Predicting Drug Response on Multi-Omics Data Using a Hybrid of Bayesian Ridge Regression with Deep Forest

Talal Almutiri[1], Khalid Alomar[2], Nofe Alganmi[3]
Department of Information Systems[1, 2], Department of Computer Science[3]
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia[1, 2, 3]

***Abstract*—An accurate drug response prediction for each patient is critical in personalized medicine. However, numerous studies that relied on single-omics datasets continue to have limitations. In addition, the curse of dimensionality considers a challenge to drug response prediction. Deep learning has remarkable prediction effectiveness compared to traditional machine learning, but it requires enormous amounts of training data which is a limitation because the nature of most biological data is small-scale. This paper presents an approach that combines Bayesian Ridge Regression with Deep Forest. BRR relies on the Bayesian approach, in which linear model estimation occurs based on probability distributions rather than point estimates. It was utilized to integrate multi-omics, a feature selection that calculates the coefficient as the feature importance. DF reduces the computational cost and hyper-parameter tuning cost. The Cancer Cell Line Encyclopedia CCLE was used as a dataset to integrate the gene expression, copy number variant, and single nucleotide variant. Root Mean Square Error, Pearson Correlation Coefficient, and the coefficient of determination were used as the evaluation metrics. The obtained findings show that the proposed model outperforms Random Forest and Convolutional Neural Network regarding regression performance; it achieved 0.175 for RMSE, 0.842 for PCC, and 0.708 for R2.**

Keywords*—Bayesian ridge regression; deep forest; deep learning; drug response prediction; machine learning; multi-omics data

I. Introduction

Personalized medicine is a cancer therapy method that aims to find the most effective therapeutic solutions for each patient. The combination of genetic and drug-sensitivity data and the subsequent creation of drug-response associations allows for this discovery [1]. While personalized medicine is not yet utilized as a regular treatment, it is possible for most cancer patients due to the progress made in multi-omics features and drug-sensitivity testing [2]. Personalized treatment regimens based on genetics are one of the primary aims of systems medicine [3]. For the development of individualized cancer therapy treatments with a projected efficacy much above existing standard-of-care methods, the inferred models' ability to correctly forecast a tumor's responsiveness to a medicine or drug combination might benefit that process [3]. However, there has been a lack of progress in cancer treatment based on single-omics datasets such as those generated by the Human Genome Project and the early genomic profiling of the Cancer Genome Atlas (TCGA) projects [4]. The multi-omics analysis that has gained prominence in cancer research over the last several decades may be the only way to get a comprehensive view of cancer behavior and uncover new therapeutic vulnerabilities [5].

Moreover, the "curse of dimensionality" or large p, small n, is one of the most challenging issues in drug response prediction and when dealing with omics data in general [6] in other words, having many features p. However, only a few available data instances n create a particular barrier to using early concatenation in multi-omics integration. For instance, the human genome has more than 20,000 protein-coding genes, a significant number. As a result of this integration, multi-omics datasets may easily contain more than 50,000 attributes when the genome, proteome, and transcriptome are all included. Regarding cancer data, the number of available tumor samples in a dataset is usually restricted, with cancer cohorts typically consisting of only a few hundred patient samples [5]. As a result, the features must be reduced through feature selection [7]. Feature selection works to identify the relevance of features and selects a collection of features or attributes based on a particular assessment criterion [8].

Despite the power of deep neural networks power, it appears to have drawbacks [9]. To begin with, it is noted that deep neural networks require enormous amounts (large-scale data) of training data are often needed, making them inapplicable to jobs having only small-scale data. Due to the high cost of class annotation, many real-world tasks currently lack adequate labeled data [9], [10], resulting in the poor effectiveness of deep neural networks in relation to those tasks [11]. Additionally, the success of deep learning is strongly dependent on carefully tuning several hyper-parameters [12]. Consequently, Zhou [11] introduced gcForest or Deep Forest (DF), which integrates multi-Grained and Cascade Forest as a deep learning alternative. gcForest is a new decision tree ensemble technique approach that outperforms deep learning across various applications.

This paper introduced a method that combines Bayesian Ridge Regression (BRR) with Deep Forest (DF) called (BRR-DF). BRR was used to integrate multi-omics which was utilized as a feature selection method that calculates the coefficient to determine the feature importance score. BRR relies on the Bayesian approach, in which linear model estimation occurs based on probability distributions rather than point estimates. In addition to the model parameters also coming from the distribution, the response is also generated from the probability distribution. The training inputs and outputs affect the posterior probability of the model parameters.

Furthermore, Ridge was embedded with Bayesian regression to reduce model complexity and multicollinearity by shrinking the coefficients. Most omics data is considered to be small-scale data; therefore, Bayesian is suitable for these cases. It integrates the prior knowledge of the parameter (prior parameter distribution) with the observed data. DF integrates Multi-Grained and Cascade Forest, which effectively capture the local features. The cascade forests utilize a network structure inspired by a multi-layer artificial neural network to continuously improve results. DF was suggested as a deep learning alternative to reduce complexity time in hyper-parameter tuning that otherwise causes a computational cost in deep learning models.

The contribution of the proposed method can be summarized in the following two points:

- Using BRR as a feature selection method for integrating gene expression, copy number variant, and single nucleotide variant to improve drug response prediction. BRR can handle inadequate data or skewed distributed data by modeling linear regression models using probability distributions. By utilizing BRR, we aim to identify the most informative features from multiple genomic data sources and improve the accuracy of drug response prediction.
- Using DF to reduce the computational cost of hyper-parameter tuning; also, when the inputs have a high degree of complexity or dimensionality, DF can boost its representational learning capabilities to improve prediction.

The rest of this paper is organized as follows. Section 2 introduced the related work. Section 3 presented the methods and materials used in this study including datasets, framework of BBR-DP, and evaluation metrics. Section 4 elaborated results and discussion. In Section 5, the conclusion and future work were presented.

II. Related Work

The current studies have introduced various machine learning techniques for predicting drug sensitivity and discovering biomarkers affecting drug response. Examples of these techniques are Support Vector Machines (SVMs) [13], Graph Networks [14], [15], Bayesian multitask multiple kernel learning [16], [17], Random Forest (RF) [19–22], and Neural Network [22] models. However, there is still a significant opportunity to improve prediction effectiveness and model generalizability regarding these computational models. Deep Learning (DL) has also been employed successfully in other drug discovery-related tasks. The prediction effectiveness of Deep Learning algorithms is comparable to, if not better, than that of the approaches for the bulk of these tasks [23]. Numerous deep learning-based techniques for drug response prediction have been proven to be successful, including DeepProfile [24], CDRscan [25], DeepCDR [14], DeepDSC [26], and GraphDPR [15].

A Bayesian ridge regression-based approach (B-GEX) was developed by Wenjian et al. [27] to infer the gene expression profiles of various organs from blood gene expression profiles. A low-dimensional feature vector was derived from the complete blood gene expression profile using feature selection for each gene in a tissue. To train the inference models to capture the cross-tissue expression correlations between each target gene in tissue and its preselected feature genes in peripheral blood, they used The Genotype-Tissue Expression (GTEx) RNA sequencing (RNA-seq) data of 16 tissues.

Velten and Huber [28] proposed a method for guiding penalization in regression using information from external covariates. Their method penalizes the feature groups defined by the covariates differentially and adjusts the relative power of penalization according to the information content of each group. Their procedure combines shrinkage with feature selection and provides a scalable optimization scheme using techniques from the Bayesian tool set. The method accurately retrieves each feature group's accurate effect sizes and sparsity patterns in simulations. They evaluated the performance of their method for drug response prediction using leukemia data. Prediction performance improves when the groups' dynamic ranges differ significantly.

Sharifi-Noghabi et al. [29] utilized deep learning to develop a method called MOLI (multi-omics late integration). MOLI integrates gene expression data, copy number alterations, and somatic mutation. Their model learns features for each omics data type by encoding subnetworks particular to it. MOLI is the first end-to-end late integration approach using deep learning that combines a "triplet loss function" and a "binary cross-entropy" to improve this representation. Responder cell lines are more comparable and distinct from non-responder cell lines, while the half maximal inhibitory concentration (IC50) values predicted by this depiction are more accurate.

Malik et al. [30] proposed a late multi-omics integration framework for robustly quantifying survival and drug response in breast cancer patients, emphasizing the relative predictive ability of the available omics datatypes. A supervised feature selection algorithm, neighborhood component analysis (NCA), was used to select the relevant features from the multi-omics datasets retrieved from The Cancer Genome Atlas (TCGA) and Genomics of Drug Sensitivity in Cancer (GDSC) databases.

A Deep Forest architecture, first presented by Zhou [31], was used by Su et al. [32] to develop the Deep-Resp-Forest anti-cancer drug response prediction model, which classifies the anti-cancer drug response as either sensitive or resistant. In Zhou et al.'s work, the Deep Forest, known as gcForest, was a cascade of forests. Su et al. achieved remarkable results when their model was tested against the Cancer Cell Line Encyclopedia (CCLE) and the Genomics of Drug Sensitivity in Cancer (GDSC). As they mentioned, regression is preferred for more accurate results.

