

A Biomathematical Note on Machine-Learning Models for Predicting Drug Response in Digital Health Systems

Prachi Juyal^{1,*}, Ms. Sonali Gupta², Dr. Prince Verma³, Mohit Kumar⁴, Dr. Dheeresh Upadhyay⁵, Dr. Charudatta Dattatraya Bele⁶

¹Associate Professor, Graphic Era Deemed to be University, India. ²Assistant Professor, Department of Aeronautical Engineering, Nitte Meenakshi Institute of Technology, Bengaluru, Karnataka, India. ³Assistant Professor, Department of Computer Science and Engineering, Lovely Professional University Specialization: Deep Learning and Big Data Analytics, India. ⁴Assistant Professor, Teerthanker Mahaveer College of Pharmacy, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, 244001, India. ⁵Associate Professor, Department of Electrical and Electronics Engineering, Mangalayatan University, Aligarh-202146, Uttar Pradesh, India. ⁶Associate Professor, Department of Mathematics Shri Shivaji College, Parbhani (MH), India.

ABSTRACT

The growing integration of machine learning into industries like digital health systems has made interest in computational models for the prediction of drug response using high-dimensional molecular data more intense. The mathematical form, interpretability and strength of such models are, however, important issues. We provide a biomathematical analysis of machine-learning models of drug response prediction using a standardized cross-study pharmacogenomic benchmark of drug-cell line pairs with molecular-feature representations and log-normalized IC50 values in this study. A comparative framework was used to test regularized linear models (Ridge, Lasso, ElasticNet), robust regression and a few of the selected nonlinear baselines with identical data splits and preprocessing conditions. The error-based and rank-based measures of model performance were used, with the support of sparsity analysis, residual diagnostics, and bootstrap stability analysis. The findings indicate that sparsity-favoring linear models can be used to predict drug response as well as or better than more complicated nonlinear models, and find a smaller effective feature space. The residual and stability tests also show that the model has strong and unbiased predictive power and has minimal heteroscedasticity. These results indicate that mathematically constrained, interpretable frameworks are useful to predict drug responses, which justifies the usefulness of biomathematical machine-learning models as accurate elements of digital health decision-support systems.

ARTICLE HISTORY

Received September 20, 2025

Accepted December 13, 2025

KEYWORDS

Biomathematical Modeling,
Drug Response Prediction,
Machine Learning,
Digital Health Systems,
Pharmacogenomic

1. Introduction

The growing digitisation of healthcare systems has led to the availability of biological, molecular and clinical data on a large scale, which has opened up the possibility of computational approaches to inform therapeutic decision-making. Within this digital health context, the prediction of drug response has become quite central as an issue, especially in the field of oncology and pharmacogenomics, where heterogeneous biological systems show very variable responses to pharmacological interventions. Predicting drug response from molecular data is by nature a biomathematical problem, involving the construction of quantitative models that relate high-dimensional feature spaces to continuous phenotypic responses with variations in biological variability, noise and uncertainty (Al-Zirkan et al., 2025).

Recent developments in artificial intelligence have elevated the pace of research in healthcare analytics and their applications can range anywhere from diagnoses to optimisation of treatments (Chatterjee et al., 2024). However, the increasing dependence on complex machine-learning architectures has raised issues in aspects of interpretability, robustness and reproducibility - key requirements to enable their deployment within digital health systems. Explainable and mathematically-based models are increasingly being acknowledged as a necessity especially in the precision medicine scenarios where the model outputs may be used to design an experiment or even to present further clinical hypotheses (Allen, 2024). As such, there is renewed interest in biomathematical approaches which place an emphasis onto structured modeling, regularization, and analysis of error in a transparent manner.

Drug response prediction has been explored widely using machine learning techniques, and in many cases using

extensive pharmacogenomic data sets containing drug-cell line response pairs (Firoozbakht et al., 2022). These datasets generally contain thousands of molecular features obtained from gene expression or other omics modalities, and present great mathematical challenges because of high dimensionality, multicollinearity and low sample sizes. Although deep learning and ensemble techniques have been suggested to represent complicated nonlinear interactions (Hajim et al., 2024; Hakami, 2024), they can be moderately better than the current methods but require interpretability and are more susceptible to noise.

The central question of a biomathematical interest is this: are apparent complexity of molecular feature spaces representative of the real dimensionality of biological processes underlying drug response? Mathematical modeling of healthcare has been focusing on the concept of reduced-order structures for a long time, in which the system behavior is characterized by a few dominant variables (Liu et al., 2023). In drugs response modeling, sparsity-promoting regression methods offer a natural hypothesis test where it is explicitly assumed that only a limited number of features are used in the prediction. Such constraints are in agreement with biological intuition in which only subsets of genes or pathways are to be expected to play decisive roles in mediating drug sensitivity or resistance (AL-Rajabi et al., 2025).

Simultaneously, the concepts of robustness and stability have become a major concern in computational healthcare studies. Biological data is noisy and heterogeneous in nature and predictive models are required to be robust to sampling variability and outliers to be useful in practice. One of the points made by Hassan et al. (2024) is the importance of combining machine learning and mathematical modeling techniques, such as regularization and uncertainty quantification, to ensure reliable predictions. The use of residual diagnostics, bootstrap resampling, and sensitivity analyses is thus a vital part of a rigorous biomathematical study, especially in the digital health system.

Although a lot of research has been done, there are still a number of limitations in the existing literature. Predictive accuracy is often the focus of many studies without enough analysis of the model structure or error behavior so that the validity of underlying assumptions can be evaluated. Also, the comparisons between linear and nonlinear models are usually performed in dissimilar experimental conditions, which restrict their interpretation. Lastly, such models are not necessarily discussed as a part of digital health infrastructures where transparency, reproducibility, and regulatory concerns are the primary concerns (Hassan et al., 2024; Hassan et al., 2022).

