Methods 166 (2019) 91–102

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Deep-Resp-Forest: A deep forest model to predict anti-cancer drug response

Ran Su[a], Xinyi Liu[a], Leyi Wei[b], Quan Zou[c,*]

[a] School of Computer Software, College of Intelligence and Computing, Tianjin University, Tianjin, China
[b] School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin, China
[c] Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China

ARTICLE INFO

Keywords:
Deep learning
Anti-cancer
Drug response
Prediction
Deep forest

ABSTRACT

The identification of therapeutic biomarkers predictive of drug response is crucial in personalized medicine. A number of computational models to predict response of anti-cancer drugs have been developed as the establishment of several pharmacogenomics screening databases. In our study, we proposed a deep cascaded forest model, Deep-Resp-Forest, to classify the anti-cancer drug response as "sensitive" or "resistant". We made three contributions in this study. Firstly, diverse molecular data could be effectively integrated to provide more information than single type of data for the classification. Combination of two types of data were tested here. Secondly, two structures based on the multi-grained scanning to transform the raw features into high-dimensional feature vectors and integrate the diverse data were proposed in our study. Thirdly, the original deep and time-consuming architecture of cascade forest was improved by a feature optimization operation, which emphasized the most discriminative features across layers. We evaluated the proposed method on the Cancer Cell Line Encyclopedia (CCLE) and Genomics of Drug Sensitivity in Cancer (GDSC) data sets and then compared with the Support Vector Machine. The proposed Deep-Resp-Forest has demonstrated the promising use of deep learning and deep forest approach on the drug response prediction tasks. The R implementation for running our experiments is available at https://github.com/RanSuLab/Deep-Resp-Forest.

1. Introduction

The growth in the field of genetics has led researchers to find genetic differences in human and their cancers. The traditional therapies that use the similar treatments to patients who have the same type and stage of cancer may not be effective due to the cancer's heterogeneity and genetic diversity even within a single type or subtype [14]. Therefore, the cancer treatment should be individualized, and target the cancer-causing genetic changes in each individual's cancer. Personalized medicine or precision medicine, or genomic medicine, customizes the health-care such as medical decisions, practices, therapies and products being tailored to the individual patient [26,20]. It takes into account an individuals genetic makeup and disease history before a treatment regimen is generated [43]. The identification of critical biomarkers to understand the pathogenesis is a major challenge in personalized medicine of cancer [12]. With the establishment of some large-scale pharmacogenomics screening data sets, such as the NCI-60 [35], Genomics of Drug Sensitivity in Cancer (GDSC) [51] and Cancer Cell Line Encyclopedia (CCLE) [6], it is possible to identify biomarkers useful for predicting the anti-cancer drug response of a given patient through building drug-response prediction model. The molecular profiles and the pharmacological information are integrated to explore the individual cell line's response to anti-cancer drugs and the relevant biomarkers.

Over decades, a variety of models have been established for drug sensitivity prediction. Most of the studies were based on genomic information. Staunton et al. used the gene expression profiles of NCI-60 to build an anti-cancer drug sensitivity prediction model [38]. They used a weighted voting scheme to classify each cell line as sensitive or resistant under certain drug treatment. John et al. used the gene profiles of colon cancer cells to predict and distinguish response to multiple chemotherapeutic agents [33]. Aben et al. found it was hard to interpret the genetic models and proposed a TANDEM algorithm which used "upstream" data composed of mutations, copy number, methylation and cancer type and "downstream" data composed of gene expressions to predict the drug response and maximize the interpretability of the drug response models [1]. Recently, heterogeneous information in addition to the genomic information was integrated to provide more views of the predictors. For instance, chemical structures and gene expression profiles or their interactions were combined to predict the drug response [3,55]; Biological context-specific functional modules were used in cell line specific drug efficacy prediction in [24]. There are a number

* Corresponding author.
E-mail address: zouquan@nclab.net (Q. Zou).

https://doi.org/10.1016/j.ymeth.2019.02.009
Received 13 November 2018; Received in revised form 13 January 2019; Accepted 10 February 2019
Available online 14 February 2019

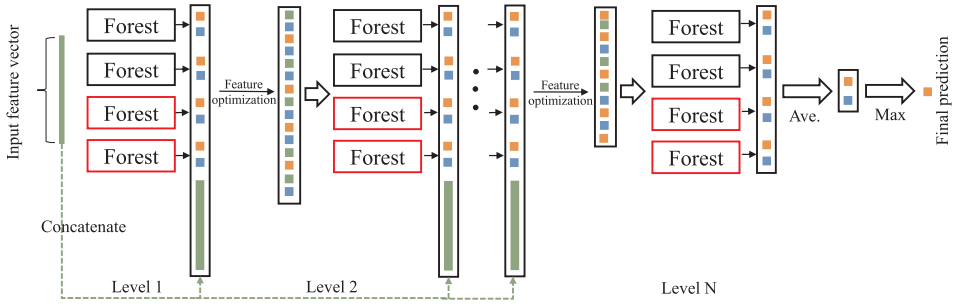


Fig. 1. The basic structure of the cascade forest with feature optimization. Each level is composed of two random forests (black) and two completely random forests (green). All the concatenated features will go through a feature optimization to obtain a more compact feature set and then are passed to the next level. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

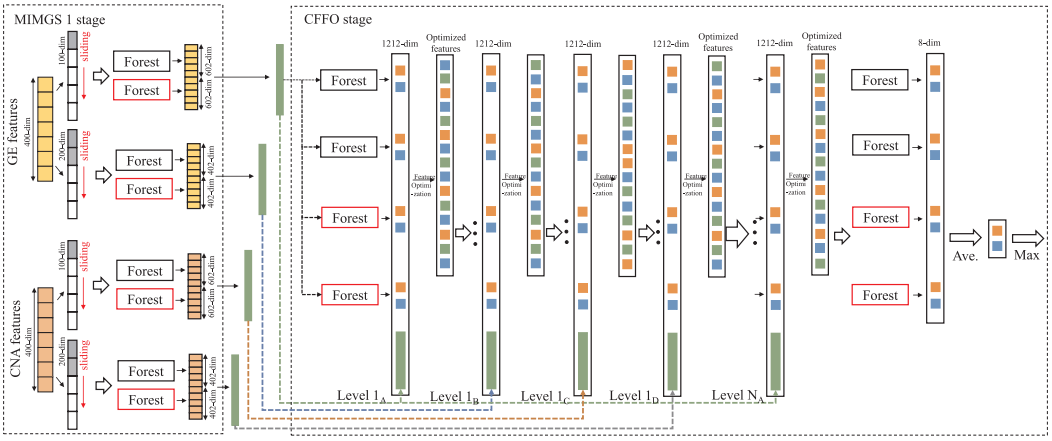


Fig. 2. The overall framework of the Deep-Resp-Forest. The left part shows the multi-info multi-grained scanning (MIMGS) stage; and the right part shows the cascade forest with feature optimization (CFFO) stage. (a) shows the MIMGS 1 stage; (b) shows the MIMGS 2 stage. Here we used two types of data, the gene expression (GE) and copy number alternation (CNA) as examples, and two sizes of windows as example.

computational methods to solve the drug response prediction problem. Among them, Elastic net (EN) [59], Support Vector Regression (SVR) [11] and Random Forest (RF) [9] are three of the most popular methods. In a literature survey made by [25], EN has been recommended as one of the best performing algorithms. However, it has been reported that the correlations between the predicted response and the observed response using these algorithms are rather low, with 0.20 (RF) and 0.23 (SVR) in GDSC, and 0.49 (RF) and 0.38 (SVR) in CCLE [55]. Therefore, obtaining an accurate and biologically interpretable prediction model is still a challenge.

Deep learning have shown impressive performance on a diverse set of biomedical molecular-level tasks. For instance, Sekhon et al. used deep learning to predict differential gene expression from histone modification signals [36]; Lanchantin et al. used a LSTM to learn the interaction and binding between the Transcription Factors (TFs) and the genomic sequences [28]. As for the pharmaceutical industry, a vast number of attentions have been paid to solve the issues in the drug discovery with deep learning [10], such as drug toxicity prediction [34,50], drug repositioning [2,49,5], drug-side effect [16,17], drug-target interaction [15,37], etc. As for the anti-cancer drug response study, not much has been reported with deep learning method. Chang et al. reported a deep learning model named CDRscan based on

Table 1
The accuracy comparison between MIMGS 1 and MIMGS 2 tested on 15 drugs of CCLE.

