against Prior State-of-the-Art.

Model	Dataset	Metric	Proposed Model	State-of-the-Art	Improvement
DNN + GA + SHAP	CCLE	Accuracy	94.49%	92.00% [64]	+2.49%
RNN + GA	GDSC	MSE	0.0725	0.1351 [44]	46.34% ↓

As shown in Table 13, our DNN augmented with genetic-algorithm-based feature selection and SHAP achieved 2.49% higher accuracy on CCLE compared to the best prior model 1, while our RNN with GA reduced the mean squared error on GDSC by 46.34% relative to the previous benchmark.

6.5. Translational Potential and Clinical Execution

The proposed framework identified several promising repositioned drugs, including 5-fluorouracil, 5-azacytidine, Dactinomycin, and Docetaxel, which demonstrated sensitivity profiles across multiple cancer types beyond their conventional indications. Validation was primarily computational and supported through consistency with known pharmacogenomic findings and biomarker–drug relationships reported in the literature. The reported “85% oncologist approval” reflects qualitative expert agreement regarding the biological plausibility and clinical relevance of the top-ranked predictions based on pathway consistency, literature evidence, therapeutic feasibility, and translational potential.

Through investigator-initiated studies and extended access initiatives, our medication repositioning study finds immediate clinical translation potential. With already having proven safety profiles, and clinical experience, the medications displaying the greatest repositioning potential—Sorafenib, Erlotinib, and Palbociclib—dramatically lower the challenges to clinical implementation. Our explainable artificial intelligence study's mechanistic findings clearly reveal biomarker options for patient selection, and response tracking.

These results reflect a paradigm change from conventional sequential drug development to smart repositioning techniques using safety data, and current pharmacological understanding. Particularly for uncommon or treatment-resistant cancers, the significant decrease in development times—from perhaps decades for new drug development to months for repositioning validation—could greatly speed access to effective treatments for cancer patients [63].

7. Limitations

There are a few things that the current Separate XAI study should be aware of. First, all model construction and assessment are based only upon in vitro cell-line datasets (GDSC and CCLE). These datasets cannot completely capture the complexity of patient tumors since they do not include the tumor microenvironment, immunological interactions, or pharmacokinetics, which makes it harder to apply the results to real patients. Second, even while cross-dataset validation shows that GDSC, and CCLE are both strong, they both use the same profiling platforms, and have the same biases. This means that models may not work as well upon independent cohorts or upon new pharmacogenomic datasets that use other measurement methods. Third, the mutation, CNV, and expression characteristics employed here only analyze a small part of the molecular factors that might affect how a medicine works. For example, they do not analyze epigenetic changes, noncoding RNA impacts, or post-translational regulation, which could leave out important resistance mechanisms. Fourth, the genetic algorithm, and deep architectures need a lot of computer power to train and fine-tune hyperparameters. This might make it hard to use them in places with inadequate infrastructure. Fifth, SHAP gives explanations after the fact, however, it does not ensure that they are causative. Some highly rated traits may be related to sensitivity rather than driving it, and these biomarkers still need to be tested within experiments. Lastly, our predictions for medication repositioning are based upon statistical links, and expert evaluation; however, they have not been tested within a clinical or experimental setting yet. Without in vitro or in vivo confirmation, it is hard to say how they will affect patient treatment. Within future studies, these problems should be addressed through adding data arising from patients, increasing molecular modalities, improving computing performance, and doing wet-lab validation. This will make the framework more useful within the clinic, and more widely applicable.

8. Future Directions

In this work will focus on improving the robustness, generalizability and clinical applicability of the proposed Separate XAI framework. One important direction is to incorporate formal statistical hypothesis testing, such as paired *t*-tests, Wilcoxon signed-rank tests, and Friedman/Nemenyi post-hoc analysis across cross-validation folds, to rigorously evaluate the statistical significance of performance differences between competing models.

Generalization capability was assessed through cross-dataset validation between GDSC and CCLE, where models demonstrated stable predictive performance with limited degradation under dataset-shift conditions. The current framework primarily evaluates compounds represented within the training datasets; therefore, prediction on entirely unseen drugs remains a limitation of the present study. Performance variability across cancer types was observed due to biological heterogeneity and sample imbalance. While repeated cross-validation and cross-dataset evaluation partially assessed robustness, broader domain-shift analyses using external cohorts and institutional variability were beyond the scope of this work. Future studies will investigate graph-based molecular representations, domain adaptation, federated learning, and external validation cohorts to further improve generalization beyond GDSC and CCLE.