Table I shows some recent studies that were focused on applying ML and DL methods in drug response prediction by focusing on methods/techniques, contributions/advantages, and limitations/disadvantages.

TABLE I. SUMMARY OF SOME RECENT STUDIES IN DRUG RESPONSE PREDICTION AND THEIR CONTRIBUTIONS AND LIMITATIONS

Resource and Year	Methods/ Techniques	Merits, Contribution Advantages	Demerits and limitations Disadvantages	Datasets
Sharifi-Noghabi et al. [29], 2019	DNN Ridge regression	They indicated that MOLI outperforms early integration multi-omics and single-omics techniques. They mentioned it was the first strategy to employ pan-drug transfer learning for targeted drugs, and it improved prediction effectiveness relative to drug-specific inputs.	Although their research only employed DNA mutation, CNA, and gene expression profile, MOLI may be expanded to include other omics information and drug chemical structure. While they only explored the triplet loss for improving the concatenated representation, they observed that similar losses such as the contrastive loss function employed in the Siamese network [33] can be utilized instead. All utilized datasets have substantially skewed or imbalanced class distributions due to the few number of respondents' vs non-responders. They solved that by oversampling minorities. However, this method typically leads to overfitting, especially for deep neural networks.	GDSC PDX TCGA
Liu et al. [14], 2020	UGCN CNN	Insufficient or imbalanced training examples can be supplemented with the proposed UGCN by random selection of multiple complementary graphs for each medication. In the classification task, they randomized the feature matrix, connected complementary networks at random, and positive training examples were augmented five times. DeepCDR may be utilized with molecular generation processes. Existing chemical generation models based on the Recurrent neural network (RNN) technique[66], generative adversarial networks (GANs) [34] and deep reinforcement learning [35] Concentrate on broad chemicals while ignoring the characteristics of specific cancer cells. Methods for cancer-specific or disease-specific innovative drug design may be presented by employing DeepCDR predicted CDR as prior knowledge or a reward score for driving chemical production.	Top DL algorithms like DeepCDR and GraphDPR [15] perform better. For a drug-blind test, the examined deep learning approaches do not act as well as the SRMF, a matrix factorization-based method. To improve DL methods for predicting drug reactions, obtaining differentiating information from drug profiles is critical. Either create novel drug target fingerprint systems or use sophisticated "graph neural networks" to extract latent properties from drug data [36]. In future work, researchers can use huge amounts of omics data analyzed before and after treatment to determine how the tested drugs affect their molecular profiles.	GDSC CCLE TCGA
Jia et al.[37], 2021	VAE Elastic Net PCC PCA	Accurate drug sensitivity data prediction in cancer samples would allow recapitalization of recognized and new biomarkers, which are commonly missing owing to cell line methods or limitation of sample size. Their categorization of chemicals by reaction profiles showed distinct groupings and signatures. Using TCGA data, they discovered a link between medicines and TMB that was previously infeasible using cell line models. To find pan-cancer genomic markers, they explored DNA mutations, CNVs, and gene expression. The positive correlations between AZD6244 and the earlier published 18-gene signature demonstrated how their results are robust.	For some drugs, including LBW242, couldn't enhance prediction accuracy by fitting models. The model-fitting parameters of VAE-based models could not compete with PCA methods for several drugs (in-sample PCC and holdout R2). However, given insufficient data, the PCA-based model for paclitaxel failed to distinguish between pCR and non-pCR patients. So future validation is necessary to validate these prediction models. Moreover, while certain drugs had good prediction results in the cell line method, their response in cancer examples was variable. So studying drug response in cancer samples is substantially more difficult and involves various contexts and variables.	GDSC CCLE TCGA
Pouryahya et al. [38], 2022	Wasserstein distance Spearman's correlation PCC Hierarchical clustering Random forest regression	Using the optimal mass transport (OMT) theory and unsupervised and supervised ML models in conjunction with the CDCN model, they were able to show that random forest approaches in the consequent distinct pairs of cell-line and drug clusters can deliver more satisfactory predictive ability than the CDCN model used in previous studies. Using Wasserstein distances, which are calculated between invariant measurements of gene expression patterns, the researchers discovered that cell lines that were comparable in terms of Wasserstein distances responded similarly to (structurally identical) medicines.	In the clustering of drugs, unsupervised removal of strongly correlated cheminformatic features while maintaining non-redundant informative features. Despite the elimination of this feature, their strategy outperformed other approaches in terms of predictive power. Using mutation, CNV, and hyper-methylation data may enhance prediction results or provide new findings.	GDSC HPRD CCLP PubChem OncoKB
Wang et al. [39] 2023	GCNs AEs	In order to overcome some of the shortcomings of recent studies, including ignoring the correlation between drug cell line pairs (DCPs), the GADRP was developed. Additionally, the issue of over-smoothing, in which the representation of each node becomes more similar as the number of layers grows, was not considered in recent research that used GCNs. So they built a sparse drug cell line pair (DCP) network incorporating data on drug, cell line, and DCP similarity before using a stacked deep AE to extract low-dimensional representations from cell line attributes. Later, to learn DCP features, initial residual and layer attention based GCN (ILGCN), which can resolve over-smoothing issues, was used. Finally, the prediction was performed using a fully connected network.	First, ILGCN can only be regulated within five levels due to the scale of the DCP network and the constraints of computer storage capacity. Second, the GADRP deep learning model lacks biological entities like targets and disorders, which contribute to its level of inexplicability. Consideration should be given to including more entities and associations in cancer medication response prediction. Additionally, despite GADRP's potent prediction capabilities, its use in the clinic remains a significant issue because it is trained using in vitro data.	PubChem PRISM CCLE

GDSC: Genomics in Drug Sensitivity in Cancer
UGCN: Uniform Graph Convolutional Network
PDX: Patient-Derived tumor Xenograft
CNN: Convolutional Neural Network
CCLE: Cancer Cell Line Encyclopedia
VAE: Variational Autoencoder
TCGA: The Cancer Genome Atlas
PCC: Pearson Correlation Coefficient
CCLP: COSMIC Cell Line Project
PCA: Principal Component Analysis
HPRD: Protein Reference Database (HPRD)
GCN: Graph Convolutional Network
OncoKB: Precision Oncology Knowledge Base (OncoKB)
AE:Autoencoder
DNN: Deep Neural Network

Therefore, the drug response methods showed remarkable results when multi-omics were integrated. However, integration causes a curse of dimensionality which negatively affects prediction. In addition, multi-omics data is small-scale data, which needs a method to handle inadequate data or skewed distribution.

III. Methods and Materials

The proposed solution works to reduce dimensionality and integrate the three omics, before using Deep Forest to improve the drug response prediction. The solution consists of four phases: datasets preparation, integrating multi-omics using Bayesian Ridge Regression, the Deep Forest phase, and the evaluation phase. The general framework is shown in Fig. 1; more details for each phase are discussed in the following points. Each single omics was processed independently; Bayesian Ridge Regression was utilized for each single data type.

A. Datasets

More than 1000 human cancer cell lines were gathered and molecularly described in the Cancer Cell Line Encyclopedia (CCLE) project [40] that has acquired and molecularly characterized over 1000 human cancer cell lines. The investigation discovered 24 anti-cancer drug sensitivity profiles among 504 cell lines. The CCEL [21], [40] dataset was used in this research. The half-maximal inhibitory concentration IC50 was used as the drug response for cell lines across the drugs (denoted by $y_{res,c}$) c for a cell line. Three omics were used, including single-nucleotide mutation (denoted by $x_{snv,g}$) g for gene, gene expression (denoted by $x_{exp,g}$), and copy number alternation/variation (denoted by $x_{cnv,g}$) Gene expression and copy number alternation are real values, the single-nucleotide mutation use binary values, "1" used for mutation and "0" for wild type. There are no missing values in the gene expression data. For copy number alternation and single-nucleotide mutation, rows with more than half of the cells missing values were removed. The mean weight approach was used to compensate for the missing values for the remaining cell lines.

The distance was calculated to select the nearest k, which was used to impute the gene expression missing value, defined as follows:

$$dis(c,k) = \left\| x_{exp,c} - x_{exp,k} \right\|_2^2 \quad (1)$$

where c is the cell line, k is the nearest cell line, and x is the gene expression value for each cell line.