	AEW541	AZD0530	Erlotinib	LBW242
MIMGS 1	0.789	0.824	0.973	0.930
MIMGS 2	0.691	0.811	0.973	0.930
	Lapatinib	**Nilotinib**	**PD-0325901**	**PD-0332991**
MIMGS 1	0.922	0.935	0.721	0.957
MIMGS 2	0.922	0.935	0.679	0.950
	PF2341066	**PLX4720**	**RAF265**	**Sorafenib**
MIMGS 1	0.871	0.963	0.765	0.979
MIMGS 2	0.884	0.963	0.765	0.979
	TAE684	**TKI258**	**Topotecan**	**Average**
MIMGS 1	0.517	0.825	0.975	0.863
MIMGS 2	0.602	0.825	0.975	0.859

Table 2
The accuracy comparison between MIMGS 1 and MIMGS 2 tested on 15 drugs of GDSC.

	Erlotinib	Sunitinib	AZ628	Tozasertib
MIMGS 1	0.800	0.790	0.885	0.945
MIMGS 2	0.800	0.804	0.870	0.945
	Imatinib	**Crizotinib**	**Saracatinib**	**S-Trityl-L-cysteine**
MIMGS 1	0.800	0.850	0.900	0.796
MIMGS 2	0.800	0.850	0.900	0.796
	Z-LLNle-CHO	**GNF-2**	**A-770041**	**WH-4-023**
MIMGS 1	0.952	0.850	0.643	0.606
MIMGS 2	0.952	0.850	0.715	0.639
	BMS-536924	**Pyrimethamine**	**Entinostat**	**Average**
MIMGS 1	0.742	0.987	0.946	0.833
MIMGS 2	0.754	0.987	0.946	0.841

genomic data [10]. This is the handful relevant work based on deep learning method as far as we know. Therefore, using the deep architecture to predict the drug response is still a great challenge and there is a high probability that the performance may be greatly improved with the deep learning approaches. In this study, taking the molecular profiles as the input features, we developed an anti-cancer drug response prediction model based on a deep forest architecture, which was originally proposed by Zhou [58], and we name it as Deep-Resp-Forest. The deep forest named gcForest in Zhou et al.'s work was an cascade of forests. Our contributions can be summarized as following: 1) Diverse molecular information can be effectively incorporated into the proposed model. They were transformed them into feature vectors, generated the class vectors and sent them to different layers to train the cascade trees afterwards. Richer information from diverse aspects was utilized. 2) Two structures based on the multi-grained scanning to transform the original feature vector and integrate the diverse data were proposed in our study. We conducted experiments on different data sets to show the performance of each structure. 3) The gcForest went deeply using the outputs at the preceding level as the current level's inputs. This deep and time-consuming architecture was improved by a feature optimization operation, which emphasized the most discriminative features across layers and greatly reduce the computational cost. The Deep-Resp-Forest does not only utilize the strengths of the gcForest, such as easy training and exploiting, as well as the ability to handle small scale data, but it also integrates information from multiple aspects, which provides more information for representation learning, also, the improvement of the cascade forest structure greatly reduces the computation burden so it is applicable on large-scale data. From the patient's perspective, it is more preferred for them to learn if one drug works or not, rather than a specific number [18]. Predicting "sensitive" or "resistant" in this case is more meaningful. This becomes a binary classification problem to label a drug as "sensitive" and "resistant".

2. Materials and methods

2.1. Materials and data sets

We used two data sets to build the drug response prediction models. The first data set we used is the Cancer Cell Line Encyclopedia (CCLE) [6] database. It contains a range of 33–275 cancer cell lines screened for 24 drugs with gene expression profiles, copy number alteration (CNA) and drug sensitivity measurements. The second data set is the Genomics of Drug Sensitivity in Cancer (GDSC) [19] database. It comprises a panel of cancer cell lines screened for 265 drugs. 12–156 cell lines were documented with gene profiles, CNA data and drug response. The expression profiles which were extracted from hybridized RNA in HT-HGU133A Affymetrix whole genome array and normalized by robust multi-array averaging (RMA), and copy number alterations (CNAs) which are binary values were used as the classification features. Activity area and IC50 were used for both database for sensitivity measurement. Here, according to [55], we used activity area for CCLE and IC50 for GDSC. In order to save the computation time, we randomly selected 400 genes and 400 CNA for CCLE and 400 genes and 400 CNA for GDSC. The CCLE data is publicly available from www.broadinstitute.org/ccle/ and the CGP is available from www.cancerrxgene.org/.

2.2. Drug sensitivity classification

We z-score normalized the drug sensitivity values to zero mean and unit variance across all the cell lines. According to [18], for CCLE data, we categorized the response of one cell line to one drug as sensitive if the normalized active area was higher than 0.8; and resistant if the normalized active area was lower than -0.8. Cell lines with activity area between -0.8 and 0.8 were considered to be intermediate and eliminated. For the GDSC data, we categorized the response of one cell line to one drug as sensitive if the normalized active area was lower than -0.8; and resistant if the normalized active area was higher than 0.8.

2.3. The multi-info multi-grained scanning (MIMGS) stage

There are two important stages in the gcForest, the multi-grained scanning stage and the cascade forest stage. The multi-grained stage transformed the original feature vector to a high dimensional feature vector and the cascade forest performed the classification using the high dimensional feature vector as well as the class vectors obtained from the preceding level. We improved the multi-grained scanning stage through effectively adding more types of information and we named it multi-info multi-grained scanning (MIMGS). Here we proposed two sub-structures for MIMGS and introduced them in the following sections.

2.3.1. MIMGS 1: sending different feature vectors to different layers

As mentioned earlier, the design of the multi-grained scanning in the gcForest aims to enhance the cascade forest stage. It converted the raw features into feature vectors which were delineated by the sliding windows. Then the feature vectors were used to train a completely-random tree forest and a random forest to generate the class vectors. In

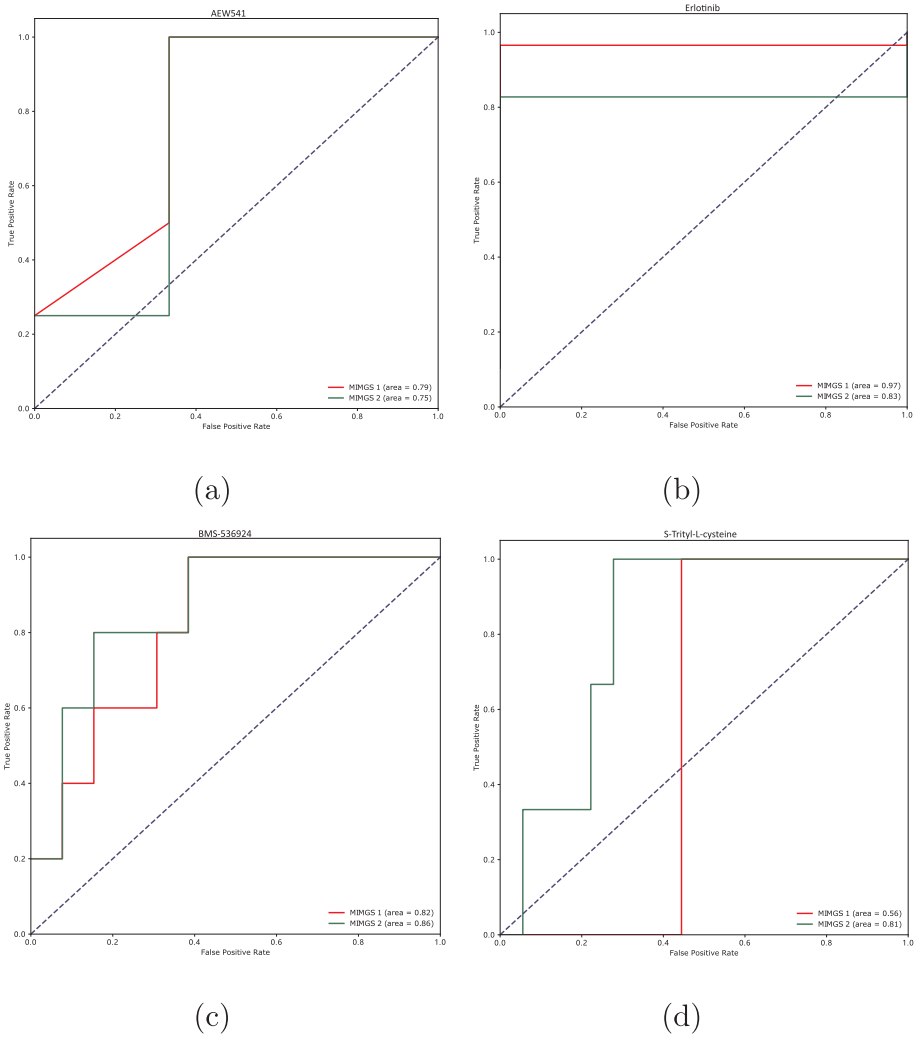


Fig. 3. The ROC curves using to show the performance of MIMGS 1 and MIMGS 2. (a) and (b) are drugs from CCLE; (c) and (d) are drugs from GDSC.

our study, this stage was performed with different types of information and multiple sizes of windows was used to generate different feature vectors. Here we used two types of data, the gene expression and CNA as examples, and two sizes of windows as example, more types of data and more sizes of windows can be incorporated. Firstly, we fed the gene expression (GE) profiles and the copy number alternations (CNA) into the Deep-Resp-Forest. Both types of data were used as the features for the classifier. We call the gene expression profiles as GE feature group and the copy number alternations is called CNA feature group. At the MIMGS stage, each feature group was scanned with multiple sized windows, extracting a number of feature vectors, and class vectors were generated and concatenated to form the input for the cascade forest stage.