Future work will incorporate systematic pathway enrichment analysis using KEGG, (Kyoto Encyclopedia of Genes and Genomes, release 2024). Reactome, and Gene Ontology databases, as well as network-level explainability analysis, to further validate the biological consistency of SHAP-derived biomarkers.

Furthermore, future work will include external validation with independent pharmacogenomic cohorts and real-world clinical datasets to further assess model generalization across heterogeneous biological environments. Further extension of this framework to a

wider multi-omics integration including proteomics, metabolomics and epigenomics may improve the predictive accuracy and biological interpretability even further. An alternative promising direction is to combine the transformer-based and graph neural network architectures to better capture the complex gene-drug interactions as well as the pathway-level dependencies. Future work will also investigate federated and privacy-preserving learning techniques to allow secure multi-institutional collaboration in precision oncology research.

Although graph neural networks (GNNs) have recently demonstrated strong performance in pharmacogenomics by explicitly modeling biological interaction networks and pathway relationships, the current study focused on establishing an interpretable and computationally efficient multi-architecture framework using DNN, CNN, and RNN models. Future work will investigate pathway-aware and graph-based neural architectures to further improve biological representation learning and predictive generalization.

Finally, the translational potential of the proposed framework will be evaluated in real clinical settings through prospective clinical evaluation and expert-guided validation to support personalized cancer therapy selection and AI-assisted drug repositioning.

9. Conclusions

In this study, we propose a computational framework for cancer drug response prediction by integrating independent dataset-specific preprocessing, deep learning models, and explainable artificial intelligence on CCLE and GDSC datasets. Strict separation of datasets is enforced by the proposed approach, ensuring preservation of biological and experimental specificity, leading to better predictive performance and model robustness.

The experimental results demonstrate that the proposed framework has strong performance compared to existing state-of-the-art methods. Our best model (DNN + GA + XAI) achieves 94.49% accuracy on the CCLE dataset, and the proposed framework reaches a mean squared error (MSE) of 0.0725 on the GDSC dataset, outperforming other competitive methods such as consDeepSignaling and PathDNN. These results demonstrate the effectiveness of the proposed architecture for *in silico* drug response prediction.

Moreover, the integration of SHAP-based explainability offers interpretable explanations of important genomic features and signaling pathways, including important biomarkers TP53 and KRAS. This enhances the transparency of the model and facilitates biological interpretation of the predictions.

It should be noted that all results are obtained only by computational experiments. The implications for drug repositioning or clinical use, if any, remain speculative and require further study in both experimental and clinical settings. This work should be interpreted as a methodological contribution towards improved prediction modelling and explainability in cancer pharmacogenomics, not a clinically validated decision support system.

Author Contributions: H.M.N. conceived and designed the study. F.A.M. performed the experiments and data preprocessing. O.M.B. analyzed the data and prepared the figures. H.M.N. wrote the main manuscript text. A.G.O. provided valuable support and expert validation of the clinical and biological aspects of the proposed framework. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used in this study are publicly accessible. The Genomics of Drug Sensitivity in Cancer (GDSC) dataset can be accessed at <https://www.cancerrxgene.org> accessed on 1 May 2026, and the Cancer Cell Line Encyclopedia (CCLE) dataset is available at <https://portals.broadinstitute.org/ccle> accessed on 1 May 2026. Model implementation code used for analysis are available from the corresponding author upon reasonable request.