These gaps are the reasons why the current research is driven and takes a consciously narrow focus. Instead of introducing new algorithms, the paper is a biomathematical note that methodically compares a collection of traditional and popular machine-learning models in the context of drug response prediction using a single and reproducible framework. The study evaluates the performance of linear models with various regularization schemes (ridge, lasso, elastic net), a robust regression model (Huber), as well as nonlinear baselines of interest (k-nearest neighbors and random forests) on a cross-study benchmark dataset of monotherapy drug response prediction. Training and evaluation of all the models are done with the same data splits, feature representations and performance metrics, which makes them fair.

The range of this study is narrow. The analysis is limited to in vitro data of drug-cell line response with the drug sensitivity measured in log-normalized IC50. There is no patient-level clinical data to be taken into account and no statement about direct clinical applicability. The models are rather placed in a role of a part of computational pipelines in digital health research, including hypothesis generation, benchmarking, and decision-support during early-stage drug development. This breadth indicates that its approach is methodological as opposed to translational directness.

This scope has a number of limitations. The representations of the features are preprocessed and latent and do not allow the direct biological interpretation at the level of individual genes. It only analyzes monotherapy responses and fails to capture the dynamic over time and drug-drug interactions. Moreover, linear and shallow nonlinear models might not be sufficient in high-order biological interactions. These constraints are recognized as part of the benchmark environment and are presented as guidelines on how the biomathematical and computational extensions can be made in the future.

Nevertheless, the research has a significant contribution to the knowledge on drug response prediction on a biomathematical perspective despite these limitations. The analysis empirically supports the hypothesis that drug response is controlled by an effective structure of lower dimensions hidden in a high-dimensional molecular space by showing that sparsity-promoting linear models can be as effective as more complex nonlinear methods. The use of residual diagnostics and bootstrap stability analysis also enhances the validity of the results and highlights the need of the uncertainty-conscious model in digital health systems.

In this context, the aims of the study are threefold, i.e. (i) to compare a set of linear, sparsity-constrained, robust, and nonlinear machine-learning models systematically in a common benchmark framework (ii) to examine the mathematical characteristics of these models, i.e. sparsity, residual behavior, and stability; and (iii) to interpret the obtained results in the context of digital health systems, where interpretability, robustness and reproducibility are crucial.

In this biomathical perspective, the current work will help in explaining the importance of classical machine-learning models in predicting drug responses and in advancing the creation of transparent and mathematically-based computational aids in digital health studies.

2. Literature Review

The use of machine learning and mathematical modeling in healthcare has grown at a very fast pace due to the necessity to process complex and high-dimensional biomedical data in digital health systems. Initial research was focused on the creation of powerful computing models that would be able to process vast amounts of healthcare data. [Praveen et al. \(2022\)](#) suggested a machine learning-based framework of healthcare data management with the help of big data engineering methods, which emphasizes the scalability and robustness of the system. Although these frameworks were useful in integrating data, they did not have a clear mathematical form to be interpreted and analyze uncertainty, which restricted their application in biomathematical inference.

With the development of research towards precision and personalized medicine, the combination of mathematical modeling and machine learning gained more and more popularity. As shown by [Ramaswamy and Keidar \(2023\)](#), personalized plasma medicine approaches can be informed by the use of mathematical models alongside machine learning to guide the cancer treatment. Their article highlights the need to incorporate physical and biological limits into predictive models, which supports one of the key biomathematical principles, namely predictive accuracy based on mathematically coherent representations of biological systems and not just on data-driven heuristics. The increased use of artificial intelligence in medicine has led to attempts to formalize the conceptual and mathematical basis of it. [Rashidi et al. \(2025\)](#) also reviewed the generative and non-generative AI methods in pathology and medicine with a focus on the trade-offs between the model complexity, interpretability, and clinical utility. Their discussion points out that more intricate architectures do not only fail to enhance predictive reliability especially in the biomedical contexts where predictability is dominated by heterogeneity and noise in the data. This finding is in line with biomathematical arguments that support parsimonious models with clear assumptions.

Mathematical modeling has been also very important in epidemiology and public health and has presented transferable information on drug response prediction. [Sabherwal et al. \(2024\)](#) assessed the mathematical models in predicting the dynamics of COVID-19 transmission, emphasizing the significance of assumptions in the models, sensitivity of the parameters, and quantifying the uncertainty. Even though they are concerned with the modeling of infectious diseases, their conclusions emphasize a more general biomathematical point: predictive models should not be judged solely on the basis of their performance, but on the basis of their mathematical structure and error behavior.

Machine learning has found more and more applications in drug discovery and optimization of therapy decisions in order to shorten the time spent on prediction and decision making. In their study, [Stergiou et al. \(2022\)](#) created machine learning predictive models of epidemics and accelerated drug discovery and showed how predictive analytics can be applied to pharmaceutical research. Nevertheless, their effort also indicates difficulties associated with generalizability and strength especially when models are used outside their training distributions. These challenges are also highlighted in systematic reviews. According to [Uddin et al. \(2024\)](#), interpretability, data quality, and validation were the barriers that persist in making machine learning in healthcare effective. On the same note, in their survey of machine learning applications in geriatric medicine, [Woodman and Mangoni \(2023\)](#) observed that predictive models can assist with clinical decision-making, but their use is limited by a lack of transparency and mathematical explanation. Machine learning has been extensively studied in oncology in the area of prognosis and treatment choice. In [Zhang et al. \(2023\)](#), the authors carried out a critical analysis of AI-assisted cancer prediction models and warned of overfitting and the lack of transparency of the algorithm. They claim that more easily interpretable, mathematically simple models can provide more accurate information especially when there are few biological signals compared to the dimensionality of the molecular data.