We used a sliding window with size w to scan the GE and CNA groups respectively. The scanning on the whole feature sets produced a number of feature subsets. We assume there are totally N_{GE} and N_{CNA} features in the two groups and the stride of the sliding is s. After the scanning, we obtained $(N_{\mathrm{GE}} - w)/s + 1$ feature vectors or instances which contain w_{GE} dimensional features for GE and $(N_{\mathrm{CNA}} - w)/s + 1$ instances containing w_{CNA} dimensional features for CNA. The label of the feature vector was determined by its origin's label, that is, if the feature vector was obtained from a "sensitive" case, its label will be "sensitive" and vice versa. Then the instances would be used to train completely-random tree forests and random forests. This would generate class vectors which are an estimate of class distribution. We flattened the class vectors and a d_{GE}-dimensional or d_{CNA}-dimensional feature vector was obtained. d_{GE} and d_{CNA} were calculated as following:

$$d_{\mathrm{GE}} = ((N_{\mathrm{GE}} - w_{\mathrm{GE}})/s_{\mathrm{GE}} + 1)\cdot c\cdot n_0$$
$$d_{\mathrm{CNA}} = ((N_{\mathrm{CNA}} - w_{\mathrm{CNA}})/s_{\mathrm{CNA}} + 1)\cdot c\cdot n_0 \quad (1)$$

where c is the number of classes and n_0 is the number of forests (sum of both completely-random tree forest and random forest) in each feature group at MIMGS stage. In our study, since our problem is a binary classification problem, c was set to 2 and the class vectors were two-dimensional. We used the generated instances and entered them to a completely-random tree forest and a random forest so n_0 was also set to 2. Here we assume the number of forests in each feature group is the same. To capture feature vectors with diverse sizes, in our studies, we used m_{GE} w_{GE} and m_{CNA} w_{CNA} values and generated multiple class vectors for each feature group. Together we obtained $m_{\mathrm{GE}} + m_{\mathrm{CNA}}$ feature vectors. Then each of the transformed feature vector was sent to each layer in the cascade forest sequentially. Using this operation, we provided the cascade forests diverse information for learning and trained more powerful classifier. The MIMGS 1 is shown in Fig. 2a.

2.3.2. *MIMGS 2: concatenating different feature vectors and sending to all the layers*

We also tested another structure to produce the transformed feature vectors. We concatenated all the class vectors from all the feature groups, formed a rather large feature vector and sent to the cascade forest. It has the length or dimension D as following:

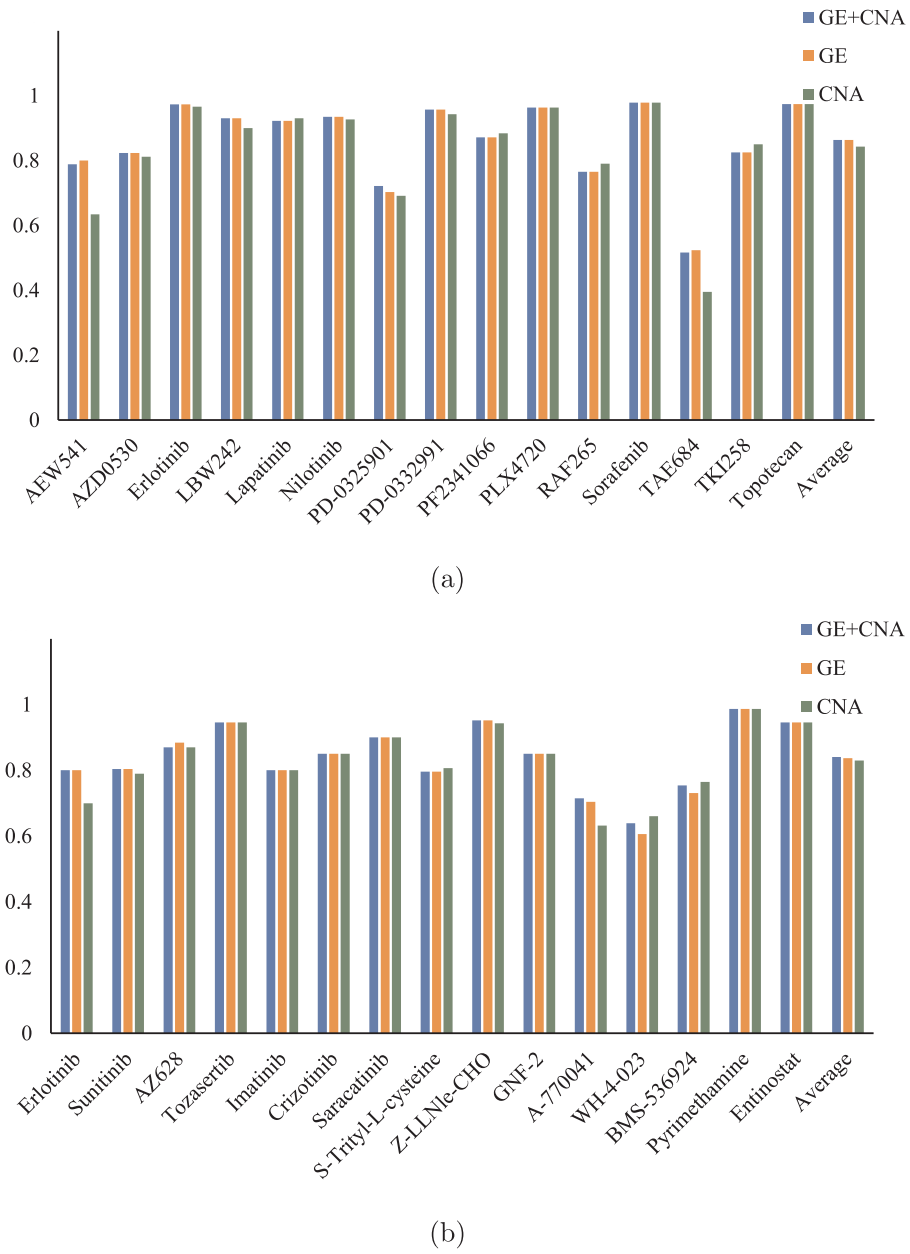


Fig. 4. Performance comparison between the use of GE and CNA together, GE only and CNA only.

$$D = m_{GE} \cdot d_{GE} + m_{CNA} \cdot d_{CNA} \quad (2)$$

where m_{GE} and m_{CNA} are the number of the windows used in GE and CNA respectively. Here they were both set to 2. The D-dimensional feature vector would be used for the cascade forest prediction. Through the MIMGS, the original N_{GE}-dimensional and N_{CNA}-dimensional feature space was transformed to $((N_{GE} - w)/s + 1) \cdot c \cdot n_0$ and $((N_{CNA} - w)/s + 1) \cdot c \cdot n_0$ dimension. Normally, the transformed feature space is much larger than the original feature space. We denoted the transformed features with F_{t0}. The design of multiple information and the multi-grained scanning enhance the representation learning ability of cascade forest. Here we tested on four hundred dimensional feature space, if the input is in high dimension, the multi-grained scanning is quite similar to the 1D or 2D convolution operations in DNN, which extracts features in contextual or structural manner. The concatenation of all the features of diverse types and multiple window sizes in MIMGS is different with the original gcForest architecture, where the transformed features corresponding to different window size were fed to different layers. We show the MIMGS 2 in Fig. 2b.