Acknowledgments: The authors acknowledge the Arab Academy for Science, Technology and Maritime Transport (AASTMT) for providing the computational resources and research environment necessary for this work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ferlay, J.; Laversanne, M.; Ervik, M.; Lam, F.; Colombet, M.; Mery, L.; Piñeros, M.; Znaor, A.; Soerjomataram, I.; Bray, F. *Global Cancer Observatory: Cancer Tomorrow (Version 1.1)*; International Agency for Research on Cancer: Lyon, France, 2024; Volume 2024.
2. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2024**, *74*, 229–263. [[CrossRef](#)] [[PubMed](#)]
3. Bizuayehu, H.M.; Ahmed, K.Y.; Kibret, G.D.; Dadi, A.F.; Belachew, S.A.; Bagade, T.; Tegegne, T.K.; Venchiarutti, R.L.; Kibret, K.T.; Hailegebireal, A.H.; et al. Global disparities of cancer and its projected burden in 2050. *JAMA Netw. Open* **2024**, *7*, e2443198. [[CrossRef](#)] [[PubMed](#)]
4. Siegel, R.L.; Kratzer, T.B.; Giaquinto, A.N.; Sung, H.; Jemal, A. Cancer statistics, 2025. *CA Cancer J. Clin.* **2025**, *75*, 10–45. [[CrossRef](#)] [[PubMed](#)]
5. Carli, F.; Di Chiaro, P.; Morelli, M.; Arora, C.; Bisceglia, L.; De Oliveira Rosa, N.; Cortesi, A.; Franceschi, S.; Lessi, F.; Di Stefano, A.L.; et al. Learning and actioning general principles of cancer cell drug sensitivity. *Nat. Commun.* **2025**, *16*, 1654. [[CrossRef](#)] [[PubMed](#)]
6. Meng, W.; Xu, X.; Xiao, Z.; Gao, L.; Yu, L. Cancer Drug Sensitivity Prediction Based on Deep Transfer Learning. *Int. J. Mol. Sci.* **2025**, *26*, 2468. [[CrossRef](#)] [[PubMed](#)]
7. Chowdhury, S.; Rajaganapathy, S.; Sun, L.; Wang, L.; Yang, P.; Cerhan, J.R.; Zong, N. SensitiveCancerGPT: Leveraging Generative Large Language Model on Structured Omics Data to Optimize Drug Sensitivity Prediction. *bioRxiv* **2025**. [[CrossRef](#)] [[PubMed](#)]
8. Ghassemi, M.; Naumann, T.; Schulam, P.; Beam, A.L.; Chen, I.Y.; Ranganath, R. A review of challenges and opportunities in machine learning for health. *AMIA Summits Transl. Sci. Proc.* **2020**, *2020*, 191–200. [[PubMed](#)]
9. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [[CrossRef](#)] [[PubMed](#)]
10. Obermeyer, Z.; Emanuel, E.J. Predicting the future—Big data, machine learning, and clinical medicine. *N. Engl. J. Med.* **2016**, *375*, 1216–1219. [[CrossRef](#)] [[PubMed](#)]
11. Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. A survey on bias and fairness in machine learning. *ACM Comput. Surv.* **2021**, *54*, 1–35. [[CrossRef](#)]
12. Badgeley, M.A.; Zech, J.R.; Oakden-Rayner, L.; Glicksberg, B.S.; Liu, M.; Gale, W.; McConnell, M.V.; Percha, B.; Snyder, T.M.; Dudley, J.T. Deep learning predicts hip fracture using confounding patient and healthcare variables. *npj Digit. Med.* **2019**, *2*, 31. [[CrossRef](#)] [[PubMed](#)]
13. Zhang, Z. Generalizability in clinical AI models: Lessons from deployment in healthcare settings. *J. Biomed. Inform.* **2022**, *128*, 104036.
14. Wang, Y.; Yu, X.; Gu, Y.; Li, W.; Zhu, K.; Chen, L.; Tang, Y.; Liu, G. XGraphCDS: An explainable deep learning model for predicting drug sensitivity from gene pathways and chemical structures. *Comput. Biol. Med.* **2024**, *168*, 107746. [[CrossRef](#)] [[PubMed](#)]
15. Sotudian, S.; Paschalidis, I.C. ITNR: Inversion Transformer-based Neural Ranking for cancer drug recommendations. *Comput. Biol. Med.* **2024**, *172*, 108312. [[CrossRef](#)] [[PubMed](#)]
16. Mohammadzadeh-Vardin, T.; Ghareyazi, A.; Gharizadeh, A.; Abbasi, K.; Rabiee, H.R. DeepDRA: Drug repurposing using multi-omics data integration with autoencoders. *PLoS ONE* **2024**, *19*, e0307649. [[CrossRef](#)] [[PubMed](#)]
17. Li, Y.; Guo, Z.; Gao, X.; Wang, G. MMCL-CDR: Enhancing cancer drug response prediction with multi-omics and morphology images contrastive representation learning. *Bioinformatics* **2023**, *39*, btad734. [[CrossRef](#)] [[PubMed](#)]
18. Bang, D.; Koo, B.; Kim, S. Transfer learning of condition-specific perturbation in gene interactions improves drug response prediction. *Bioinformatics* **2024**, *40*, i130–i139. [[CrossRef](#)] [[PubMed](#)]
19. Wang, C.; Zhang, M.; Zhao, J.; Li, B.; Xiao, X.; Zhang, Y. The prediction of drug sensitivity by multi-omics fusion reveals the heterogeneity of drug response in pan-cancer. *Comput. Biol. Med.* **2023**, *163*, 107220. [[CrossRef](#)] [[PubMed](#)]
20. Hajim, W.I.; Zainudin, S.; Daud, K.M.; Alheeti, K. Optimized models and deep learning methods for drug response prediction in cancer treatments: A review. *PeerJ Comput. Sci.* **2024**, *10*, e1903. [[CrossRef](#)] [[PubMed](#)]
21. Dong, Y.; Zhang, Y.; Qian, Y.; Zhao, Y.; Yang, Z.; Feng, X. ASGCL: Adaptive Sparse Mapping-based graph contrastive learning network for cancer drug response prediction. *PLoS Comput. Biol.* **2025**, *21*, e1012748. [[CrossRef](#)] [[PubMed](#)]
22. Wang, Y.; Jin, X.; Qiu, R.; Ma, B.; Zhang, S.; Song, X.; He, J. Developing and validating a drug recommendation system based on tumor microenvironment and drug fingerprint. *Front. Artif. Intell.* **2025**, *7*, 1444127. [[CrossRef](#)] [[PubMed](#)]