The recent developments in the systems biology modeling also exemplify the importance of biomathematical integration. The article by [Živanović and Filipović \(2024\)](#) investigated the use of system biology in drug optimization in in silico clinical trials, which involves mathematical modeling and machine learning to model complex physiological dynamics. Their work is the example of how structured mathematical models can be used to supplement data-driven approaches to improve interpretability and strength. Altogether, the literature demonstrates that there is a strong necessity to provide biomathematically-based studies that critically assess machine-learning models in healthcare. Although sophisticated algorithms are still on the rise, there is still a gap in research that focuses on the structure of models, sparsity, and stability and analysis of errors in standardized benchmark conditions. The gap should be addressed to develop not only accurate but also interpretable and reliable drug response prediction models that can be used in digital health systems.

3. Methodology

3.1 Research Design

The research design chosen in this study is quantitative, computational, and biomathematical because it seeks to systematically assess the machine-learning models of drug-response prediction in the context of a unified and reproducible experimental design. It is not maximizing the predictive performance itself that is of interest but instead an analysis of the

mathematical structure, stability, and interpretability of typical predictive models applied to high-dimensional molecular data applicable to digital health systems.

The research design, which views drug response prediction as a high-dimensional regression problem, supposes that observed phenotypic responses (drug sensitivity) are caused by an underlying lower-dimensional biological process in a large space of molecular features. This is the reason why regularized and sparsity promoting models are used and may be viewed as structural restrictions on the solution space of the regression problem.

Formally, the predictive task is defined as learning a mapping

$$f: \mathbb{R}^p \rightarrow \mathbb{R}, \quad (1)$$

Where $\mathbf{x}_i \in \mathbb{R}^p$ denotes the molecular representation of the i -th drug-cell line pair, $p = 958$ is the feature dimension, and the scalar response $y_i \in \mathbb{R}$ represents the log-normalized half-maximal inhibitory concentration (IC50). The learned function f is evaluated on unseen samples to assess generalization performance.

The experiment will be based on a benchmark-based comparative design, which is the training and analysis of several machine-learning algorithms with the same data splits and preprocessing. This design has the advantage of allowing observed differences in performance to be explained by the structure of the models and mathematical assumptions, instead of by data leakage or non-experimental consistency.

3.2 Data Collection Methods

3.2.1 Data Source

The data used in this paper is a curated pharmacogenomic benchmark of drug response measurements in large sets of cancer cell lines. The benchmark was built by means of a standardized pipeline that aligns drug sensitivity data and molecular profiles data across various experimental studies, which makes these data consistent and reproducible across diverse experimental sources (Hider et al., 2024). A drug-cell line pair is associated with each observation, and the response variable is the logarithm of half-maximal inhibitory concentration of the drug, denoted IC50, which is a popular quantitative measure of the drug activity. The use of log-normalization is done to stabilize the variance and overcome skewness in the response distribution, which is in line with the established pharmacometric modeling techniques and statistical assumptions of regression-based biomathematical analysis.

3.2.2 Feature Representation

Each drug-cell line pair is provided with a fixed length feature vector of dimension $p = 958$. These properties are a condensed molecular profile of high-dimensional omics measurements, such as gene expression and corresponding molecular descriptors. Although the original biological measurements can represent tens of thousands of variables, the benchmark representation transforms them into a low-dimensional latent space that can be used to regression model.

Let the dataset be denoted as

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N, \quad (2)$$

Where $N = 9519$ is the total number of samples, $\mathbf{x}_i \in \mathbb{R}^{958}$, and $y_i \in \mathbb{R}$.

3.2.3 Data Splitting

The dataset is divided into training, validation and test sets by using predefined splits provided with the benchmark:

- Training set: $N_{\text{train}} = 7616$
- Validation set: $N_{\text{val}} = 952$
- Test set: $N_{\text{test}} = 951$

These divides are posed in all experiments to remove randomness in model evaluation and can compare algorithms directly.

3.3 Population and Sampling

3.3.1 Population Definition

The interest population is the in vitro drug response experiments done on human cancer cell lines. The cells lines are biologically different in terms of their genome and the drugs manipulate the cellular processes in a certain way based on the mechanisms of action at the molecular level. Biomathematically, the population can be considered to be a group of biological dynamical systems and each respond to pharmacological intervention.

It should be stressed that such population is not related to human patients. It is instead a proxy population that is common in drug discovery early-stage and systems pharmacology.

3.3.2 Sampling Strategy

The sampling method is deterministic and benchmark-based and not probabilistic. All the drug-cell line pairs that pass through the inclusion criteria of the benchmark are utilized. Further sub-sampling and resampling is not done in the collection of data.

Nevertheless, in model testing, bootstrap resampling is used on the test data to measure uncertainty in predictive accuracy. Given a test set $\mathcal{T}$ of size n , bootstrap samples $\mathcal{T}^{(b)}$ are generated by sampling with replacement, and performance metrics are computed for each bootstrap replicate.

It is a method that gives an opportunity to estimate empirical confidence intervals without the need to make assumptions about parametric distributions of prediction errors.

3.4 Data Analysis Techniques

3.4.1 Problem Formulation

The core analytical task is formulated as a regularized regression problem. Given the design matrix $X \in \mathbb{R}^{N \times p}$ and response vector $y \in \mathbb{R}^N$, the objective is to estimate a coefficient vector $\beta \in \mathbb{R}^p$ such that

$$\hat{y} = X\beta \quad (3)$$

Approximates the observed responses under suitable regularization constraints.