The mean value of the nearest cell lines was used to impute the missing value of cell line c in copy number alteration of genes g.

$$misCNV(c,g) = \sum_{k=1}^{K} \frac{dis(c,c_k)}{\sum_{k=1}^{K} dis(c,c_k)} \; misCNV(c_k,g) \quad (2)$$

The values of the single-nucleotide mutation features are binary, with 1 indicating mutation and 0 indicating wild type. The mean feature value for cell line c among the k-nearest cell lines was used to compensate for the missing SNV (single-nucleotide mutation or variation) value of gene g as follows:

$$misSNV(c,g) = \begin{cases} 1 \; if \; \left(\sum_{k=1}^{K} misSNV(c_k,g) > \sum_{k=1}^{K} (1 - misSNV(c_k,g)) \right) \\ 0 \; otherwise \end{cases} \quad (3)$$

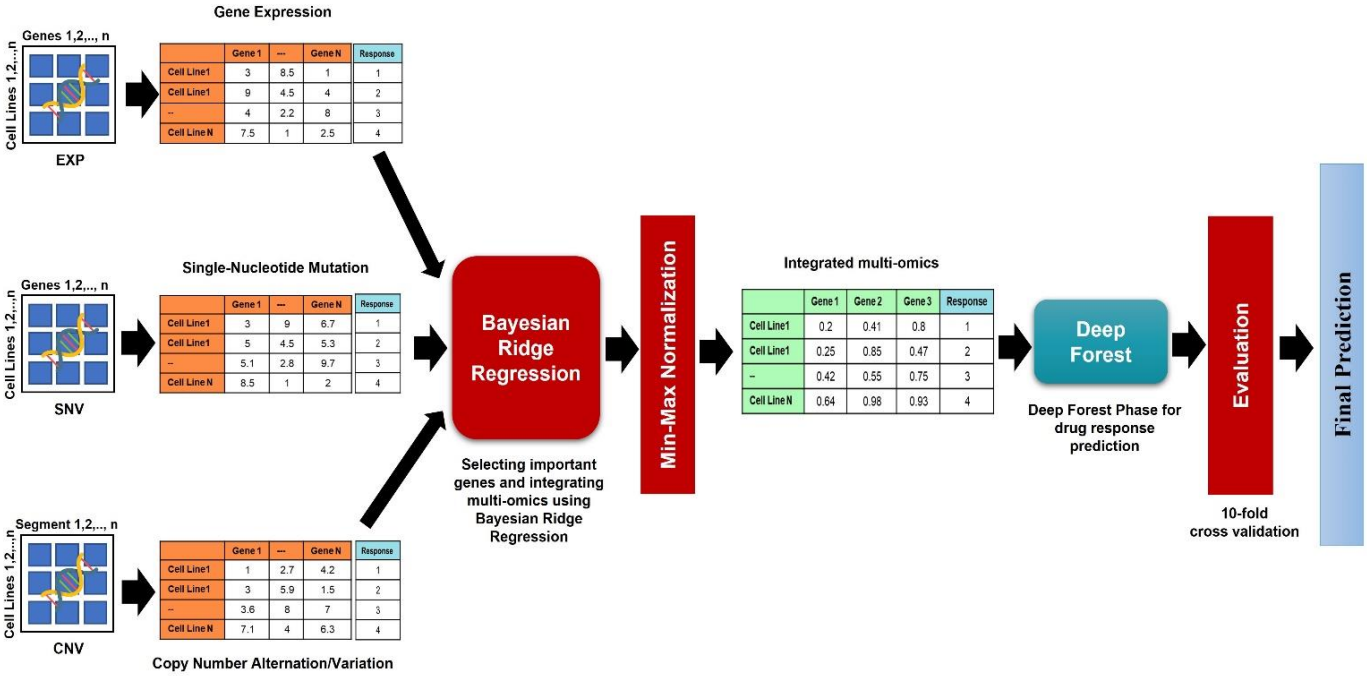

Fig. 1. The framework of BBR-DP to integrate multi-omics data for predicting drug response.

In a similar way, the missing value of IC50 was imputed in the same way as the copy number alteration manner. The mean value of the nearest cell lines was used to impute the missing value of cell line c.

$$misIC50(c,g) = \sum_{k=1}^{K} \frac{dis(c,c_k)}{\sum_{k=1}^{K} dis(c,c_k)} \; misIC50(c_k,g) \quad (4)$$

K=10 was selected for preparing the CCEL dataset.

The IC50 matrix was converted to a tabular form which had 8712 rows, then all of the drug responses of each cell line were grouped by the mean for each cell line to be considered as the drug response that needed to be predicted in a regression problem. The final total samples were 363 for the IC50 data. Table II shows the total number of samples for the CCEL data.

TABLE II. THE TOTAL NUMBER OF SAMPLES FOR THE CCEL DATA

Type	Raw Data	After Preprocessing
Cell Lines	1061	363
Drugs	24	24
Gene expression	20049 (1061)	19,389 (363)
Single-nucleotide mutation	1667 (1061)	1667 (363)
Copy number alteration	24960 (1061)	24960 (363)

B. Integrating Multi-Omics Bayesian Ridge Regression

Bayesian Ridge Regression (BRR) was used as the feature selection method to reduce dimensionality and integrate multi-omics. This method, which is based on a Bayesian approach, is concerned with selecting subsets of the independent variables in linear regression to predict a response variable. The response variable is first assigned a probability distribution via the specification of a family of prior distributions for the unknown parameters in the regression model. However, because the data influence this family's ultimate choice of the prior distribution, the independent variables are assumed to be distinct observables, and the corresponding regression coefficients are assigned independent prior distributions [41]. BRR fits a model where the weighted sum of the independent variables can predict the response variable. It works by determining a set of coefficients to utilize in the weighted sum to perform a prediction. These coefficients were used as feature importance scores to select the best features of each single omics data.

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \varepsilon \quad (5)$$

where y is the dependent variable (also known as the response variable) β is the coefficient or model parameter, x is the value of a predictor variable, and there is also an error term describing the effect of variables not included in a model or random sampling noise.

From a Bayesian perspective, probability distributions rather than point estimates are used to build linear regression. The response, y, should be chosen from a probability distribution rather than evaluated as a single number. The goal of Bayesian Linear Regression is to ascertain the posterior distribution for the model parameters rather than to identify the one 'best' value of the model parameters. In addition to the model parameters also coming from a distribution, the response is also generated from a probability distribution. The training inputs and outputs affect the posterior probability of the model parameters [42], [43]. However, Ridge was embedded with Bayesian regression to reduce the model complexity and multicollinearity by shrinking the coefficients. Most omics data is considered to be small-scale; therefore, Bayesian is suitable for these kinds of cases. It integrates the prior knowledge of the parameter (prior parameter distribution) with the observed data.

The three omics Exp, SNV, and CNV were tested as a single item of data and integrated in different combinations as follows: 1) EXP, 2) SNV, 3) CNV, 4) EXP and SNV, 5) EXP and CNV, 6) SNV and CNV, 7) EXP, SNV, and CNV. in addition, three experiments were implemented to evaluate the proposed solution and to study the multi-omics in various scenarios.

In the first scenario, the Baseline, all features of each single/multi omics were included in the model. It was implemented to test how the three omics affect the drug response without feature selection methods. In the second scenario, BRR was utilized to select the essential features of each single-cell omic - integrating them with the other one. According to the literature review, as a common practice, the mean value of all coefficients was used as a threshold to select features with coefficients higher than or equal to the calculated mean [44]. Also, the coefficients computed by the BRR were used to select important features for each single/multi-omics. In this last scenario, the top 10% of coefficients higher than or equal to the calculated mean were selected as informative features. After implementing various experiments for the different ratios, the ratio of the top 10% was selected, and it was noted that this 10% achieved the best results. This ratio was also used to reduce the computational cost of the model.

C. Drug Response Prediction Using Deep Forest

Deep Forest is a new ensemble Random Forest or decision tree approach that integrates multi-Grained Cascade Forest. This approach utilizes a cascade ensemble to create a deep forest as an alternative to deep learning that supports representation learning in gcForest. When the inputs have a high degree of complexity or dimensionality, multi-grained scanning can boost its representational learning capabilities, possibly helping gcForest to be contextually or structurally knowledgeable [10], [31]. gcForest allows a model complexity to be automatically defined, it performs very well even on small-scale data, and the number of cascade stages may be adjusted adaptively. Additionally, the developers/researchers can tailor their training expenses to their available computing resources. While deep neural networks have many hyper-parameters, gcForest has just a few. Its performance is relatively stable according to the hyper-parameter settings, it can achieve remarkable performance in most scenarios, even across datasets from diverse domains, by utilizing the default option. Through the use of external neural networks, gcForest may be trained and the theoretical analysis made more straightforward It is noted that ensemble methods or cascade trees are more accessible to analyze than deep learning [10], [45].