23. Baek, B.; Jang, E.; Park, S.; Park, S.H.; Williams, D.R.; Jung, D.W.; Lee, H. Integrated drug response prediction models pinpoint repurposed drugs with effectiveness against rhabdomyosarcoma. *PLoS ONE* **2024**, *19*, e0295629. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Adam, G.; Rampášek, L.; Safikhani, Z.; Smirnov, P.; Haibe-Kains, B.; Goldenberg, A. Machine learning approaches to drug response prediction: Challenges and recent progress. *npj Precis. Oncol.* **2020**, *4*, 19. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Nguyen, T.; Nguyen, G.T.; Nguyen, T.; Le, D.H. Graph convolutional networks for drug response prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *19*, 146–154. [\[CrossRef\]](#)
26. Jin, J.; Yu, Y.; Wang, R.; Zeng, X.; Pang, C.; Jiang, Y.; Li, Z.; Dai, Y.; Su, R.; Zou, Q.; et al. iDNA-ABF: Multi-scale deep biological language learning model for the interpretable prediction of DNA methylations. *Genome Biol.* **2022**, *23*, 219. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Chiu, Y.C.; Chen, H.I.H.; Zhang, T.; Zhang, S.; Gorthi, A.; Wang, L.J.; Huang, Y.; Chen, Y. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med. Genom.* **2019**, *12*, 18. [\[CrossRef\]](#)
28. Zulfikar, H.; Huang, Q.L.; Lv, H.; Sun, Z.J.; Dao, F.Y.; Lin, H. Deep-4mCGP: A deep learning approach to predict 4mC sites in *Geobacter pickeringii* by using correlation-based feature selection technique. *Int. J. Mol. Sci.* **2022**, *23*, 1251. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Xu, J.; Xu, J.; Meng, Y.; Lu, C.; Cai, L.; Zeng, X.; Nussinov, R.; Cheng, F. Graph embedding and Gaussian mixture variational autoencoder network for end-to-end analysis of single-cell RNA sequencing data. *Cell Rep. Methods* **2023**, *3*, 100382. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Yan, K.; Lv, H.; Guo, Y.; Peng, W.; Liu, B. sAMPpred-GAT: Prediction of antimicrobial peptide by graph attention network and predicted peptide structure. *Bioinformatics* **2023**, *39*, btac715. [\[PubMed\]](#)
31. Corsello, S.M.; Nagari, R.T.; Spangler, R.D.; Rossen, J.; Kocak, M.; Bryan, J.G.; Humeidi, R.; Peck, D.; Wu, X.; Tang, A.A.; et al. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* **2020**, *1*, 235–248. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 17–48. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Baptista, D.; Ferreira, P.G.; Rocha, M. Deep learning for drug response prediction in cancer. *Brief. Bioinform.* **2021**, *22*, 360–379. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Borisov, V.; Leemann, T.; Seßler, K.; Haug, J.; Pawelczyk, M.; Kasneci, G. Deep neural networks and tabular data: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 7499–7519. [\[CrossRef\]](#)