3.4.2 Linear and Regularized Models

3.4.2.1 Ridge Regression

Ridge regression estimates coefficients by solving

$$\hat{\beta}_{\text{ridge}} = \arg \min_{\beta} \{\|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2\}, \quad (4)$$

where $\lambda > 0$ controls the degree of ℓ_2 -regularization. This formulation penalizes large coefficients and improves numerical stability in high-dimensional settings.

3.4.2.2 ElasticNet Regression

ElasticNet combines ℓ_1 and ℓ_2 penalties:

$$\hat{\beta}_{\text{EN}} = \arg \min_{\beta} \{\|y - X\beta\|_2^2 + \alpha(\rho \|\beta\|_1 + (1 - \rho) \|\beta\|_2^2)\}, \quad (5)$$

Where $\alpha > 0$ controls overall regularization strength and $\rho \in [0,1]$ balances sparsity and smoothness. From a biomathematical viewpoint, the ℓ_1 term enforces feature selection, consistent with the hypothesis that drug response depends on a limited subset of molecular factors.

3.4.2.3 Additional Models

In order to test robustness and nonlinearity, other models are tested, such as Lasso regression, Huber regression (robust to heavy-tailed errors), k-nearest neighbors' regression, and random forest regression. These models offer complimentary views to the bias-variance trade-off and nonlinear effects.

3.4.3 Model Evaluation Metrics

Performance is quantified using multiple metrics:

- Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

- Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

- Pearson correlation coefficient, measuring linear association.

- Spearman rank correlation coefficient, measuring monotonic association

The incorporation of rank-based measures is especially applicable in the pharmacological setting, in which relative values of drug sensitivity are usually more valuable than absolute ones.

3.4.4 Residual and Stability Analysis

Residuals $r_i = y_i - \hat{y}_i$ are analyzed to assess model adequacy. Distributional analysis and plot involving residual versus prediction are some residual diagnostics that are used to identify bias, heteroscedasticity or systematic deviations. The analysis of bootstrap stability is done by resampling the test set repeatedly and re-computing the performance measures, which gives empirical confidence intervals that indicate the variability of the sampling.

3.5 Ethical Considerations

This paper only makes use of publicly available, de-identified in vitro data. No human participants, patient records or personal identifiable information is used. In this regard, the research will not need the institutional review board (IRB) approval. Ethically, the study is based more on transparency and reproducibility, using all documented and reproducible computational interactions, in which model decisions and analysis procedures remain clear and determinable. The interpretation of the results is done in the responsible manner, where predictive outputs are presented as computational decision-support tools and not in the form of clinical or patient-level predictions. In addition, the study expressly does not engage in algorithmic overclaiming by admitting the inherent limitations of the in vitro drug response models and make no extrapolations of the results to other contexts. Through the following principles, the study will comply with the current ethical standards in computational biology and digital health research, which will encourage responsible machine-learning model usage in biomedical practices.

4. Results

4.1 Data Presentation and Experimental Overview

The experimental analysis was performed on an experimental pharmacogenomic dataset of 9519 drug-cell line response pairs, each with a 958-dimensional molecular feature vector and a scalar response variable of log-normalized IC50. The data was divided into fixed training (7,616 samples), validation (952 samples), and test (951 samples) sets, which ensures the reproducibility and the uniform comparison between models.

All models were trained using only the training set, hyperparameters were fixed using validation performance where appropriate and all reported results are to the held-out test set, unless otherwise. The evaluation of performance was carried out based on the combination of error-based (RMSE, MAE) and association-based (Pearson and Spearman correlation) metrics, which present complementary views regarding predictive accuracy and consistency between biological ranking.

4.2 Comparative Model Performance

In order to give a detailed quantitative comparison between models, Table 1 summarizes validation and test-set performance both using error-based and association-based metrics. The measurement in this tabular form supplements the graphical comparisons presented in the following figures and allows straightforward numerical evaluation of predictive accuracy, consistency rank and behavioral so-called generalization among model classes.

Table 1: Quantitative Performance Comparison of Evaluated Models

Model	VAL RMSE	VAL MAE	VAL Pearson r	VAL Spearman ρ	TEST RMSE	TEST MAE	TEST Pearson r	TEST Spearman ρ
Lasso ($\alpha = 0.001$)	0.1531	0.1192	0.1558	0.1895	0.1631	0.1242	0.1562	0.1881
Huber ($\epsilon = 1.35$, $\alpha = 0.0001$)	0.1565	0.1132	0.1482	0.1938	0.1691	0.1211	0.1469	0.1849
Random Forest (n = 400, leaf = 2)	0.1564	0.1213	0.1316	0.1826	0.1663	0.1259	0.1219	0.1669
KNN (k = 15)	0.1565	0.1215	0.1331	0.1846	0.1661	0.1255	0.1285	0.1759

Table 1. Quantitative comparison of machine-learning models on validation and test sets based on error-based metrics (RMSE, MAE) and association-based metrics (Pearson and Spearman correlation). Reduced values of RMSE and MAE mean higher predictive accuracy whereas increased values of correlation mean stronger agreement between the predicted and observed drug response.

4.2.1 Error-Based Performance

In order to provide a quantitative comparison of the predictive accuracy of the compared machine-learning models,

the test-set performance was measured with the help of the root mean squared error (RMSE), a common error-based measure of performance that punishes large deviations between predicted and measured drug response values. This comparison gives an idea of the degree to which model structure and regularization affect absolute prediction error in high-dimensional pharmacogenomic problems.