2.4. Cascade forest with feature optimization (CFFO)

Reported by Breiman, usually the ensemble algorithm obtained the best performance among all the single predictors [8]. So the cascade forest stage in the standard gcForest employed an ensemble-ensemble structure, which means it has the architecture of ensemble of tree forests and the tree forest is the ensemble of decision trees. Also according to [8], ensemble method outperforms its each base model and the most improvement occurs when ensembling more dissimilar predictors [8]. To ensure the diversity, two types of forests are used, one is the completely-random tree forest and the other is the random forest.

As mentioned in Section 2.3, class vectors which represent the class distribution were generated through the forests. The class distribution computed the proportion of each class at the leaf node of one tree for the particular instance and then the values of all the trees was averaged as the distribution of that class. At each level, for each type of information and each sized window, n_1 forests were trained. In our studies, n_1 equaled to 4, which included two completely-random tree

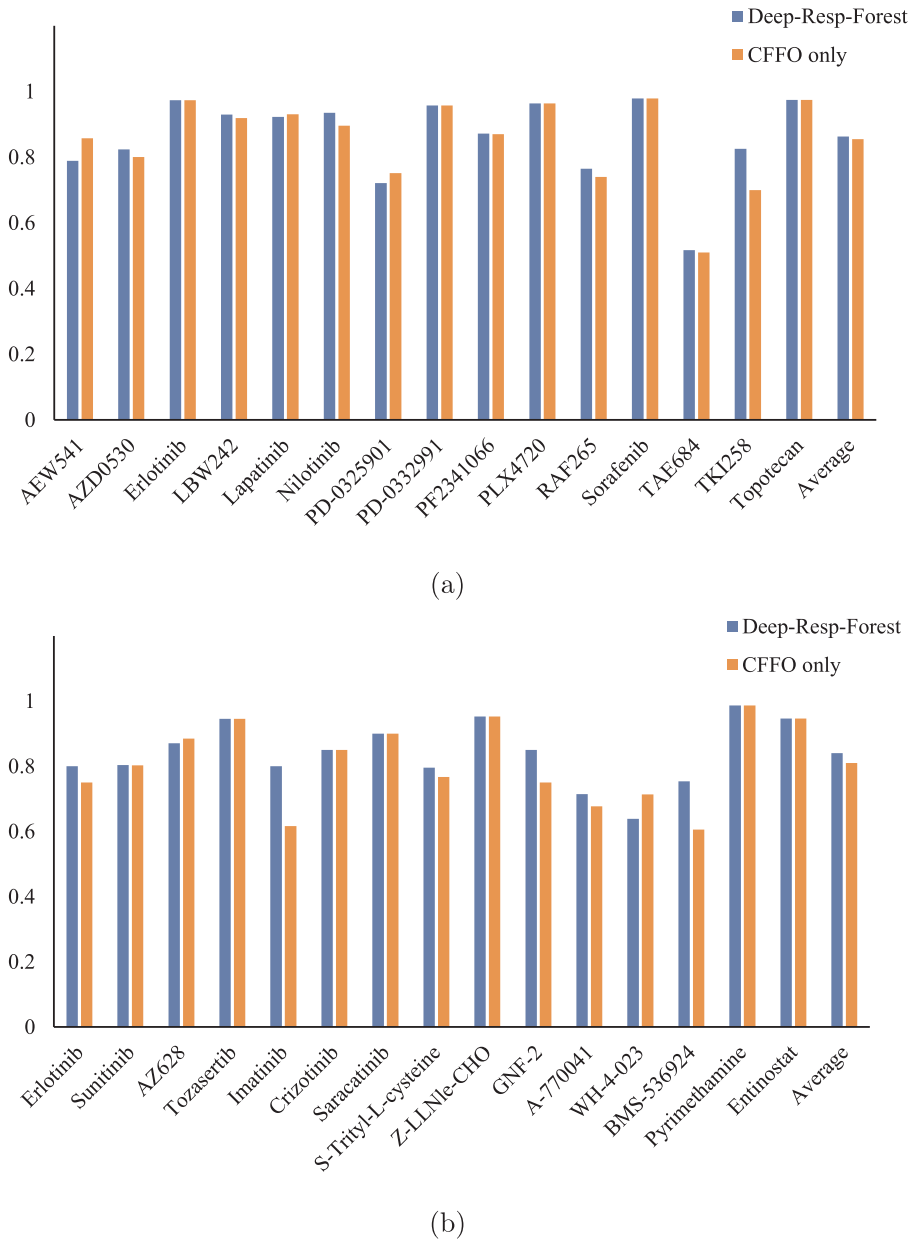
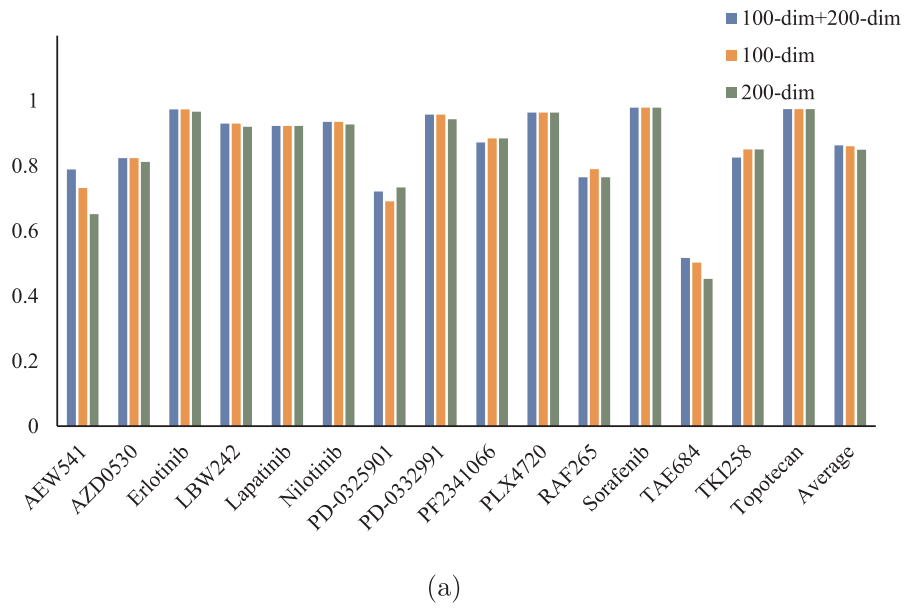


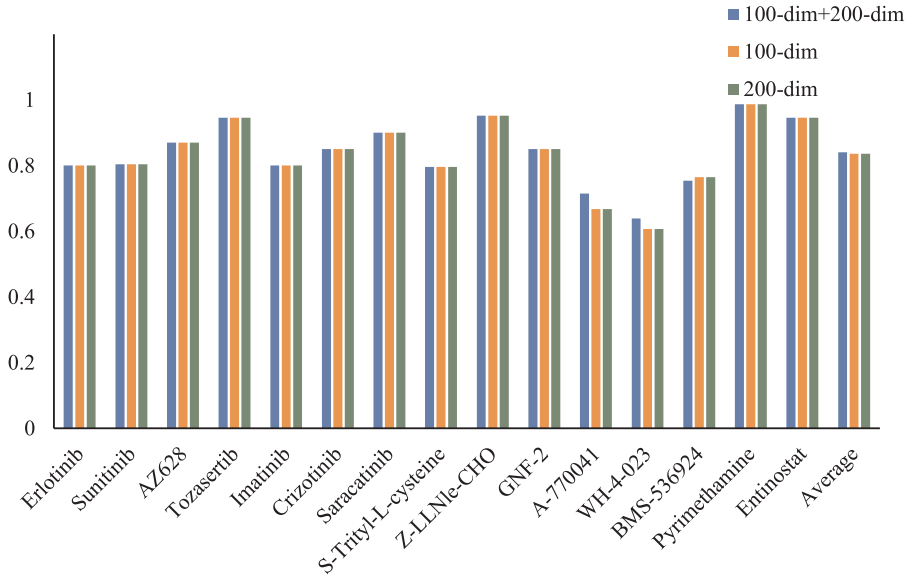
Fig. 5. Performance comparison between the full model and the CFFO only. Here the MIMGS 1 was used in CCLE for comparison and MIMGS 2 was used in GDSC for comparison.

forests and two random forests. The class vectors produced by the $c \cdot n_1$ forests were concatenated and then "old" features were combined and formed a new high-dimensional feature vector. In the standard gcForest, the new feature vector was used again to train the four forests at the next level. This operation uses the artificial high-dimensional feature space across all the layers, which may lead to heavy computational burden for the whole structure considering the very deep structure and the performance is sensitive to the number of decision trees. Guo et al. proposed a BCDForest (Boosting Cascade Deep Forest) model for the prediction of cancer subtype [22]. They boosted the important features at each level of the cascade forest through constructing a new feature for each forest using the standard deviation of the most important features. In our studies, we performed a feature optimization before each layer and the most informative feature were emphasized with this strategy. We show the basic structure of the cascade forest with feature optimization in Fig. 1. It shows from the figure, after each layer, the class vectors together with the old features either from MIMGS 1 or MIMGS 2 prepared the input for the next layer. Then we picked the most discriminative p features measured by their feature importance and fed them into the next layer. The feature importance was measured by the predictive difference between the original data set and the permuted data set in the random forest algorithm. We kept the features which have valid importance values and removed features with zero importance values. The feature optimization was performed at each level. The removal of non-important features across layers reduces the computation cost and the emphasis of discriminative features improves the classification performance.