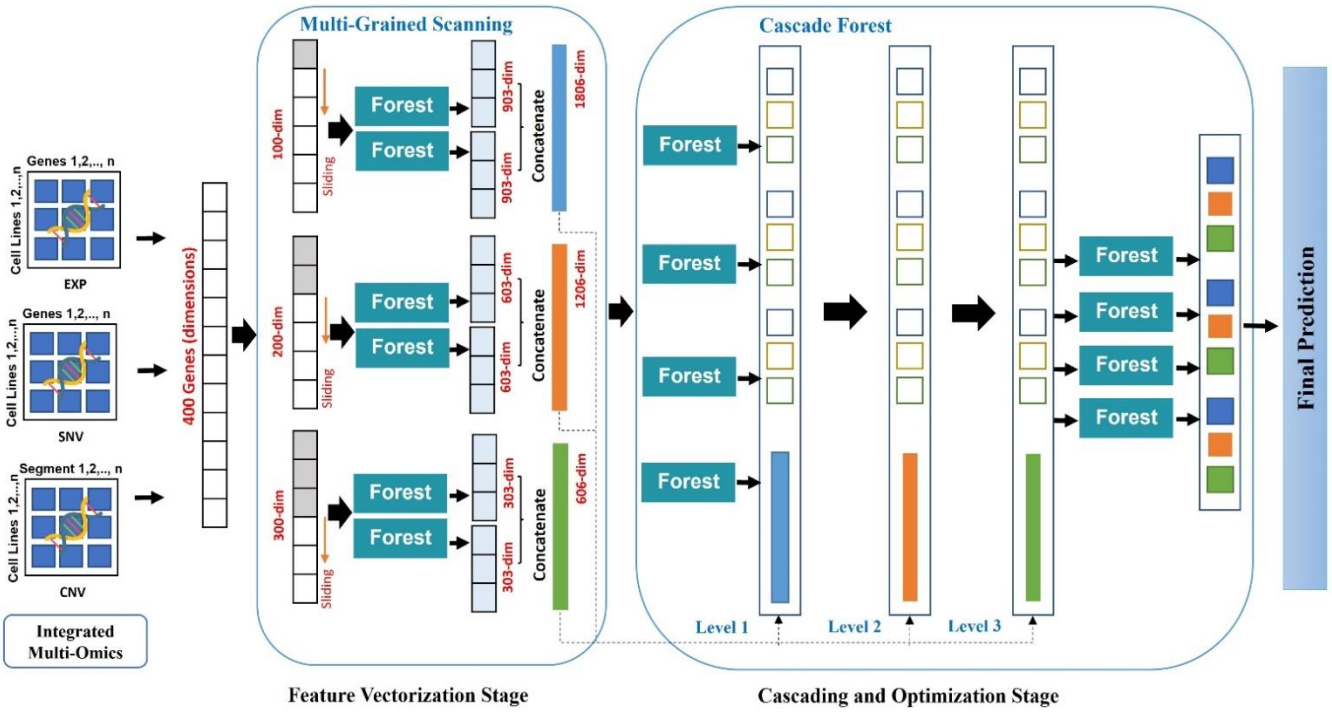

Fig. 2. The flowchart of BBR to integrate multi-omics data.

This phase consists of two stages, shown in Fig. 2: the Feature Vectorization Stage and Cascading and Optimization Stage. Deep Forest relies on multi-grained scanning, which effectively captures the local features. Also, the cascade forests utilize a network structure inspired by a multi-layer artificial neural network for continuously improving results.

1) Feature vectorization stage: Multi-grained scanning, motivated by the multi-convolution kernels used in Convolutional Neural Networks (CNNs), can discover and handle feature relationships in the subsequent cascade forests. the sliding window technique utilizes a scanning process to determine the local features and convert the raw data into a chain or set of low-dimensional local feature vectors [9], [31]. These low-dimensional vectors are then used to train a series of forests to get the class distributions for the input vectors.

Therefore, the raw features will be transformed into a high-dimensional feature vector using the multi-grained scanning, in this stage. Multi-Grained Scanning is utilized by sliding the windows to scan the original features and convert them into feature vectors. Suppose there is gene expression data as the sequence data; 400 raw features (dimensions) will be selected. Three sizes of sliding window will be used 100, 200, and 300. After scanning for a window size of 100 features, we will get 301 feature vectors according to this formula (total dimensions Nd - window size w)/ the stride of the sliding s + 1. The distance the window moves in each step is named the stride. Therefore, 400 - 100 / 1 + 1 = 301 feature vectors. A window with 100 dimensions will be generated, then 201 feature vectors will be processed for a window size of 200 and 101 feature vectors for a 300- dimensional. The final sample for training will be 903 instances for a window size of 100 features, one random forest, and three classes for prediction, as an example. If there are two random forests, the total number of samples will be 1806 instances when they are concatenated [31], [46].

2) Cascading and optimization stage: Cascade forests employ a network structure similar to a multi-layer artificial neural network, with each layer connected to the layer before it in the network hierarchy. It may be thought of as a collection of randomly generated forests that have been joined together [31]. Several random forests are used to construct each layer, and each decision tree inside a particular forest produces a drug response prediction independent of the others. In the following step, an overall drug response vector for the forest is created by taking the average of the drug responses provided by the decision trees in the forest. In the process of decomposition, the representation vector can be used as an input for the next cascade level in the process of decomposition. Processing is carried out in stages per each layer of the cascade, with each layer sending its results to the next and the processing results being passed on layer by layer until the prediction performance in the next level of the cascade does not increase [10].

D. Evaluation Metrics

The 10-fold cross-validation approach was utilized to evaluate the performance of the proposed solution. The mean of 10 iterations was recorded as the final result. Three evaluation metrics were used in this research, Root Mean Squared Error (RMSE), Pearson Correlation Coefficient (PCC), and the coefficient of determination R-Squared (R^2).

RMSE [25], [26] was utilized to measure the error, which is the difference between the actual drug response values and the predicted values. It is defined as:

$$RMSE = \frac{\sqrt{\sum(y_i - \tilde{y}_i)^2}}{N} \quad (6)$$

where y is the real value of drug response, $\tilde{y}_i$ is the predicted value of drug response, and N is the sample size.

The PCC [47] value was used to measure the degree of relationship or correlation between the drug response and predictors produced due to the multi-omics integration, defined as:

$$PCC = \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \quad (7)$$

where x_i is the value of the predictors, y_i is the drug response value, and $\bar{x}$ and $\bar{y}$ indicate the mean of the values.

R^2 [48]was used to measure how much of the variability of drug response can be explained by its relationship to the other predictors which could be formulated as:

$$R^2 = 1 - \frac{\sum(y_i - \tilde{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \quad (8)$$

where y is the actual value of drug response, and $\tilde{y}_i$ is the predicted value of drug response.

E. Comparison Criteria

The proposed Deep Forest was compared to Random Forest (RF) as a traditional machine learning method and CNN as a deep learning method. In addition, RF showed remarkable results according to various studies [19], [21], [49]. Furthermore, these algorithms were selected because CNN and an ensemble RF inspired Deep Forest. Therefore, they were employed to study which algorithm affects the Deep Forest the most.

IV. Results and Discussion

The performance of the proposed solution was demonstrated using three scenarios: baseline, coefficient higher than or equal to the mean, and the top 10% of coefficients higher than or equal to the mean.

A. Baseline Scenario

In this scenario, all features were included in the model. It was used as a baseline result for measuring the effectiveness of the suggested BRR method. Table III shows the results of the three methods for the baseline.

TABLE III. The Results of the Baseline Scenario for the CCEL Data

Type	Features	RMSE	R^2	PCC	Time (Sec)
Deep Forest (DF)					
EXP	19389	**0.196**	**0.639**	**0.8**	204.6
SNV	1667	0.238	0.002	0.029	28.1
CNV	24960	0.221	0.161	0.398	242.4
EXP, SNV	21056	0.196	0.624	0.79	138
EXP, CNV	44349	0.198	0.588	0.767	447
SNV, CNV	26627	0.222	0.159	0.396	328.8
EXP, SNV, CNV	46016	0.198	0.587	0.766	550.2
Random Forest (RF)					
EXP	19389	0.184	0.547	0.74	25.6
SNV	1667	0.237	0.016	0.119	0.859
CNV	24960	0.23	0.071	0.261	38.6
EXP, SNV	21056	**0.182**	**0.565**	**0.752**	27.6
EXP, CNV	44349	0.189	0.474	0.688	61.8
SNV, CNV	26627	0.223	0.131	0.357	38
EXP, SNV, CNV	46016	0.187	0.519	0.72	63.6
CNN					
EXP	19389	**0.212**	0.329	0.572	186
SNV	1667	0.252	0.004	0.047	21.4
CNV	24960	0.243	0.032	0.168	240
EXP, SNV	21056	0.206	0.383	0.618	195
EXP, CNV	44349	0.244	0.297	0.544	426
SNV, CNV	26627	0.237	0.05	0.217	251.4
EXP, SNV, CNV	46016	0.24	**0.407**	**0.637**	443.4

In general, DF showed the highest computational cost. Regarding R^2 and PCC, DF achieved the best results, and CNN showed the worst results. RF showed the lowest results from the perspective of RMSE and computational time. When multi-omics were integrated, DF achieved no effect, with EXP having the highest score. RF displayed a 4% improvement ratio when EXP and SNV were integrated. Also, CNN showed a 24% improvement ratio when EXP, SNV, and CNV were integrated.

Fig. 3 to 5 show the differences between the R2 training and testing results in the Baseline scenario. Those figures measure the differences between prediction results on training data compared to testing data in the term of R2. There was overfitting in the three models; this usually happens with small-scale data. The average ratio of overfitting for DF, RF, and CNN was 60%, 62%, and 72%, respectively. In which DF had the smallest differences between training and testing results and CNN showed the highest ratio. This means the models cannot be generalized and needs more data or other techniques to handle overfitting.

Fig. 3. R^2 results for the training and testing of the DF in the Baseline scenario.