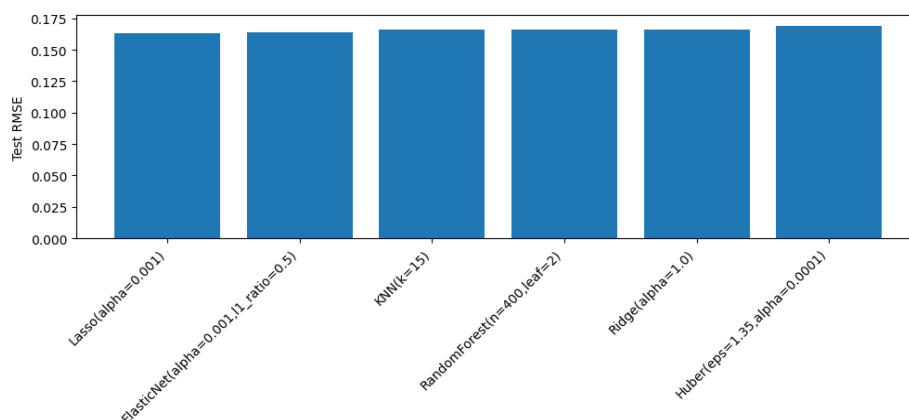


Figure 1: Model Comparison on Test Set (RMSE).

Comparison of test-set root mean squared error (RMSE) of assessed machine-learning models (Figure 1), such as Lasso, ElasticNet, Ridge, Huber regression, k-nearest neighbors (KNN), and random forest regression. When the RMSE values are smaller, it means that the predictive accuracy of the unobserved drug-cell line response data is better.

It reveals that there is a similar range of the RMSE values in all models (around 0.16 to 0.17), with sparsity-favored linear models like Lasso and ElasticNet having slightly smaller errors than Ridge regression. Nonlinear models, such as KNN and random forests, on the contrary, fail to show significant gains in RMSE, implying that the additional complexity of the model does not bring significant gains given the current feature representation.

Biomathematically, these findings suggest that the effects of drugs can be well-modeled by linear function of a limited set of molecular characteristics, which is a restatement of the applicability of regularized linear modeling models to high-dimensional biological processes.

4.2.2 Rank-Based Association and Biological Relevance

Whereas error-based measures of prediction accuracy are a measure of absolute accuracy, rank-based measures are especially significant in pharmacological modeling, in which relative ranking of drug sensitivity may have more biological significance than absolute values. Figure 2 provides the comparison of test-set Spearman rank correlation among the tested machine-learning models, in terms of their capacity to maintain monotonic relationships between predicted and observed drug responses.

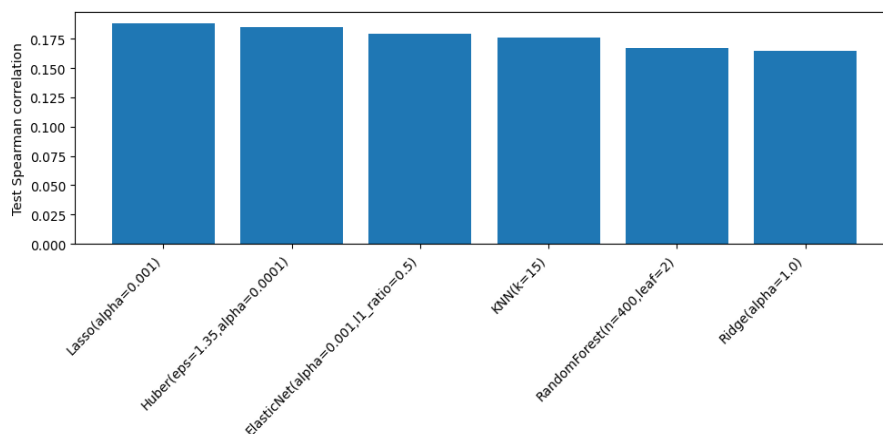


Figure 2: Model Comparison on Test Set (Spearman Correlation).

Comparison of test-set Spearman rank correlation between assessed machine-learning models (Figure 2), such as Lasso, ElasticNet, Ridge, Huber regression, k-nearest neighbors (KNN) and random forest regression. Greater values of Spearman correlation denote greater monotonic agreement between predicted and observed drug sensitivity.



Lasso regression has the best Spearman rank correlation on test set, closely followed by Huber regression and ElasticNet, and Ridge regression has the worst rank correlation of the methods evaluated. This trend shows that sparsity-inducing and robust linear models are more efficient at maintaining the relative ranking of drug sensitivity across samples. It is interesting to note that nonlinear methods like KNN and random forests do not have a higher rank correlation than linear methods, which supports the conclusion that the main signal that is being represented by the benchmark feature representation is monotonic and relatively linear. Biomathematically, these findings indicate that using mathematical structure by regularization and robustness restrictions may be more useful than the inclusion of other nonlinear flexibility.

4.3 Observed vs. Predicted Drug Response

In order to graphically analyze the consistency between predicted and observed values of drug responses, Figure 3 shows a scatter plot of observed and predicted IC50 values on the test set of the most successful sparsity-promoting model, Lasso regression with regularization parameter $\alpha = 0.001$. This visualization is a qualitative measure of the monotonicity and dispersion of model predictions, not just based on numerical measures of performance.