At each level, the forests received feature information processed by its preceding level, and output the results to the next level. Also, when training on a new level, a training set was used for parameter tuning and a validation set was used to estimate the performance. The training procedure terminated until achieving almost a stable performance. Therefore, it is not necessary to pre-define the number of cascade levels in advance and the depth of the cascade can be learned automatically, which makes the Deep-Resp-Forest adapt to different scales of data, not limited to the huge scaled data. Besides, each level used a supervised



(a)



(b)

Fig. 6. Performance comparison between using multiple-sized sliding windows and single-sized sliding windows. We compared the performance between 100-dim combined with 200-dim, 100-dim alone and 200-dim alone. (a) is the result of CCLE and (b) is the result of GDSC. We used MIMGS 1 for CCLE and MIMGS 2 for GDSC.

strategy and the initially combined features, the F_{t0}, were fused into the new features. Therefore, the amount of hyper-parameters were not as huge as the DNN, where a careful parameter tuning is highly demanded. For a test instance, it firstly goes into the MIMGS stage to obtain the transformed and augmented feature representations. Then the transformed features are represented by each level in the CF stage till the last level.

2.5. Performance metrics

We evaluated the performance of the Deep-Resp-Forest using accuracy (ACC). It is defined as following:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \times 100\% \tag{3}$$

where TP, TN, FN and FP are true positive, true negative, false negative, false positive respectively. We used 5-fold cross validation to evaluate the predictive model. There are three kinds of cross-validations: n-fold cross-validation, jackknife cross-validation and independent data test [13]. Among the three tests, jackknife test has been widely used in bioinformatics because it could produce a unique outcome [41,32,44,40,7]. A number of studies used jackknife as a validation method [31,15,30,47,46,48,45]. However, it is time- and source-consuming. Thus, in this paper, we used 5-fold cross-validation to examine the proposed models. The mean values of the ACC across all the folds is used for the final performance evaluation.

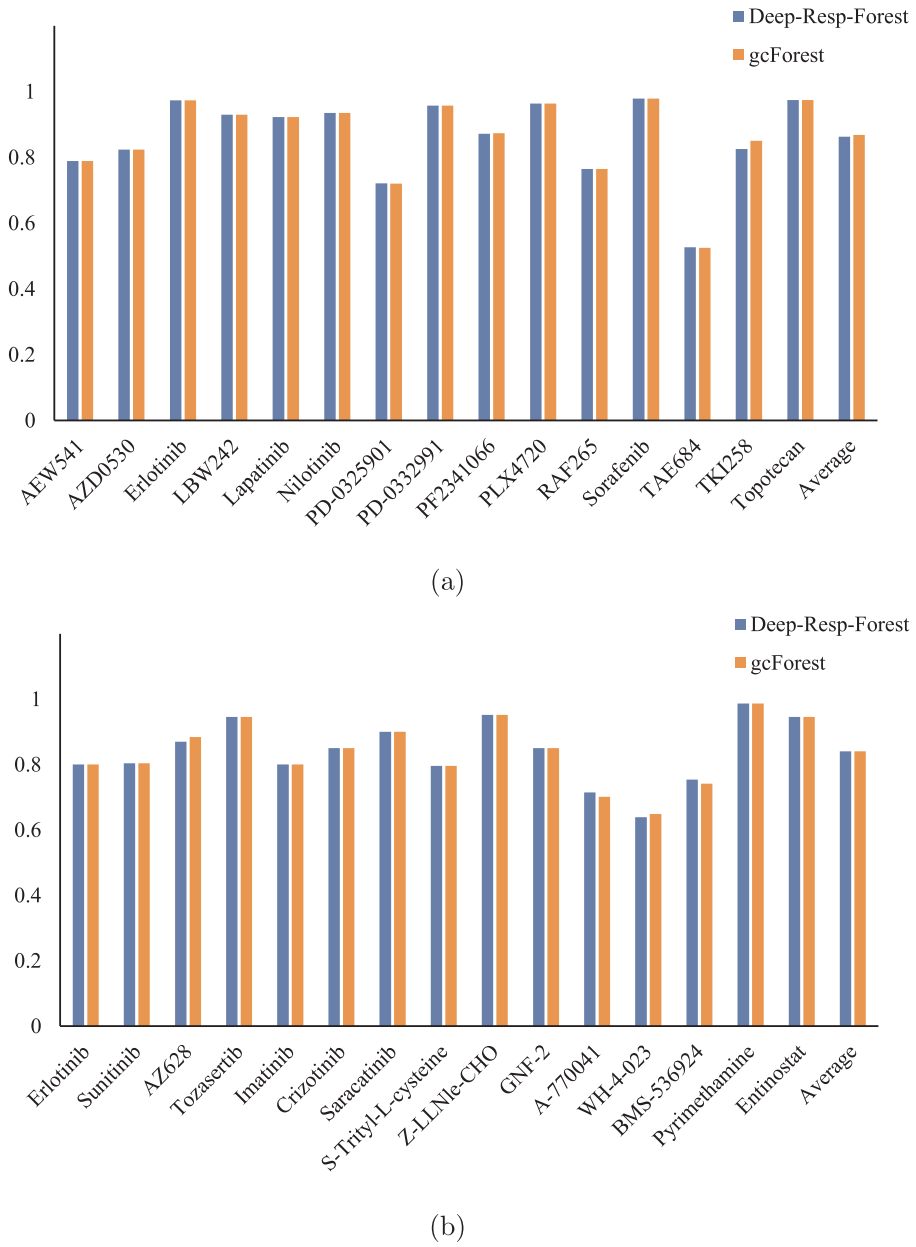


Fig. 7. Performance comparison between the proposed model and the standard gcForest. (a) is the result of CCLE and (b) is the result of GDSC. We tested MIMGS 1 for CCLE and MIMGS 2 for GDSC.

Table 3
Average time consumption of the proposed method and gcForest.

Time cost on average	Deep-resp-Forest	gcForest
CCLE 1	300 s	320 s
GDSC 2	212	262 s

3. Results

3.1. Computational framework of the Deep-Resp-Forest

The success of machine learning algorithm lies largely upon its representation learning capability [29]. Deep neural network (DNN) is composed of multiple layers to learn representations of input data and has achieved great success in many areas, such as image processing [27,23], speech recognition [21,57] and even drug discovery [4]. However, it suffers from several problems. First, a huge amount of data is required for training, but in real life, it is not so easy to see data in that huge scale; Second, the deep neural network is like a “black box” and it is difficult to exploit inside; Last, too many hyper-parameters make the training tricky, so it is hard for replication. Zhou et al. proposed the deep forest, the multi-grained cascade forest, also named gcForest to overcome these problems [58]. In our study, we predicted the anti-cancer drug response with an improved deep forest structure. We effectively integrate diverse information and predict in a more efficient manner. We show the overall architecture in Fig. 2.

From Fig. 2, it shows there are two important stages in the proposed Deep-Resp-Forest, the multi-info multi-grained scanning (MIMGS) and the cascade forest with feature optimization (CFFO). The MIMGS produces the feature vectors for the CFFO stage; Then the CFFO performs the classification in a level by level manner. The representational learning ability is enhanced both by the MIMGS and the CFFO. We built models on CCLE and GDSC using the IC50 and active area respectively. The response of “resistant” was labeled 0 and “sensitive” was labeled 1.