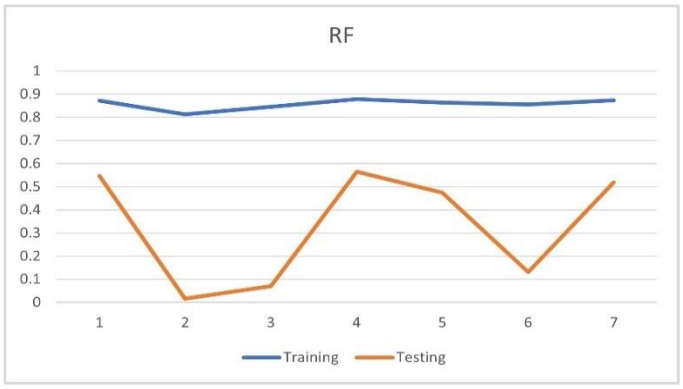

Fig. 4. R^2 results for the training and testing of the RF in the Baseline scenario.

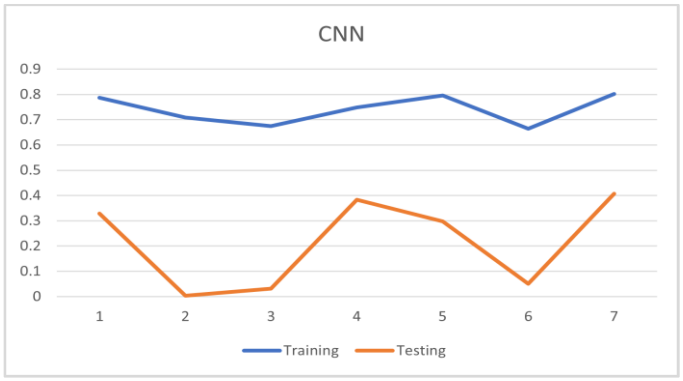

Fig. 5. R^2 results for the training and testing of the CNN in the Baseline scenario.

B. The Mean of Coefficients as a Threshold

In this scenario, the coefficients of each feature were calculated using BRR. Each feature was considered significant when its coefficient was higher than or equal to the mean of all coefficients. Table IV shows the results of this scenario.

TABLE IV. THE RESULTS OF THE MEAN OF COEFFICIENTS SCENARIO FOR THE CCEL DATA

Type	Features	RMSE	R^2	PCC	Time (Sec)
Deep Forest (DF)					
EXP	10204	0.199	0.594	0.771	89.4
SNV	783	0.241	0.012	-0.097	47.1
CNV	12538	0.226	0.095	0.304	130.8
EXP, SNV	10987	**0.197**	**0.643**	**0.802**	122.4
EXP, CNV	22742	0.201	0.526	0.725	211.2
SNV, CNV	13321	0.226	0.094	0.303	137.4
EXP, SNV, CNV	23525	0.201	0.53	0.728	370.2
Random Forest (RF)					
EXP	10204	**0.192**	**0.479**	**0.692**	13.4
SNV	783	0.258	0.021	-0.14	0.375
CNV	12538	0.234	0.053	0.226	18.4
EXP, SNV	10987	0.196	0.476	0.689	13.5
EXP, CNV	22742	0.2	0.381	0.616	31.9
SNV, CNV	13321	0.233	0.055	0.23	18.5
EXP, SNV, CNV	23525	0.196	0.468	0.683	32.2
CNN					
EXP	10204	**0.171**	**0.664**	**0.814**	85.2
SNV	783	0.234	0.067	0.258	13.5
CNV	12538	0.272	0.007	-0.069	122.4
EXP, SNV	10987	0.166	0.617	0.785	90
EXP, CNV	22742	0.184	0.519	0.72	204.6
SNV, CNV	13321	0.251	0.015	0.114	129.6
EXP, SNV, CNV	23525	0.206	0.595	0.771	208.8

CNN showed the best results in terms of RMSE, R^2, and PCC; it achieved 0.171, 0.664, and 0.814, respectively. Then DF came second best and finally RF. Both RF and CNN exhibited no effect when multi-omics were integrated because gene expression caused the highest results. DF presented an 8% improvement ratio when EXP and SNV were integrated. Exp played an essential factor in achieving remarkable results, CNV, and SNV.

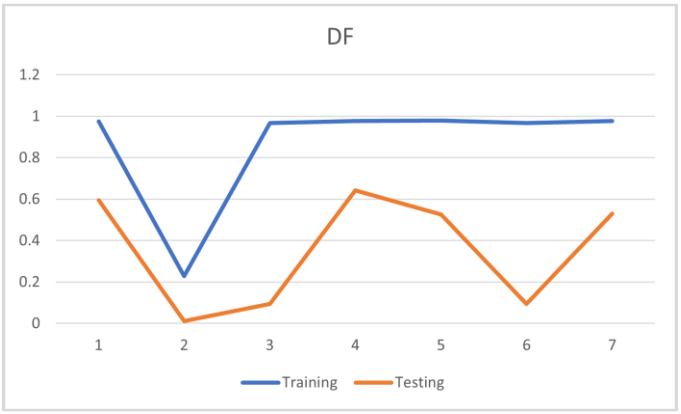

Fig. 6. R^2 results for the training and testing of the DF in the Mean of Coefficients scenario.

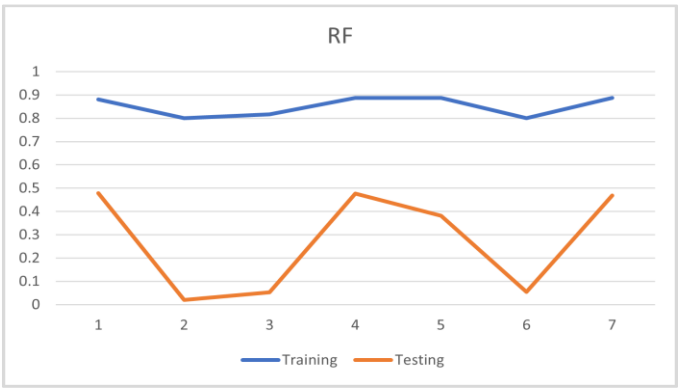

Fig. 7. R^2 results for the training and testing of the RF in the Mean of Coefficients scenario.

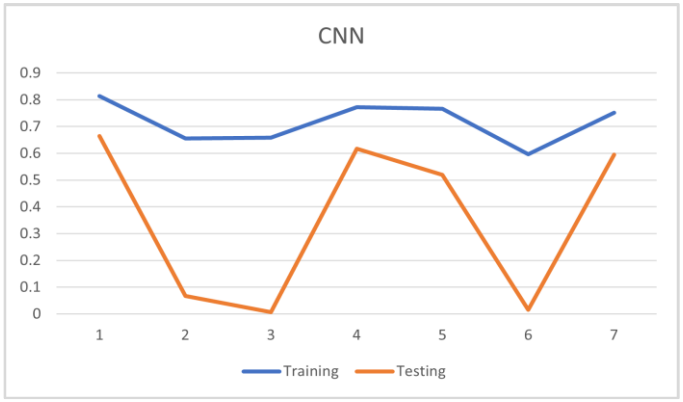

Fig. 8. R^2 results for the training and testing of the CNN in the Mean of Coefficients scenario.

Fig. 6 to 8 show the differences between training and testing results in The Mean of Coefficients as a Threshold scenario. There is still overfitting even though feature selection, cross-validation, and Dropout were utilized. The average ratio of overfitting for DF, RF, and CNN was 63%, 69%, and 54%, respectively. CNN showed the lowest differences ratio between the R2 training and testing results. CNN overfitting ratio was reduced from 72% to 54% compared to the Baseline scenario. However, RF and DF did not show a positive effect of applying BBR.

C. The Top 10% of Coefficients

In this scenario, the coefficient of each feature was calculated using BRR. Then, the mean of all coefficients was used as a threshold, and the top 10% of features were selected from the features that passed the threshold. Table V shows the results of this scenario.

TABLE V. THE RESULTS OF THE TOP 10% OF COEFFICIENTS SCENARIO FOR THE CCEL DATA

Type	Features	RMSE	R^2	PCC	Time (Sec)
Deep Forest (DF)					
EXP	1020	0.176	0.706	0.84	47.5
SNV	78	0.227	0.087	0.289	23.2
CNV	1253	0.225	0.107	0.322	51.7
EXP, SNV	1098	**0.175**	**0.708**	**0.842**	45.1
EXP, CNV	2273	0.188	0.584	0.764	35
SNV, CNV	1331	0.222	0.127	0.352	66
EXP, SNV, CNV	2351	0.184	0.618	0.786	56
Random Forest (RF)					
EXP	1020	0.192	0.495	0.703	1.48
SNV	78	0.245	0.016	0.116	0.11
CNV	1253	0.234	0.067	0.256	1.88
EXP, SNV	1098	0.19	0.509	0.713	1.53
EXP, CNV	2273	0.193	0.469	0.685	3.01
SNV, CNV	1331	0.235	0.062	0.247	2
EXP, SNV, CNV	2351	**0.183**	**0.583**	**0.763**	3.02
CNN					
EXP	1020	0.171	0.544	0.737	15.5
SNV	78	0.234	0.035	0.182	6.53
CNV	1253	0.229	0.083	0.283	18
EXP, SNV	1098	0.194	0.495	0.703	16.4
EXP, CNV	2273	0.161	0.564	0.751	26.8
SNV, CNV	1331	0.217	0.167	0.407	18.7
EXP, SNV, CNV	2351	**0.159**	**0.585**	**0.764**	27.4

In this scenario, the highest results were noticed when multi-omics were integrated. DF achieved RMSE 0.175, R^2 0.708, and PCC 0.842 because of combining EXP and SNV. RF and CNN displayed the best scores when the three omics were integrated. The lowest RMSE -0.159- was achieved by CNN.