35. Zhang, H.; Chen, Y.; Li, F. Predicting Anticancer Drug Response With Deep Learning Constrained by Signaling Pathways. *Front. Bioinform.* **2021**, *1*, 639349. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Choi, J.; Park, S.; Ahn, J. RefDNN: A reference drug based neural network for more accurate prediction of anticancer drug resistance. *Sci. Rep.* **2020**, *10*, 1861. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Kothari, S.; Sharma, S.; Shejwal, S.; Kazi, A.; D'Silva, M.; Karthikeyan, M. An Explainable AI-Assisted Web Application in Cancer Drug Value Prediction. *MethodsX* **2024**, *12*, 102696. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Proietti, M.; Ragno, A.; Rosa, B.L.; Moreau, Y.; Fariselli, P.; Raimondi, D. Explainable AI in Drug Discovery: Self-Interpretable Graph Neural Network for Molecular Property Prediction Using Concept Whitening. *Mach. Learn.* **2024**, *113*, 2013–2044. [\[CrossRef\]](#)
39. Sharifi-Noghabi, H.; Zolotareva, O.; Collins, C.C.; Ester, M. MOLI: Multi-Omics Late Integration with Deep Neural Networks for Drug Response Prediction. *Bioinformatics* **2019**, *35*, i501–i509. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Tang, J.; Gottlieb, A. PathDSP: Explainable drug sensitivity prediction through pathway enrichment. *Bioinformatics* **2021**, *37*, 2426–2434. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Deng, L.; Cai, Y.; Zhang, W.; Yang, W.; Gao, B.; Liu, H. Pathway-Guided Deep Neural Network toward Interpretable and Predictive Modeling of Drug Sensitivity. *J. Chem. Inf. Model.* **2020**, *60*, 4497–4505. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Wu, J.; Lai, J.; Zhao, X.; Wang, Z.; Zhang, Y.; Wang, L.; Su, Y.; He, Y.; Li, S.; Jiang, Y.; et al. DeepCCDS: Interpretable Deep Learning Framework for Predicting Cancer Cell Drug Sensitivity through Characterizing Cancer Driver Signals. *Adv. Sci.* **2025**, *12*, e2416958. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Yu, G.; Fan, Q. Deep Learning-Driven Drug Response Prediction and Mechanistic Insights in Cancer Genomics. *Sci. Rep.* **2025**, *15*, 20824. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Codicè, F.; Pancotti, C.; Rollo, C.; Moreau, Y.; Fariselli, P.; Raimondi, D. The Specification Game: Rethinking the Evaluation of Drug Response Prediction for Precision Oncology. *J. Cheminform.* **2025**, *17*, 33. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Partin, A.; Brettin, T.S.; Zhu, Y.; Narykov, O.; Clyde, A.; Overbeek, J.; Stevens, R.L. Deep learning methods for drug response prediction in cancer: Predominant and emerging trends. *Front. Med.* **2023**, *10*, 1086097. [\[CrossRef\]](#)
46. DeLong, L.N.; Gadiya, Y.; Galdi, P.; Fleuriot, J.D.; Domingo-Fernández, D. MARS: A Neurosymbolic Approach for Interpretable Drug Discovery. *arXiv* **2024**, arXiv:2410.05289. Available online: <https://arxiv.org/abs/2410.05289> (accessed on 1 May 2026).
47. Han, Y.; Kim, S.; Jeong, E.; Lee, S.; Yun, S.; Lim, S. DiSPA: Differential Substructure-Pathway Attention for Drug Response Prediction. *arXiv* **2026**, arXiv:2601.14346. Available online: <https://arxiv.org/abs/2601.14346> (accessed on 1 May 2026).
48. Giloni, A.; Grolman, E.; Hagemann, T.; Fromm, R.; Fischer, S.; Elovici, Y.; Shabtai, A. BENN: Bias estimation using a deep neural network. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 117–131. [\[CrossRef\]](#)