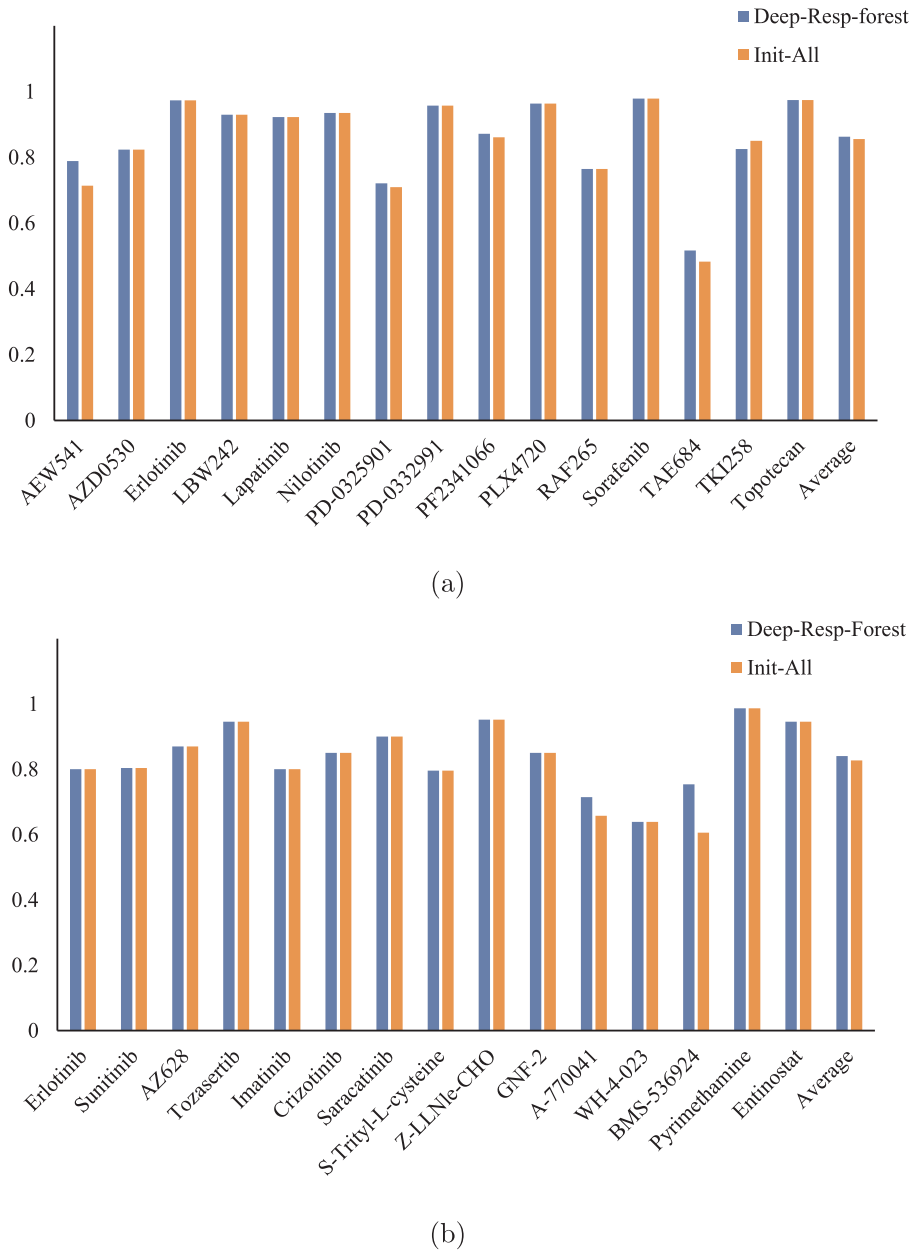


Fig. 8. Performance comparison between the proposed model and the model with all the features concatenated at the beginning. (a) is the result of CCLE and (b) is the result of GDSC. Here we used MIMGS 1 for CCLE and MIMGS 2 for GDSC.

The CCLE has totally 24 drugs and GDSC has 265 drugs. We only kept the drugs which had full gene, CNA and label information. We tested on 15 drugs in CCLE and 15 drugs in GDSC.

3.2. Comparison between two structures in MIMGS stage

We have tried two structures in the MIMGS stage. The MIMGS 1 used m_{GE} sizes of windows and m_{CNA} sizes of windows and generated multiple class vectors for each feature group. Here the sliding window was sized at 100 and 200 so m_{GE} and m_{CNA} both equal to 2. Here we denote the four transformed feature vectors as F_{tGE_100}, F_{tGE_200}, F_{tCNA_100}, F_{tCNA_200}. And they were sent to different layers of the CFFO stage. The MIMGS 2 concatenated all the four feature vectors together and this concatenated feature vector was sent to each level of the CFFO. Here we show the performance using the two structures in Table 1 and Table 2.

From Table 1, it shows that the MIMGS1 performed slightly better than MIMGS 2 for CCLE. 13 out of 15 drugs of MIMGS 1 performed better than or the same with MIMGS 2. For GDSC, shown in Table 2, MIMGS 2 performed better than MIMGS 1. This tells that different MIMGS stage can be chosen for different data, yet not huge different is brought.

Here we also show the receiver operating characteristic (ROC) curve in Fig. 3. For CCLE, the MIMGS 1 curve lay at the left top of the MIMGS2 curve; while for GDSC, the MIMGS 2 curve was at the left and top of the MIMGS 1.

3.3. Comparison with each single type of data

It was reported that the use of heterogenous information in anti-cancer drug response prediction tasks can greatly improve the performance [54]. We used the combination of gene profiles and the copy number alternation in order to classify the drug sensitivity with respect to heterogenous views. Here we compared the performance using the

Table 4
The accuracy comparison between SVM and Deep-Resp-Forest tested on 15 drugs of CCLE.

	AEW541	AZD0530	Erlotinib	LBW242
SVM	0.674	0.824	0.973	0.93
Deep-Resp-Forest	0.789	0.824	0.973	0.930
	Lapatinib	**Nilotinib**	**PD-0325901**	**PD-0332991**
SVM	0.922	0.935	0.733	0.957
Deep-Resp-Forest	0.922	0.935	0.721	0.957
	PF2341066	**PLX4720**	**RAF265**	**Sorafenib**
SVM	0.884	0.963	0.790	0.979
Deep-Resp-Forest	0.871	0.963	0.765	0.979
	TAE684	**TKI258**	**Topotecan**	**Average**
SVM	0.426	0.850	0.975	0.854
Deep-Resp-Forest	0.517	0.825	0.975	0.863

Table 5
The accuracy comparison between SVM and Deep-Resp-Forest tested on 15 drugs of GDSC.

	Erlotinib	Sunitinib	AZ628	Tozasertib
SVM	0.800	0.804	0.885	0.945
Deep-Resp-Forest	0.800	0.804	0.870	0.945
	Imatinib	**Crizotinib**	**Saracatinib**	**S-Trityl-L-cysteine**
SVM	0.800	0.850	0.900	0.796
Deep-Resp-Forest	0.800	0.850	0.900	0.796
	Z-LLNle-CHO	**GNF-2**	**A-770041**	**WH-4-023**
SVM	0.952	0.850	0.619	0.639
Deep-Resp-Forest	0.952	0.850	0.715	0.639
	BMS-536924	**Pyrimethamine**	**Entinostat**	**Average**
SVM	0.742	0.987	0.946	0.834
Deep-Resp-Forest	0.754	0.987	0.946	0.841

GE alone, the CNA alone and the combination of both. Similarly, we constructed the model with single type of input with two window sizes and tested on CCLE data. The results can be found in Fig. 4. From the figure, it shows that the integration of multiple sources of data performs better than using one type of information alone.

3.4. Impact of the MIMGS stage and the use of multiple sizes of windows

It was proved by Zhou et al. that the cascade forest was enhanced by the multi-grained scanning [58]. In our study, similarly, the MIMGS stage provided a high-dimensional feature vector for the CFFO stage. The raw feature vector was greatly augmented. In order to test the impact of MIMGS, we compared the performance using the full model and the model with only CFFO in Fig. 5. The raw features in each feature group were transformed into high-dimensional feature vectors using the MIMGS. This greatly augmented the feature space. Our results show that the MIMGS indeed enhances the cascade forest stage afterwards.

The MIMGS stage scanned the whole feature set with multiple sliding window sizes. We conducted another experiment to show the impact of the multiple sizes of the sliding windows. In each feature group, we used 100-dim and 200-dim windows to scan the raw features. Here we also tested the performance of single dimension and show in Fig. 6. From Fig. 6, we can see that using multiple sized windows achieve the highest performance among the three.