Fig. 9 to 11 demonstrate the last scenario: The Top 10% of Coefficients. The average ratio of overfitting for DF, RF, and CNN was 54%, 64%, and 39%, respectively. In this scenario, the overfitting of the DF was reduced from 60% to 54% compared to the baseline and from 72% to 39% for CNN. RF had the highest difference between training and testing for all scenarios. Therefore, this scenario showed the best effect of utilizing BRR in the three models.

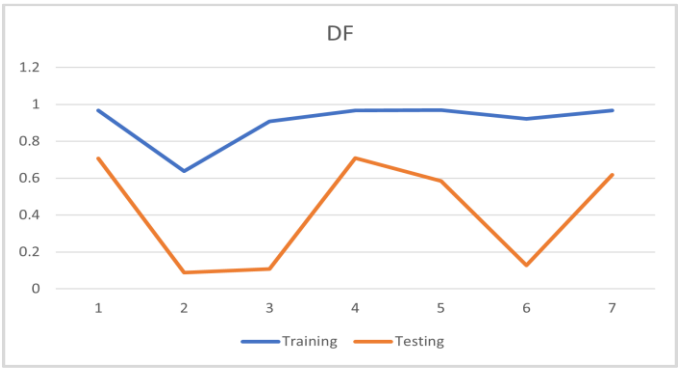

Fig. 9. R^2 results for the training and testing of the DF in the Top 10% of Coefficients scenario.

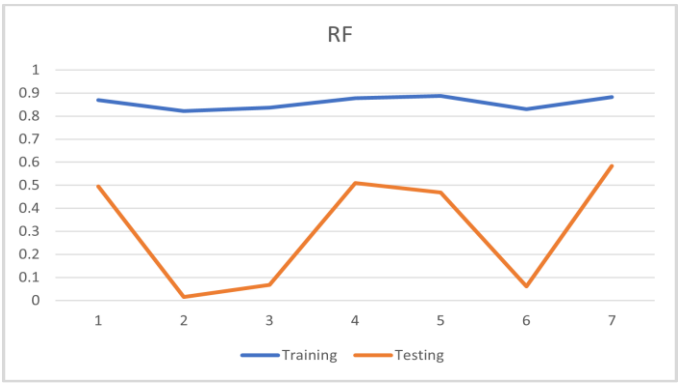

Fig. 10. R^2 results for the training and testing of the RF in the Top 10% of Coefficients scenario.

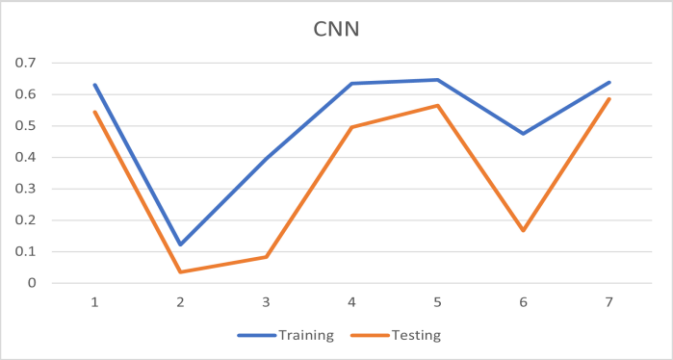

Fig. 11. R^2 results for the training and testing of the CNN in the Top 10% of Coefficients scenario.

D. Summary of the Scenarios and Algorithms

This section presents the comparisons when evaluating which algorithms performed better. The DF, RF, and CNN algorithms were compared in terms of R^2 and RMSE regardless of a specific scenario.

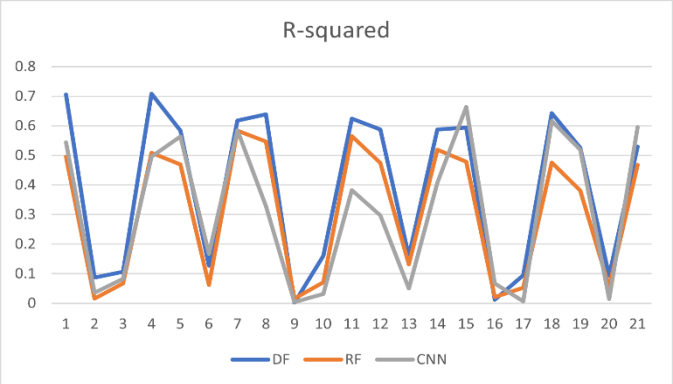

Fig. 12. R^2 results for the DF, RF, and CNN algorithms in all scenarios.

According to Fig. 12, the suggested algorithm Deep Forest (DF) has the best results as it achieved 39% for the average of the R^2 results of all scenarios. Both RF and CNN obtained 31% as the average of R^2.

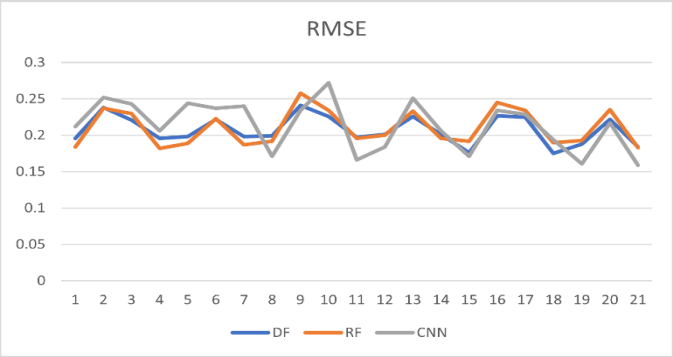

Fig. 13. RMSE results for the DF, RF, and CNN algorithms in all scenarios.

Regarding RMSE, DF obtained the lowest score, as shown in Fig. 13; it was 0.207 which is the average of all scenarios. RF obtained 0.210, and CNN obtained 0.213.

Therefore, the results and performance of the proposed method can be summarized by the following points:

- Bayesian Ridge Regression as a feature selection method for integrating multi-omics showed an 11% improvement ratio in terms of R^2 compared to the Baseline scenario. Also, the complexity time was reduced by 90%.
- The proposed method BRR-DF has the best results in terms of R^2 and RMSE in all three scenarios. The Top 10% scenario exhibited the best performance regardless of the specific algorithm.
- The drug response was mainly explained by the gene expression data more than the copy number and single nucleotide variants.
- Random Forest showed an 18% improvement when the three omics were integrated. Also, it was the fastest algorithm when dealing with these omics data.
- In Deep Forest, integrating gene expression and single nucleotide variant has a better result than integrating all three omics. Both Random Forest and CNN showed better results when all three omics were integrated.

E. Comparison with Related Studies

To evaluate the performance and robustness of the proposed model using the CCLE data, BRR-DF was compared with three state-of-the-art models as follows:

- WGRMF [50]: Weighted Graph Regularized Matrix Factorization is applied for predicting the anti-cancer drug response in cell lines. Their model used the CCLE, which contains 491 cell lines and 23 drugs with 10,870 known responses. WGRMF utilizes gene expression and drug fingerprints as the input for the model.
- DeepDSC [26]: Gene expression is employed to extract the features of cell lines using a stacked deep autoencoder, before the chemical structure is integrated with gene expression to predict the drug response. DeepDSC uses the CCLE, which contains 491 cell lines and 23 drugs with 10,870 known responses.
- SRMF [51]: A Similarity-Regularized Matrix Factorization model combining gene expression with chemical structures for drug response prediction. The CCLE, with 10,870 known responses, contains 491 cell lines and 23 drugs.

Table VI shows a comparison between the proposed method and the three models in the CCLE dataset.

TABLE VI. COMPARISON OF PERFORMANCES WITH OTHER RELATED MODELS

Model	RMSE	PCC	R^2
WGRMF	0.56	0.72	-
DeepDSC	0.23	-	**0.78**
SRMF	0.57	0.71	-
BBR-DF (Proposed)	**0.17**	**0.84**	0.70

The proposed model showed an improvement in terms of achieving the lowest RMSE and highest PCC compared with the other models. R^2 needs to be enhanced, and some of the limitations are discussed in the following section.

F. Effectiveness of Gene Expression in Drug Response

The main challenge in drug development research and clinical trials is the lack of understanding regarding how individuals respond to drugs, which can significantly impact their efficacy and tolerability [52]. The present study aimed to investigate this issue by analyzing gene expression data, copy number alterations (CNA), and single nucleotide variants (SNV) to determine their role in drug response. The results indicated that gene expression data was the most significant factor in describing drug response, as compared to CNA and SNV. To identify the genes that are most important in explaining drug response, the study utilized Bayesian Ridge Regression (BRR) to calculate a coefficient for each gene based on its mRNA expression data. Genes with high coefficients were considered potential candidates for explaining drug response. Table VII presents the top five genes based on their coefficient for 24 drugs, while the complete list of ranked genes for each drug is provided in the Supplementary Materials.