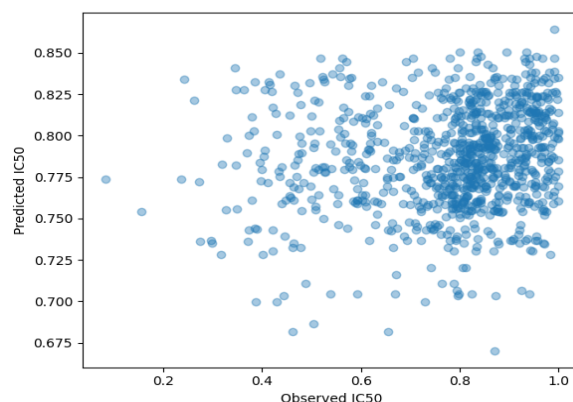


Figure 3: Observed vs. Predicted IC50 Values on the Test Set (Lasso Regression).

Observed and predicted IC50 values on the test set of the Lasso regression model ($\alpha = 0.001$) (Figure 3). The points indicate drug-cell line pairs, and the deviations of the identity line indicate prediction error. The values of the predicted IC50 values have a distinct monotonic relationship with the observed values, which means that the model reflects the global response gradients between drug-cell line pairs. Though the points are not clustered closely along the identity line, the general trend is similar to the moderate but statistically significant Spearman correlations that have been reported above. The standard deviation of the diagonal is indicative of the inherent complexity of prediction of drug responses due to biological heterogeneity, noise in experiments and unmeasured confounding variables. Notably, the lack of severe clustering or orderly deviation indicates that the model does not reduce meaningful variability in predictions to a small range, which is why the model is appropriate in ranking-based pharmacological analysis.

4.4 Residual Diagnostics and Error Structure

4.4.1 Residuals versus Predicted Values

Figure 4 provides residual diagnostics of the test set to determine the adequacy of the model and test the validity of assumptions of underlying regression. In particular, the plot of residuals (difference between observed and predicted IC50) vs. predicted IC50 is used to test bias, structure of variance and possible violations of the assumption of linear modeling.

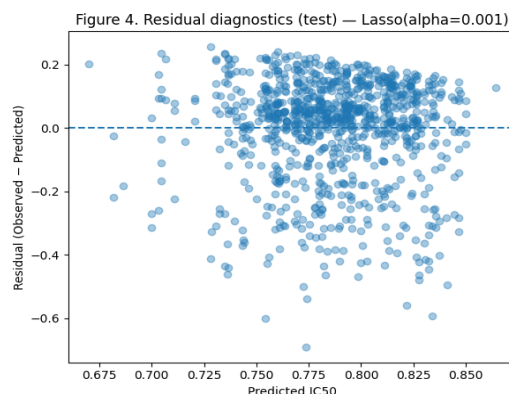


Figure 4: Residual Diagnostics on the Test Set (Lasso Regression).

Test set residual diagnostics of the Lasso regression model ($\alpha = 0.001$) (Figure 4). The residuals are calculated and are the difference between observed and predicted IC50 values versus the predicted IC50 and plotted. The general trend of the residuals around zero is observed over the entire spectrum of the predicted IC50 values, which means that there is no significant systematic bias in the model predictions. There is no significant funnel-like structure, which implies that heteroscedasticity is limited and that the error variance of prediction is not very large and is relatively constant at different levels of responses. Nevertheless, there is a clear imbalance in the residual distribution, which is a heavier negative tail. It means that the model tends to underestimate IC50 in a group of highly resistant drug-cell line pairs. Biomathematically, this tendency is indicative of the intrinsic complexity of the representation of extreme phenotypes by linear approximations and it indicates that there may exist nonlinear or threshold-like biological processes which are not well represented by the given feature representation.

4.4.2 Residual Distribution

These observations are also supported by the distribution of the residuals as depicted in the residual histogram. It is unimodal but concentrated around zero with negative skewness, which is in line with the residual-versus-predicted plot. This is a typical asymmetry that is reported in pharmacogenomic modeling studies and does not nullify the model; instead, it can define the boundaries of linear predictive models.

Most of the residuals are bounded (approximately within ± 0.25) implicates that the predictions of the model are not usually ill-calibrated, and extreme errors are very rare.

4.5 Sparsity and Effective Dimensionality

One of the biomathematical findings made by the ElasticNet sparsity analysis is that only 278 of the 958 coefficients are non-zero which covers about 29 percent of the feature space. The outcome of this finding is that the prediction of drug responses is done on a lower-dimensional effective manifold embedded in the high-dimensional space of molecular features.

Mathematically, the ElasticNet solution can be interpreted as identifying a sparse vector $\beta \in \mathbb{R}^{958}$ with support size $|\text{supp}(\beta)| \ll 958$. This confirms the hypothesis that biological drug response can be controlled by a small set of strong molecular determinants, which is in line with systems biology and network control views.

This interpretation is further reinforced by the similar or better performance of sparse linear models as compared to more complicated nonlinear models.

4.6 Stability Analysis and Robustness

Bootstrap resampling of the test set was done to evaluate the strength of the reported performance. In 100 bootstrap replicates, the performance measures had small confidence intervals, with RMSE and MAE having the smallest variation and correlation measures being positive throughout.

The analysis of stability validates the fact that the observed results are not a result of some specific test split, but rather indicative of systematic structural characteristics of the data and models. Statistically, this increases the degree of confidence with regard to the generalizability of the findings in the benchmark environment.

5. Discussion

The current paper will present a biomathematical evaluation of machine-learning models to predict drug response in a standardized digital health framework. Instead of focusing on algorithmic novelty or maximum predictive performance, the results focus on model structure, sparsity, stability and error behavior, which are fundamental issues in biomathematics and systems-level healthcare modeling. The results provide some information on the mathematical essence of drug response prediction and its effects on digital health systems.