3.5. Comparison with the standard gcForest

We made three modifications to the standard gcForest. Firstly, we incorporated diverse molecular information to enhance the multi-grained scanning stage. Secondly, we constructed two structures of MIMGS based on multi-grained scanning. Lastly, we added feature optimization in the cascade forest in order to capture more discriminative features. Here we also compared the results between the Deep-Resp-Forest with the standard gcForest, which is illustrated in Fig. 7. Our algorithm optimized the input features in each layer and put more emphasis on the discriminative features. From Fig. 7, it shows that a higher accuracy has been achieved but the improvement is limited. Then we tested the running time for the two algorithms and show in Table 3. It shows that Deep-Resp-Forest spent less time than the gcForest, which is more applicable to promote to large-scale data.

3.6. Comparison with features concatenated initially

We used two strategies to combine diverse types of features. The first one is the so-called MIMGS 1, which transforms feature vectors from each feature type and then send to different layers of the cascade forest; The second one is the MIMGS 2, which concatenates all the transformed feature vectors from all the feature groups and feed to each layer of the cascade forest. Some prediction studies simply used the concatenation of diverse features at beginning as input [42]. Here we show the comparison between performance using integration of features initially at the entrance of the model and the proposed strategies in Fig. 8. The two methods have quite similar performance for most of the drugs. But our integration strategy is efficient in discriminating the 'tough cases' which are difficult to separate.

3.7. Comparison with support vector machine

Here we also compared Deep-Resp-Forest with the support vector machine (SVM) [11], which is used in a wide range of applications in pharmaceutical industry [39,19]. The results are shown in Table 4 and Table 5. Here the MIMGS 1 was used in CCLE for comparison and MIMGS 2 was used in GDSC for comparison. From the results, we see that Deep-Resp-Forest performs slightly better than SVM.

4. Conclusions

In our studies, we proposed a deep learning model for the prediction of anti-cancer response. The proposed model was developed based on the gcForest, which was an ensemble-ensemble algorithm. We made three contributions to the two stages of the gcForest respectively. Firstly, in MIMGS stage, instead of single feature input, diverse molecular data can be effectively incorporated into the model to provide more information for the classification. Here we tested using the gene profiles and the copy number alternations. Secondly, all the input features were transformed to class vectors and concatenated into a high-dimensional feature vector. We used two structures to transform feature vectors. One formed the feature vector using each feature group and each window size. The other one formed the feature vector concatenating the feature vectors using all feature groups and all window sizes. Lastly, in the second stage, the CFFO stage, combining with the predicted class vectors generated by the preceding level in the cascade forest stage, these features were used as the input for the current level after going through a feature optimization procedure, which improves the performance efficiency through emphasizing on more important

features. We tested the proposed model on two publicly available data sets, the CCLE and GDSC and we conducted a series number of comparison experiments including the comparison with SVM. The results show quite a high prediction accuracy, which proves the discriminative ability of the proposed model.

In the future, we aim to improve the Deep-Resp-Forest as following: Firstly, more types of data will be incorporated to provide richer information for the classification; Secondly, we would like to improve the deep forest to obtain the exact sensitivity values, which becomes a regression problem; Lastly, the cascade forest stage will be further improved by creating new features to supervise the classification. For example, network based features, which have been widely used in disease gene [52], microRNA [56,60], circleRNA [53], can be added for prediction of anti-cancer response.

5. Funding

The work was supported by the National Key R & D Program of China (SQ2018YFC090002), the National Natural Science Foundation of China (Nos. 61701340, 61702361, and 61771331) and Natural Science Foundation of Tianjin (Nos. 18JCQNJC00800 and 18JCQNJC00500).