TABLE VII. THE TOP FIVE GENES SELECTED BY THE MODEL FOR 24 DRUGS IN CCLE

Drug	Gene	Drug	Gene
17-AAG	LIN28B	Paclitaxel	ABCB1
	TNFAIP6		UPK1B
	MAGEA4		SLC6A14
	VCAM1		PITX2
	MMP7		PLAC8
AEW541	IFITM2	Panobinostat	CYP1B1
	CD69		CPVL
	CXorf61		TM4SF18
	CLEC2B		VSNL1
	MAGEA11		NMI
AZD0530	MT1E	PD-0325901	DSE
	AC093323.3		COL1A2
	WWC3		MMP1
	CPVL		CXCR7
	SERPINE1		KLHL13
AZD6244	TUBB2B	PD-0332991	SCRN1
	DSE		KLHL13
	ARL4C		AIM2
	CSTA		CSDA
	SRPX2		S100A16
Crizotinib	GSTP1	PHA-665752	CMBL
	CXorf61		FABP4
	TM4SF18		BST2
	TFPI2		CST6
	CR2		LIMCH1
Erlotinib	DKK1	PLX4720	GOLGA8A
	CMBL		TDRD9
	MT1E		MMP1
	GNE		IFITM2
	MUC4		KLHL13
Irinotecan	IFI27	RAF265	CLEC2B
	RBM24		SDC2
	GDF15		ZNF83
	COL11A1		CXCL5
	CA2		PEG10
L-685458	GTSF1	Sorafenib	PRSS21
	CLEC2B		ROBO1
	CSDA		VCAN
	ARHGEF3		ANKRD36BP2
	ABCG1		CYFIP2
lapatinib	MT1E	TAE684	MMP1
	FBP1		HSPA1A
	BST2		RASGRP1
	SPARC		IFI27
	ALKBH3		CHN1
LBW242	AKAP12	TKI258	RGS4
	WASF3		HEY1
	CACHD1		LIN28B
	SLC10A4		SGCE
	EPS8		IGF1R
Nilotinib	DHRS9	Topotecan	MAGEA4
	CPVL		COL11A1
	DDX3Y		NGFRAP1
	PLOD2		CYP24A1
	TFPI2		KLK6
Nutlin-3	BIK	Vandetanib	PODXL
	SAMSN1		CHI3L1
	G0S2		DKK1
	SERPINB1		MT1E
	HLA-DQA1		CRNDE

MT1E (Metallothionein 1E) appears as an important predictor for 4 drugs: AZD0530, Erlotinib, Lapatinib, and Vandetanib. It is a Protein Coding gene. Frontometaphyseal Dysplasia 1 and Bladder Cancer are some of the diseases that are associated with MT1E [53]. In addition, CLEC2B (C-Type Lectin Domain Family 2 Member B) has a high score as an informative gene for 3 drugs: AEW541, L-685458, and RAF265. Several cancers, such as pancreatic adenocarcinoma, melanoma, and clear cell renal cell carcinoma, have been linked to CLEC2B as a marker [54]. However, those genes and their effect on each drug need to be validated in the biological context, which is an essential point in our future work.

V. CONCLUSION

Bayesian Ridge Regression (BRR) was combined with Deep Forest (DF) to enhance drug response prediction by integrating multi-omics data. BRR was used to select informative features for every type based on the coefficient value before integrating it with the other omics. DF works effectively to capture the local features and utilizes a network structure inspired by CNN for continuous improvement. Three scenarios were implemented. In each scenario, three models were utilized to evaluate the proposed model: Deep Forest (proposed), Random Forest, and CNN. BRR-DF displayed an 11% improvement ratio in terms of R^2 compared to the Baseline scenario. Also, the complexity time was reduced by 90%. DF showed the best results in all three scenarios in which it obtained 0.175, 0.842, and 0.708 regarding RMSE, PCC, and R^2, respectively. The Top 10% scenario exhibited the best performance regardless of the specific algorithm. In DF, integrating gene expression and a single nucleotide variant showed a better result than integrating all three omics. Both Random Forest and CNN exhibited better results when all three omics were integrated. Regarding the multi-omics that were used, the drug response was mainly explained by the gene expression data more than the copy number and single nucleotide variants.

There are some limitations to the proposed solution. Firstly: the experiments showed overfitting even though cross-validation and feature selection were utilized. Techniques such as bootstrapping, ensemble methods, and synthetic oversamples could be investigated. Secondly, we only focused on cell line data. Future work will utilize drug information such as chemical structure and the drug target. Thirdly, selecting the best features was implemented manually, in which the top 10% were chosen. However, an automatic method, such as a voting-based will be studied in future work.

REFERENCES

[1] Gambardella et al., "Personalized Medicine: Recent Progress in Cancer Therapy," Cancers (Basel), vol. 12, no. 4, p. 1009, Apr. 2020, doi: 10.3390/cancers12041009.

[2] A. Goodspeed, L. M. Heiser, J. W. Gray, and J. C. Costello, "Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics," Molecular Cancer Research, vol. 14, no. 1, pp. 3–13, Jan. 2016, doi: 10.1158/1541-7786.MCR-15-0189.

[3] R. Kurilov, B. Haibe-Kains, and B. Brors, "Assessment of modelling strategies for drug response prediction in cell lines and xenografts," Sci Rep, vol. 10, no. 1, p. 2849, Feb. 2020, doi: 10.1038/s41598-020-59656-2.

[4] E. S. Lander et al., "Initial sequencing and analysis of the human genome," Nature, vol. 409, no. 6822, pp. 860–921, Feb. 2001, doi: 10.1038/35057062.

[5] Z. Cai, R. C. Poulos, J. Liu, and Q. Zhong, "Machine learning for multi-omics data integration in cancer," iScience, vol. 25, no. 2, p. 103798, Feb. 2022, doi: 10.1016/j.isci.2022.103798.

[6] G. Nicora, F. Vitali, A. Dagliati, N. Geifman, and R. Bellazzi, "Integrated Multi-Omics Analyses in Oncology: A Review of Machine Learning Methods and Tools," Front Oncol, vol. 10, p. 1030, Jun. 2020, doi: 10.3389/fonc.2020.01030.

[7] W. Zhong, "Feature selection for cancer classification using microarray gene expression data," 2014. Accessed: May 16, 2023. [Online]. Available: https://prism.ucalgary.ca/items/0530b8e6-7c69-4a6f-8090-642c9765ef32.

[8] M. Walowe Mwadulo, "A Review on Feature Selection Methods For Classification Tasks," International Journal of Computer Applications Technology and Research, vol. 5, no. 6, pp. 395–402, Jun. 2016, doi: 10.7753/IJCATR0506.1013.

[9] G. Hu, H. Li, Y. Xia, and L. Luo, "A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis," Comput Ind, vol. 100, pp. 287–296, Sep. 2018, doi: 10.1016/j.compind.2018.04.002.

[10] Z.-H. Zhou and J. Feng, "Deep forest," Natl Sci Rev, vol. 6, no. 1, pp. 74–86, 2019.

[11] Z.-H. Zhou and J. Feng, "Deep Forest: Towards An Alternative to Deep Neural Networks," in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, California: International Joint Conferences on Artificial Intelligence Organization, Aug. 2017, pp. 3553–3559. doi: 10.24963/ijcai.2017/497.

[12] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, "Hyperparameter optimization for machine learning models based on Bayesian optimization," Journal of Electronic Science and Technology, vol. 17, no. 1, pp. 26–40, 2019, doi: https://doi.org/10.11989/JEST.1674-862X.80904120.

[13] C. Huang, R. Mezencev, J. F. McDonald, and F. Vannberg, "Open source machine-learning algorithms for the prediction of optimal cancer drug therapies," PLoS One, vol. 12, no. 10, p. e0186906, Oct. 2017, doi: 10.1371/journal.pone.0186906.

[14] Q. Liu, Z. Hu, R. Jiang, and M. Zhou, "DeepCDR: a hybrid graph convolutional network for predicting cancer drug response," Bioinformatics, vol. 36, no. Supplement_2, pp. i911–i918, Dec. 2020, doi: 10.1093/bioinformatics/btaa822.

[15] T. Nguyen, G. T. T. Nguyen, T. Nguyen, and D.-H. Le, "Graph Convolutional Networks for Drug Response Prediction," IEEE/ACM Trans Comput Biol Bioinform, vol. 19, no. 1, pp. 146–154, Jan. 2022, doi: 10.1109/TCBB.2021.3060430.

[16] J. C. Costello et al., "A community effort to assess and improve drug sensitivity prediction algorithms," Nat Biotechnol, vol. 32, no. 12, pp. 1202–1212, Dec. 2014, doi: 10.1038/nbt.2877.

[17] M. Gönen and A. A. Margolin, "Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning," Bioinformatics, vol. 30, no. 17, pp. i556–i563, Sep. 2014, doi: 10.1093/bioinformatics/btu464.

[18] F. Iorio et al., "A Landscape of Pharmacogenomic Interactions in Cancer," Cell, vol. 166, no. 3, pp. 740–754, Jul. 2016, doi: 10.1016/j.cell.2016.06.017.