49. Somani, S.; Russak, A.J.; Richter, F.; Zhao, S.; Vaid, A.; Chaudhry, F.; De Freitas, J.K.; Naik, N.; Miotto, R.; Nadkarni, G.N.; et al. Deep learning and the electrocardiogram: Review of the current state-of-the-art. *EP Eur.* **2021**, *23*, 1179–1191. [[CrossRef](#)]
50. Zagidullin, B.; Aldahdooh, J.; Zheng, S.; Wang, W.; Wang, Y.; Saad, J.; Malyutina, A.; Jafari, M.; Tanoli, Z.; Pessia, A.; et al. DrugComb: An integrative cancer drug combination data portal. *Nucleic Acids Res.* **2019**, *47*, W43–W51. [[CrossRef](#)] [[PubMed](#)]
51. Zheng, S.; Aldahdooh, J.; Shadbahr, T.; Wang, Y.; Aldahdooh, D.; Bao, J.; Wang, W.; Tang, J. DrugComb update: A more comprehensive drug sensitivity data repository and analysis portal. *Nucleic Acids Res.* **2021**, *49*, W174–W184. [[CrossRef](#)] [[PubMed](#)]
52. Hegselmann, S.; Buendia, A.; Lang, H.; Agrawal, M.; Jiang, X.; Sontag, D. Tabllm: Few-shot classification of tabular data with large language models. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Valencia, Spain, 25–27 April 2023; pp. 5549–5581.
53. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.H.; Kang, J. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **2020**, *36*, 1234–1240. [[PubMed](#)]
54. Li, M.; Wang, Y.; Zheng, R.; Shi, X.; Li, Y.; Wu, F.X.; Wang, J. DeepDSC: A deep learning method to predict drug sensitivity of cancer cell lines. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *18*, 575–582. [[CrossRef](#)]
55. Fekete, J.T.; Györfy, B. A unified platform enabling biomarker ranking and validation for 1562 drugs using transcriptomic data of 1250 cancer cell lines. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 2885–2894. [[CrossRef](#)]
56. Ucak, U.V.; Ashyrmamatov, I.; Lee, J. Reconstruction of Lossless Molecular Representations from Fingerprints. *J. Cheminform.* **2023**, *15*, 26. [[CrossRef](#)] [[PubMed](#)]
57. Ha, S.; Park, J.; Jo, K. Comparative Analysis of Regression Algorithms for Drug Response Prediction Using GDSC Dataset. *BMC Res. Notes* **2025**, *18*, 10. [[CrossRef](#)] [[PubMed](#)]
58. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
59. Min, S.; Lee, B.; Yoon, S. Deep learning in bioinformatics. *Brief. Bioinform.* **2017**, *18*, 851–869. [[CrossRef](#)] [[PubMed](#)]
60. Tran, N.K.; Klau, G.W. Drug Response Prediction: A Critical Systematic Review of Current Datasets and Methods. *Pattern Recognit. Lett.* **2026**, *199*, 21–26. [[CrossRef](#)]
61. Cai, Z.; He, J.; Yu, B.; Wei, L. MR.DRP: A Parameter Adaptive Machine Learning Framework for Drug Response Prediction. In Proceedings of the 2022 4th International Conference on Applied Machine Learning (ICAML), Changsha, China, 23–25 July 2022; pp. 61–67. [[CrossRef](#)]
62. Narayan, R.S.; Molenaar, P.; Teng, J.; Cornelissen, F.M.; Roelofs, I.; Menezes, R.; Dik, R.; Lagerweij, T.; Broersma, Y.; Petersen, N.; et al. A cancer drug atlas enables synergistic targeting of independent drug vulnerabilities. *Nat. Commun.* **2020**, *11*, 2935. [[CrossRef](#)] [[PubMed](#)]
63. Xia, F.; Allen, J.; Balaprakash, P.; Brettin, T.; Garcia-Cardona, C.; Clyde, A.; Cohn, J.; Doroshov, J.; Duan, X.; Dubinkina, V.; et al. A cross-study analysis of drug response prediction in cancer cell lines. *Brief. Bioinform.* **2022**, *23*, bbab356. [[PubMed](#)]
64. Ahmed, H.; Shedeed, H.; Hamad, S.; Saad, A. Convolutional neural network models for cancer treatment response prediction. *Int. J. Intell. Comput. Inf. Sci.* **2023**, *23*, 98–105. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.