References

[1] N. Aben, D.J. Vis, M. Michaut, L.F. Wessels, Tandem: a two-stage approach to maximize interpretability of drug response models based on multiple molecular data types, Bioinformatics 32 (2016) i413–i420.
[2] A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina, A. Zhavoronkov, Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data, Mol. Pharm. 13 (7) (2016) 2524–2530.
[3] M. Ammad-ud-din, E. Georgii, M. Gönen, T. Laitinen, O. Kallioniemi, K. Wennerberg, A. Poso, S. Kaski, Integrative and personalized QSAR analysis in cancer by kernelized bayesian matrix factorization, J. Chem. Inf. Model. 54 (8) (2014) 2347–2359.
[4] M. Asada, M. Miwa, Y. Sasaki, Extracting drug-drug interactions with attention CNNS, Proc. BioNLP 2017 (2017) 9–18.
[5] M. Bahi, M.C. Batouche, Drug-target interaction prediction in drug repositioning based on deep semi-supervised learning, Computational Intelligence and Its Applications. CIIA, IFIP Advances in Information and Communication Technology vol. 522, Springer, Cham, 2018, pp. 302–313.
[6] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A.A. Margolin, S. Kim, C.J. Wilson, J. Lehár, G.V. Kryukov1, D. Sonkin, A. Reddy, M. Liu, L. Murray, M.F. Berger, J.E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F.A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I.H. Engels, J. Cheng, G.K. Yu, J. Yu, P.A. Jr, M. de Silva, K. Jagtap, M.D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R.C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J.P. Mesirov, S.B. Gabriel, G. Getz, K. Ardlie, V. Chan, V.E. Myer, B.L. Weber, J. Porter, M. Warmuth, P. Finan, J.L. Harris, M. Meyerson, T.R. Golub, M.P. Morrissey, W.R. Sellers, R. Schlegel, L.A. Garraway, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity, Nature 483 (2012) 603–607.
[7] S. Basith, B. Manavalan, T.H. Shin, G. Lee, iGHBP: computational identification of growth hormone binding proteins from sequences using extremely randomised tree, Comput. Struct. Biotechnol. J. 16 (2018) 412–420.
[8] L. Breiman, Stacked regressions, Mach. Learn. 24 (1) (1996) 49–64.
[9] L. Breiman, Random forests, Mach. Learn. 45 (1) (2005) 5–32.
[10] Y. Chang, H. Park, H.-J. Yang, S. Lee, K.-Y. Lee, T.S. Kim, J. Jung, J.-M. Shin, Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from Cancer Genomic Signature, Sci. Rep. 8 (1) (2018).
[11] C. Cortes, V.N. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[12] J. Cui, Y. Chen, W.-C. Chou, L. Sun, L. Chen, J. Suo, Z. Ni, M. Zhang, X. Kong, L.L. Hoffman, J. Kang, Y. Su, V. Olman, D. Johnson, D.W. Tench, I.J. Amster, R. Orlando, D. Puett, F. Li, Y. Xu, Personalized medicine and cancer, Nucleic Acids Res. 39 (4) (2011) 1197–1207.
[13] F.-Y. Dao, H. Yang, Z.-D. Su, W. Yang, Y. Wu, D. Hui, W. Chen, H. Tang, H. Lin, Recent advances in conotoxin classification by using machine learning methods, Molecules 22 (7) (2017).
[14] D.L. Dexter, J.T. Leith, Tumor heterogeneity and drug resistance, J. Clin. Oncol. 4 (2) (1986) 244–257.
[15] Y. Ding, J. Tang, F. Guo, Identification of drug-target interactions via multiple information integration, Inf. Sci. 418–419 (2017) 546–560.
[16] Y. Ding, J. Tang, F. Guo, Identification of drug-side effect association via semi-supervised model and multiple kernel learning, IEEE J. Biomed. Health Inf. 1 (2018) 1-1.
[17] Y. Ding, J. Tang, F. Guo, Identification of drug-side effect association via multiple information integration with centered kernel alignment, Neurocomputing 325 (2019) 211–224.
[18] Z. Dong, N. Zhang, C. Li, H. Wang, Y. Fang, J. Wang, X. Zheng, Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection, BMC Cancer 15 (489) (2015).
[19] M.J. Garnett, E.J. Edelman, S.J. Heidorn, C.D. Greenman, A. Dastur, K.W. Lau, P. Greninger, I.R. Thompson, X. Luo, J. Soares, Q. Liu, F. Iorio, D. Surdez, L. Chen, R.J. Milano, G.R. Bignell, A.T. Tam, H. Davies, J.A. Stevenson, S. Barthorpe, S.R. Lutz, F. Kogera, K. Lawrence, A. McLaren-Douglas, X. Mitropoulos, T. Mironenko, H. Thi, L. Richardson, W. Zhou, F. Jewitt, T. Zhang, P. OBrien, J.L. Boisvert, S. Price, W. Hur, W. Yang, X. Deng, A. Butler, H.G. Choi, J.W. Chang, J. Baselga, I. Stamenkovic, J.A. Engelman, S.V. Sharma, O. Delattre, J. Saez-Rodriguez, N.S. Gray, J. Settleman, P.A. Futreal, D.A. Haber, M.R. Stratton, S. Ramaswamy, U. McDermott, C.H. Benes, Systematic identification of genomic markers of drug sensitivity in cancer cells, Nature 483 (7391) (2012) 570–575.
[20] L.A. Garraway, Genomics-driven oncology: framework for an emerging paradigm, J. Clin. Oncol. 31 (15) (2013) 1806–1814.
[21] A. Graves, A. Mohamed, G.E. Hinton, Speech recognition with deep recurrent neural networks, Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013, pp. 6645–6649.
[22] Y. Guo, S. Liu, Z. Li, X. Shang, Bcdforest: a boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data, BMC Bioinf. 19 (Suppl 5) (2018) 118.
[23] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[24] W. Hwang, J. Choi, M. Kwon, D. Lee, Context-specific functional module based drug efficacy prediction, BMC Bioinf. 17 (Suppl 6) (2016).
[25] I.S. Jang, E.C. Neto, J. Guinney, S.H. Friend, A.A. Margolin, Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data, Pacific Symposium on Biocomputing (2014) 63–74.
[26] M.J. Joyner, N. Paneth, Seven questions for personalized medicine, JAMA 314 (10) (2015) 999–1000.
[27] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 1 (2017) 1097–1105.
[28] J. Lanchantin, A. Sekhon, R. Singh, Y. Qi, Prototype matching networks for large-scale multi-label genomic sequence classification, 2017. arXiv preprint arXiv:1710.11238.
[29] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[30] Z. Li, J. Tang, F. Guo, Learning from real imbalanced data of 14-3-3 proteins binding specificity, Neurocomputing 217 (2016) 83–91 sI: ALLSHC.
[31] B. Manavalan, T.H. Shin, G. Lee, Dhspred: support-vector-machine-based human dnase i hypersensitive sites prediction using the optimal features selected by random forest, Oncotarget 9 (2018) 1944–1956.
[32] B. Manavalan, T.H. Shin, G. Lee, PVP-SVM: sequence-based prediction of phage virion proteins using a support vector machine, Front. Microbiol. 9 (2018) 476.
[33] J.M. Mariadason, D. Arango, Q. Shi, A.J. Wilson, G.A. Corner, C. Nicholas, M.J. Aranes, M. Lesser, E.L. Schwartz, L.H. Augenlicht, Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin, Cancer Res. 63 (24) (2003) 8791–8812.
[34] A. Mayr, G. Klambauer, T. Unterthiner, S. Hochreiter, DeepTox: toxicity prediction using deep learning, Front. Environ. Sci. (2016).
[35] S. Rh, The NCI60 human tumour cell line anticancer drug screen. nature reviews cancer, Nat. Rev. Cancer 6 (10) (2006) 813–823.
[36] A. Sekhon, R. Singh, Y. Qi, Deepdiff: deep-learning for predicting differential gene expression from histone modifications, Bioinformatics 34 (17) (2018) i891–i900.
[37] C. Shen, Y. Ding, J. Tang, X. Xu, F. Guo, An ameliorated prediction of drug-target interactions based on multi-scale discrete wavelet transform and network features, Int. J. Mol. Sci. 18 (8) (2017) 1781.
[38] J.E. Staunton, D.K. Slonim, H.A. Coller, P. Tamayo, M.J. Angelo, J. Park, U. Scherf, J.K. Lee, W.O. Reinhold, J.N. Weinstein, J.P. Mesirov, E.S. Lander, T.R. Golub, Chemosensitivity prediction by transcriptional profiling, Proceedings of the National Academy of Sciences of the United States of America 98 (19) (2001) 10787–10792.
[39] R. Su, H. Wu, B. Xu, X. Liu, L. Wei, Developing a multi-dose computational model for drug-induced hepatotoxicity prediction based on toxicogenomics data, IEEE/ACM Trans. Comput. Biol. Bioinf. (2018).
[40] Z.-D. Su, Y. Huang, Z.-Y. Zhang, Y.-W. Zhao, D. Wang, W. Chen, K.-C. Chou, H. Lin, iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC, Bioinformatics 34 (24) (2018) 4196–4204.
[41] H. Tang, Y.-W. Zhao, P. Zou, C.-M. Zhang, R. Chen, P. Huang, H. Lin, HBPred: a tool to identify growth hormone-binding proteins, Int. J. Biol. Sci. 14 (8) (2018) 957–964.
[42] T. Urushidani, A. Sedykh, E. Muratov, V. Kuz'min, D. Fourches, Y. Low, T. Uehara, Y. Minowa, H. Yamada, Y. Ohno, H. Zhu, I. Rusyn, A. Tropsha, Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches, Chem. Res. Toxicol. 24 (8) (2011) 1251–1262.
[43] M. Verma, Personalized medicine and cancer, J. Pers. Med. 2 (1) (2012) 1–14.
[44] L. Wei, H. Chen, R. Su, M6apred-el: a sequence-based predictor for identifying n6-methyladenosine sites using ensemble learning, Mol. Ther. – Nucleic Acids 12 (2018) 635–644.
[45] L. Wei, R. Su, B. Wang, X. Li, Q. Zou, X. Gao, Integration of deep feature representations and handcrafted features to improve the prediction of n6-methyladenosine sites, Neurocomputing 324 (2019) 3–9.
[46] L. Wei, S. Wan, J. Guo, K.K. Wong, A novel hierarchical selective ensemble classifier with bioinformatics application, Artif. Intell. Med. 83 (2017) 82–90.
[47] L. Wei, P. Xing, J. Zeng, J. Chen, R. Su, F. Guo, Improved prediction of protein-protein interactions using novel negative samples, features, and an ensemble

classifier, Artif. Intell. Med. 83 (2017) 67–74.
[48] L. Wei, C. Zhou, H. Chen, J. Song, R. Su, Acpred-fl: a sequence-based predictor based on effective feature representation to improve the prediction of anti-cancer peptides, Bioinformatics (2018).
[49] M. Wen, Z. Zhang, S. Niu, H. Sha, R. Yang, Y. Yun, H. Lu, Deep-learning-based drug-target interaction prediction, J. Proteome Res. 16 (4) (2017) 1401–1409.
[50] Y. Xu, J. Pei, L. Lai, Deep learning based regression and multiclass models for acute oral toxicity prediction with automatic chemical feature extraction, J. Chem. Inf. Model. 57 (11) (2017) 2672–2685.
[51] W. Yang, J. Soares, P. Greninger, E.J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J.A. Smith, I.R. Thompson, S. Ramaswamy, P.A. Futreal, D.A. Haber, M.R. Stratton, C. Benes, U. McDermott, M.J. Garnett, Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells, Nucleic Acids Res. 41 (D1) (2013) 955–961.
[52] X. Zeng, N. Ding, A. Rodríguez-Patón, Q. Zou, Probability-based collaborative filtering model for predicting gene-disease associations, BMC Med. Genomics 76 (10) (2017) 45–53.
[53] X. Zeng, W. Lin, M. Guo, Q. Zou, A comprehensive overview and evaluation of circular rna detection tools, PLoS Comput. Biol. 13 (6) (2017) e1005420 .
[54] F. Zhang, M. Wang, J. Xi, J. Yang, A. Li, A novel heterogeneous network-based method for drug response prediction in cancer cell lines, Sci. Rep. (2018) 1–9.
[55] N. Zhang, H. Wang, Y. Fang, J. Wang, X. Zheng, X.S. Liu, Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model, PLoS Comput. Biol. 11 (9) (2015) 1–11.
[56] X. Zhang, Q. Zou, A. Rodríguez-Patón, X. Zeng, Probability-based collaborative filtering model for predicting gene-disease associations, EEE/ACM Trans. Comput. Biol. Bioinf. (2018).
[57] Y. Zhang, W. Chan, N. Jaitly, Very deep convolutional networks for end-to-end speech recognition, Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 4845–4849.
[58] Z.-H. Zhou, J. Feng, Deep forest: Towards an alternative to deep neural networks, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17), 2017, pp. 3553–3559.
[59] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc., Series B (2005) 301–320.
[60] Q. Zou, J. Li, L. Song, X. Zeng, G. Wang, Similarity computation strategies in the microRNA-disease network: a survey, Briefings Funct. Genomics 15 (1) (2015) 55–64.