[19] S. Naulaerts, C. C. Dang, and P. J. Ballester, "Precision and recall oncology: combining multiple gene mutations for improved identification of drug-sensitive tumours," Oncotarget, vol. 8, no. 57, pp. 97025–97040, Nov. 2017, doi: 10.18632/oncotarget.20923.

[20] A. K. Mitra et al., "A gene expression signature distinguishes innate response and resistance to proteasome inhibitors in multiple myeloma," Blood Cancer J, vol. 7, no. 6, pp. e581–e581, Jun. 2017, doi: 10.1038/bcj.2017.56.

[21] X. Xu, H. Gu, Y. Wang, J. Wang, and P. Qin, "Autoencoder Based Feature Selection Method for Classification of Anticancer Drug Response," Front Genet, vol. 10, p. 233, Mar. 2019, doi: 10.3389/fgene.2019.00233.

[22] M. P. Menden et al., "Machine Learning Prediction of Cancer Cell Sensitivity to Drugs Based on Genomic and Chemical Properties," PLoS One, vol. 8, no. 4, p. e61318, Apr. 2013, doi: 10.1371/journal.pone.0061318.

[23] D. Baptista, P. G. Ferreira, and M. Rocha, "Deep learning for drug response prediction in cancer," Brief Bioinform, vol. 22, no. 1, pp. 360–379, Jan. 2021, doi: 10.1093/bib/bbz171.

[24] A. B. Dincer, S. Celik, N. Hiranuma, and S.-I. Lee, "DeepProfile: Deep learning of cancer molecular profiles for precision medicine," BioRxiv, p. 278739, 2018, doi: https://doi.org/10.1101/278739.

[25] Y. Chang et al., "Cancer Drug Response Profile scan (CDRscan): A Deep Learning Model That Predicts Drug Effectiveness from Cancer Genomic Signature," Sci Rep, vol. 8, no. 1, p. 8857, Jun. 2018, doi: 10.1038/s41598-018-27214-6.

[26] M. Li et al., "DeepDSC: A Deep Learning Method to Predict Drug Sensitivity of Cancer Cell Lines," IEEE/ACM Trans Comput Biol

Bioinform, vol. 18, no. 2, pp. 575–582, Mar. 2021, doi: 10.1109/TCBB.2019.2919581.

[27] W. Xu, X. Liu, F. Leng, and W. Li, "Blood-based multi-tissue gene expression inference with Bayesian ridge regression," Bioinformatics, vol. 36, no. 12, pp. 3788–3794, Jun. 2020, doi: 10.1093/bioinformatics/btaa239.

[28] B. Velten and W. Huber, "Adaptive penalization in high-dimensional regression and classification with external covariates using variational Bayes," Biostatistics, vol. 22, no. 2, pp. 348–364, Apr. 2021, doi: 10.1093/biostatistics/kxz034.

[29] H. Sharifi-Noghabi, O. Zolotareva, C. C. Collins, and M. Ester, "MOLI: multi-omics late integration with deep neural networks for drug response prediction," Bioinformatics, vol. 35, no. 14, pp. i501–i509, Jul. 2019, doi: 10.1093/bioinformatics/btz318.

[30] V. Malik, Y. Kalakoti, and D. Sundar, "Deep learning assisted multi-omics integration for survival and drug-response prediction in breast cancer," BMC Genomics, vol. 22, pp. 1–11, 2021.

[31] Z.-H. Zhou and J. Feng, "Deep forest," Natl Sci Rev, vol. 6, no. 1, pp. 74–86, Jan. 2019, doi: 10.1093/nsr/nwy108.

[32] R. Su, X. Liu, L. Wei, and Q. Zou, "Deep-Resp-Forest: A deep forest model to predict anti-cancer drug response," Methods, vol. 166, pp. 91–102, Aug. 2019, doi: 10.1016/j.ymeth.2019.02.009.

[33] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), IEEE, 2006, pp. 1735–1742.

[34] G. L. Guimaraes, B. Sanchez-Lengeling, C. Outeiral, P. L. C. Farias, and A. Aspuru-Guzik, "Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models," arXiv preprint arXiv:1705.10843, 2017.

[35] M. Popova, O. Isayev, and A. Tropsha, "Deep reinforcement learning for de novo drug design," Sci Adv, vol. 4, no. 7, p. eaap7885, 2018.

[36] Y. Chen and L. Zhang, "How much can deep learning improve prediction of the responses to drugs in cancer cell lines?," Brief Bioinform, vol. 23, no. 1, p. bbab378, 2022.

[37] P. Jia, R. Hu, G. Pei, Y. Dai, Y.-Y. Wang, and Z. Zhao, "Deep generative neural network for accurate drug response imputation," Nat Commun, vol. 12, no. 1, p. 1740, Mar. 2021, doi: 10.1038/s41467-021-21997-5.

[38] M. Pouryahya et al., "Pan-Cancer Prediction of Cell-Line Drug Sensitivity Using Network-Based Methods," Int J Mol Sci, vol. 23, no. 3, p. 1074, Jan. 2022, doi: 10.3390/ijms23031074.

[39] H. Wang et al., "GADRP: graph convolutional networks and autoencoders for cancer drug response prediction," Brief Bioinform, vol. 24, no. 1, p. bbac501, Jan. 2023, doi: 10.1093/bib/bbac501.

[40] J. Barretina et al., "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity," Nature, vol. 483, no. 7391, pp. 603–607, Mar. 2012, doi: 10.1038/nature11003.

[41] T. J. Mitchell and J. J. Beauchamp, "Bayesian Variable Selection in Linear Regression," J Am Stat Assoc, vol. 83, no. 404, pp. 1023–1032, Dec. 1988, doi: 10.1080/01621459.1988.10478694.

[42] C. M. Bishop and M. E. Tipping, "Bayesian regression and classification," Nato Science Series sub Series III Computer And Systems Sciences, vol. 190, pp. 267–288, 2003.

[43] J. Q. Shi, R. Murray-Smith, and D. M. Titterington, "Bayesian regression and classification using mixtures of Gaussian processes," Int J Adapt Control Signal Process, vol. 17, no. 2, pp. 149–161, Mar. 2003, doi: 10.1002/acs.744.

[44] S. Ozdemir and D. Susarla, Feature Engineering Made Easy: Identify unique features from your dataset in order to build powerful machine learning systems. Packt Publishing Ltd, 2018.

[45] Y. Guo, S. Liu, Z. Li, and X. Shang, "BCDForest: a boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data," BMC Bioinformatics, vol. 19, no. S5, p. 118, Apr. 2018, doi: 10.1186/s12859-018-2095-4.

[46] C. Xin, X. Shi, D. Wang, C. Yang, Q. Li, and H. Liu, "Multi-grained cascade forest for effluent quality prediction of papermaking wastewater treatment processes," Water Science and Technology, vol. 81, no. 5, pp. 1090–1098, Mar. 2020, doi: 10.2166/wst.2020.206.

[47] M. Shahzad, M. A. Tahir, M. A. Khan, R. Jiang, and R. A. S. Malick, "EBSRMF: Ensemble based similarity-regularized matrix factorization to predict anticancer drug responses," Journal of Intelligent & Fuzzy Systems, vol. 43, no. 3, pp. 3443–3452, Jul. 2022, doi: 10.3233/JIFS-212867.

[48] A. Golbraikh and A. Tropsha, "Beware of q2!," J Mol Graph Model, vol. 20, no. 4, pp. 269–276, Jan. 2002, doi: 10.1016/S1093-3263(01)00123-1.

[49] Y. Fang, P. Xu, J. Yang, and Y. Qin, "A quantile regression forest based method to predict drug response and assess prediction reliability," PLoS One, vol. 13, no. 10, p. e0205155, Oct. 2018, doi: 10.1371/journal.pone.0205155.

[50] N.-N. Guan, Y. Zhao, C.-C. Wang, J.-Q. Li, X. Chen, and X. Piao, "Anticancer Drug Response Prediction in Cell Lines Using Weighted Graph Regularized Matrix Factorization," Mol Ther Nucleic Acids, vol. 17, pp. 164–174, Sep. 2019, doi: 10.1016/j.omtn.2019.05.017.

[51] L. Wang, X. Li, L. Zhang, and Q. Gao, "Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization," BMC Cancer, vol. 17, no. 1, p. 513, Dec. 2017, doi: 10.1186/s12885-017-3500-5.

[52] R. Herwig and H. Lehrach, "Expression profiling of drug response - from genes to pathways," Dialogues Clin Neurosci, vol. 8, no. 3, pp. 283–293, Sep. 2006, doi: 10.31887/DCNS.2006.8.3/rherwig.

[53] G. Stelzer et al., "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses," Curr Protoc Bioinformatics, vol. 54, no. 1, pp. 1–30, Jun. 2016, doi: 10.1002/cpbi.5.

[54] X. Li, X. Tao, and X. Ding, "An integrative analysis to reveal that CLEC2B and ferroptosis may bridge the gap between psoriatic arthritis and cancer development," Sci Rep, vol. 12, no. 1, p. 14653, Aug. 2022, doi: 10.1038/s41598-022-19135-2.