One of the key findings of this paper is that regularized linear models, especially Lasso and ElasticNet, have similar or better predictive ability than nonlinear models, according to RMSE and rank-based correlation values. This is consistent with the previous findings that the complexity of models is not always associated with the corresponding improvement of biomedical prediction tasks, particularly in situations where the data are noisy and high-dimensional (Zhang et al., 2023). Biomathematically, it implies that the mapping between molecular features to drug response can be well modeled by a linear mapping with suitable constraints to the hypothesis of a low-dimensional biological structure existing.

The sparsity analysis gives another testimony to this interpretation. ElasticNet regression finds about 29 percent of the features to be active predictors, which shows that the response to drugs is controlled by a lower effective dimensionality in comparison to the nominal feature space. Such outcome agrees with the systems biology considerations, including the statement that biological responses are frequently governed by a few key pathways or regulatory modules (Živanović &

Filipović, 2024). Mathematically, the sparse solution is associated with a constrained optimization problem where the support of the coefficient vector is much less than the ambient dimension, which supports the usefulness of ℓ_1 -regularization in biomolecular modeling.

Residual diagnostics can provide further information about the weaknesses and soundness of the linear modeling framework. The residual-versus-predicted plots show that the errors are concentrated around the zero value, and the heteroscedasticity is not very high, which means that the models are not highly biased. Nevertheless, the existence of a heavier negative tail of the residual distribution indicates a consistent underestimation of the highly resistant drug cell line pairs. This has been observed in earlier research on the prediction of drug responses and is an indication of the difficulty of modelling extreme phenotypes with linear approximations (Firoozbakht et al., 2022). The biological interpretation of this can be nonlinear resistance to this and threshold effects which are poorly represented in the feature representation.

The average although consistent Spearman correlations found between models point to a significant difference between absolute accuracy and biological relevance. In pharmacology the relative order of drug sensitivity is frequently more useful than accurate numerical forecasts. The postbranch performance advantage of sparse and robust linear models argues in support of the hypothesis that trustworthiness in selecting features and outlier resistance improves violates monotonicity at control should arise between forecasting and measuring (Hassan et al., 2024).

The reliability of the results is also supported by bootstrap stability analysis. The small confidence intervals of the RMSE, MAE, and correlation measures show that the observed performance is not a result of a specific test split, but rather it is a characteristic of the data and models that are not changing because of the specific split. Statistically, such robustness is essential to digital health an application, in which predictive models are required to work across people with heterogeneous biological systems, and under different experimental conditions.

In comparison with the literature, the results of the current study support an emerging body of literature that parsimonious, mathematically constrained models are very competitive in biomedical prediction exercises. Although there is evidence of promising performance with deep learning and hybrid schemes in specific environments (Hajim et al., 2024), the robustness of deep learning suffers in coincidence with standardized performance instead of where interpretability and reliability of the results are important. The current findings are a complement to these observations, giving empirical support to the notion that classical machine-learning models can be useful in understanding the information provided when interpreted with a biomathematical perspective.

These findings have implications taking over methodological considerations. Sparsity-promoting linear models have practical benefits in digital health systems where transparency, reproducibility and regulatory acceptance are the key characteristics. Their coefficients may be understood as weighted contributions of molecular features, which makes it easy to generate hypotheses and conduct subsequent biological analysis. In addition, they are mathematically simple and allow defining uncertainty and error propagation more clearly, which is consistent with explainable AI principles promoted in healthcare studies (Allen, 2024).

Although such contributions have been made, there are a number of limitations that need to be mentioned. First, the analysis is limited to in vitro data of drug to cell lines, which do not necessarily reflect the complexity of in vivo or patient level responses. Second, the feature representations are latent and preprocessed and cannot be directly interpreted biologically on the level of genes or pathways. Third, monotherapy drug response set of interest precludes any combinatorial effects and time dependence, which is becoming important in precision medicine. Lastly, although linear models can work well given the existing benchmark, they might not be able to capture higher order interactions in richer datasets.

Such constraints indicate a number of future research directions. Adding mechanistic models or pathway-based constraints would make it more interpretable, and have retained mathematical structure. It would be more informative to extend the analysis to combination therapies and longitudinal response data to have a more comprehensive picture of drug dynamics. Also, hybrid structures involving sparse linear representation and selective nonlinear block certainly can be proposed as the means of a balanced approach that entails represent other major linear influences and enables nonlinear residual effects (Ramaswamy & Keidar, 2023).

To summarize, this paper has shown that a biomathematical model assessment of machine learning models provides useful information about the design and constraints of drug response prediction. The findings, with their focus on sparsity, stability, and error analysis, can be used to develop transparent and reliable computational tools to support the development of digital health systems, in addition to current progress in data-driven healthcare modeling.

6. Conclusion

This paper has provided a biomathematical evaluation of machine-learning models to predict the response of drugs within a standardized digital health environment. The work systematically tested linear, sparsity-constrained, robust and nonlinear regression models under the same experimental conditions using a cross study pharmacogenomic benchmark.

The findings provide that regularized linear models, especially Lasso and ElasticNet can have predictive performance equivalent or superior to more complicated nonlinear methods. Notably, sparsity analysis showed that only a small set of molecular features is useful to make predictions, which is consistent with the hypothesis that drug response is determined by a smaller effective dimensionality, which could be embedded in a high-dimensional feature space. In practical terms, the findings have a lot of implications on digital health systems. The mathematical clarity and competitiveness of sparsity-promoting models highlight their advantage of making them useful in decision pipelines embracing interpretability, reproducibility, and stability requirements. Additional residual diagnostics and bootstrap results also suggest that the achieved performance is sound against sampling variability further supporting the belief in the soundness of these models to compute drug-response assessment and generate hypothesis. It is on these findings that some recommendations can be made. To begin with, sparsity-constrained linear models are recommended to researchers and practitioners as good baseline methods of predicting drug responses, especially in cases where large-dimensional molecular data is at work. Second, model analysis must go beyond the point estimates of accuracy to encompass residual analysis and quantification of uncertainty, making sure that the conclusions made about predictions are mathematically and statistically warranted. Third, such models need to be incorporated into digital health systems with transparency and interpretability of errors, as well as performance, as a top priority. Additional directions of research involve furthering the present system towards combinatorial therapies and longitudinal response measurements, pathway informed or mechanistic constraints to develop biological comprehensibility, and hybrid mode of modeling approaches that can freely choose nonlinear elements. The combination of these initiatives can also contribute to the creation of biomathematically-based, reliable machine-learning-based digital health applications.

7. References

- AL-Rajabi, M. M., Alzyod, S., Patel, A., & Teow, Y. H. (2025). A hybrid machine learning framework for predicting drug-release profiles, kinetics, and mechanisms of temperature-responsive hydrogels. *Polymer Bulletin*, 82(8), 2911-2932.
- Al-Zirkan, K., Alawadi, A. A., & Fenjan, M. N. (2025). The Correlation Between Insulin Levels and Diabetic Foot in Type 2 Diabetes Mellitus. *Letters in biomathematics*, 12(2).
- Allen, B. (2024). The promise of explainable AI in digital health for precision medicine: a systematic review. *Journal of personalized medicine*, 14(3), 277.
- Chatterjee, S., Fruhling, A., Kotiadis, K., & Gartner, D. (2024). Towards new frontiers of healthcare systems research using artificial intelligence and generative AI. In (Vol. 13, pp. 263-273): Taylor & Francis.
- Firoozbakht, F., Yousefi, B., & Schwikowski, B. (2022). An overview of machine learning methods for monotherapy drug response prediction. *Briefings in bioinformatics*, 23(1), bbab408.
- Hajim, W. I., Zainudin, S., Daud, K. M., & Alheeti, K. (2024). Optimized models and deep learning methods for drug response prediction in cancer treatments: a review. *PeerJ Computer Science*, 10, e1903.
- Hakami, M. A. (2024). Harnessing machine learning potential for personalised drug design and overcoming drug resistance. *Journal of Drug Targeting*, 32(8), 918-930.
- Hassan, J., Saeed, S. M., Deka, L., Uddin, M. J., & Das, D. B. (2024). Applications of machine learning (ML) and mathematical modeling (MM) in healthcare with special focus on cancer prognosis and anticancer therapy: current status and challenges. *Pharmaceutics*, 16(2), 260.
- Hassan, M., Awan, F. M., Naz, A., deAndrés-Galiana, E. J., Alvarez, O., Cernea, A., Fernández-Brillet, L., Fernández-Martínez, J. L., & Kloczkowski, A. (2022). Innovations in genomics and big data analytics for personalized medicine and health care: a review. *International journal of molecular Sciences*, 23(9), 4645.
- Hider, M. A., Nasiruddin, M., & Al Mukaddim, A. (2024). Early disease detection through advanced machine learning techniques: a comprehensive analysis and implementation in healthcare systems. *Revista de Inteligencia Artificial en Medicina*, 15(1), 1010-1042.
- Liu, Y., Wu, R., & Yang, A. (2023). Research on medical problems based on mathematical models. *Mathematics*, 11(13), 2842.
- Praveen, S. P., Murali Krishna, T. B., Anuradha, C., Mandalapu, S. R., Sarala, P., & Sindhura, S. (2022). RETRACTED ARTICLE: A robust framework for handling health care information based on machine learning and big data engineering

techniques. *International Journal of Healthcare Management*, 1-18.

- Ramaswamy, V. D., & Keidar, M. (2023). Personalized plasma medicine for cancer: transforming treatment strategies with mathematical modeling and machine learning approaches. *Applied Sciences*, 14(1), 355.
- Rashidi, H. H., Pantanowitz, J., Hanna, M. G., Tafti, A. P., Sanghani, P., Buchinsky, A., Fennell, B., Deebajah, M., Wheeler, S., & Pearce, T. (2025). Introduction to artificial intelligence and machine learning in pathology and medicine: generative and nongenerative artificial intelligence basics. *Modern Pathology*, 38(4), 100688.
- Sabherwal, A. K., Sood, A., & Shah, M. A. (2024). Evaluating mathematical models for predicting the transmission of COVID-19 and its variants towards sustainable health and well-being. *Discover Sustainability*, 5(1), 38.
- Stergiou, K. D., Minopoulos, G. M., Memos, V. A., Stergiou, C. L., Koidou, M. P., & Psannis, K. E. (2022). A machine learning-based model for epidemic forecasting and faster drug discovery. *Applied Sciences*, 12(21), 10766.
- Uddin, M., Islam, A., Saha, R., & Goswami, D. (2024). The role of machine learning in transforming healthcare: A systematic review. *Journal of Business Intelligence and Management Information Systems Research*, 1(01), 01-16.
- Woodman, R. J., & Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. *Aging clinical and experimental research*, 35(11), 2363-2397.
- Zhang, B., Shi, H., & Wang, H. (2023). Machine learning and AI in cancer prognosis, prediction, and treatment selection: a critical approach. *Journal of multidisciplinary healthcare*, 1779-1791.
- Živanović, M. N., & Filipović, N. (2024). System biology Modeling for drug optimization. In *In Silico Clinical Trials for Cardiovascular Disease: A Finite Element and Machine Learning Approach* (pp. 105-137). Springer.