

# The Drug Response Prediction for Cancer Patients' Personalized Treatment by Deep Learning on Omics Features

1<sup>st</sup> Zeyu Yao

*School of Computer Science and Technology  
Hainan University  
Haikou, China  
22220854000294@hainanu.edu.cn*

2<sup>nd</sup> Ruhao Liu

*School of Computer Science and Technology  
Hainan University  
Haikou, China  
22220854000297@hainanu.edu.cn*

3<sup>rd</sup> Qingchen Zhang\*

*School of Computer Science and Technology  
Hainan University  
Haikou, China  
zhangqingchen@hainanu.edu.cn  
\* corresponding author*

**Abstract—**Predicting the response of tumors to cancer treatment drugs is currently an important challenge for personalized cancer treatment. In recent years, deep learning has been widely applied in the field of drug reaction prediction. Most of existing research for drug response are based on cancer patients' omics features. This article summarizes the research on the application of deep learning(DL) method to predict drug reactions in cancer cell lines. We provide a brief overview of the multiple sets of data currently used for drug reaction prediction, as well as an introduction to the publicly available datasets used in current research experiments. Then, we will provide a detailed introduction to some models and methods for predicting drug reactions based on deep learning. Finally, we provide suggestions for future work in response to the problems in our research methods.

**Index Terms—**drug response prediction, cancer patients' personalized treatment,omics features,deep learning

## I. INTRODUCTION

Traditional cancer treatment methods are usually based on the patient's pathological type and cancer staging. However, due to differences in physical conditions among different patients, cancer patients with the same pathological type and stage may have different reactions to the same treatment plan, even at different stages of the same patient. Therefore, traditional treatment methods often cannot meet the needs of all patients. Personalized cancer treatment aims to develop more suitable treatment plans for different patients, improving their survival and quality of life. With the rapid development of genomics and bioinformatics, people are beginning to realize the importance of developing personalized treatment plans.

Drug response prediction is closely related to personalized cancer treatment. Drug reactivity refers to the physiological or pathological changes that occur when a patient takes

medication, different patients may have different reactions to the same drug. Personalized treatment is a treatment method based on individual pathological and physiological characteristics. It can develop personalized treatment plans for patients according to molecular information such as genomics and transcriptome of patients to achieve the best treatment effect. Therefore, understanding the patient's drug responsiveness is crucial for developing personalized treatment plans.

Some existing studies use machine learning for drug response prediction. For example, kESVR [1] takes mRNA data from CCLE as input, first reduces the dimensionality of gene expression data, clusters the reduced low-dimensional data according to the given drug response values, and then trains machine learning models in different cluster data to obtain a score. Finally, drug response prediction is performed. KESVR was compared with 12 other existing methods, and the experimental results showed that kESVR was superior to other methods. kESVR cannot perform feature selection on gene expression data and does not use other omics data, reflecting the limitations of machine learning in drug prediction. Deep learning models are suitable for predicting drug reactions. Drug response prediction requires a large amount of data, including drug molecular structure, gene expression data, protein sequence, etc. Deep learning models can effectively process large-scale data [2], extract features from the data, and learn complex patterns from it. Drug reactions also involve complex nonlinear relationships, and deep learning models can automatically learn these relationships and produce high-precision prediction results. At present, some DL models have been effectively applied to tasks related to drug response prediction. For example, convolutional neural network (CNN) is mainly used to process the data of drug molec-

ular structure, which can automatically extract the features in the molecule and predict its efficacy. Graph convolutional neural networks (GCN) can improve the performance of drug response prediction by learning drug features. Autoencoder (AE) and variational autoencoder (VAE) can achieve data dimensionality reduction, data compression, and more effectively visualize high-dimensional data. For example, Auto-HMM-LMF [3] uses an autoencoder to extract important omics data features, which reduces the dimensionality of the original data while improving the model's generalization ability. Nevertheless, there are still many areas for improvement in accurately extracting omics features and ensuring the accuracy of prediction results for these DL models.

In this article, we summarize some models and methods for predicting drug reactions. Firstly, we introduced the types of multi-omics data currently used in existing research, and summarized the publicly available datasets of these multi-omics data used in the experiment. Then, we have provided a specific introduction to representative models and methods in both mono-omics data and multi-omics data. Next, we summarized the models and methods used for predicting drug reactions, identified the existing problems, and provided our suggestions. Through these studies, it can be beneficial for the development of personalized cancer treatment.

## II. OMICS DATA

Through previous research, it has been demonstrated that certain mono-omic data of patients, such as gene expression data, have a certain role in drug response experiments. However, the prediction results obtained solely from mono-omic data are not ideal. In recent years, many studies have been continuously improving data integration methods to improve the accuracy of predictions. Data integration can reduce data dimensions and reduce data noise, thereby improving data quality. Therefore, integrating multi-omics data can improve the accuracy and reliability of drug response prediction. The most widely used multi-omics data in the field of drug response prediction currently include gene mutations, copy number alteration, and gene expression data. These multi-omics data can describe the patient's genomic information and help determine the patient's response to certain drugs.

### A. Common omics data

Although other types of omics data, such as proteomics, immunology and metabolomics, can also play a role in drug response prediction, compared with gene mutation, copy number alteration and gene expression data, their data processing is difficult and subject to certain technical constraints, so they have not been widely used. This chapter introduces the three omics data used in current research.

Gene mutation (Mut) refers to the change of the composition or arrangement of base pairs in the molecular structure of gene molecules, which leads to the change of protein structure or function. The genes produced after genetic mutations become mutated genes. Currently, in the medical field, gene mutation data is widely used for disease diagnosis,

predicting drug reactions, risk assessment, and more. The current methods used in the medical field to extract gene mutations mainly include single gene detection, whole exon sequencing, and targeted sequencing. These technologies can obtain different gene mutation information, such as gene rearrangement, single nucleotide variation, etc. At present, when processing gene mutation data in the computer field, data preprocessing is generally performed to extract different types of features required, such as structural domains, DNA sequences, and amino acid sequences. The main preprocessing processes include quality control, sequence alignment, mutation annotation, etc. In the pre-processing part of gene mutation data in existing research, we will select those point mutation that affect the molecular structure, and then filter out the unresponsive point mutation [4].

Copy number alteration (CNA) is a phenomenon caused by genome rearrangement, which generally refers to the occurrence of multiple copies increasing or decreasing in the genome, mainly manifested as deletion and duplication at the submicroscopic level. In the field of medicine, copy number alteration data is widely used for predicting drug reactions, diagnosing genetic diseases, and evaluating tumor risk. The main methods used to extract copy number alteration data in the current medical field are based on microarray chips and next-generation sequencing technology. When using copy number alteration data for data analysis, it is necessary to preprocess and extract features from the original copy number alteration data. At present, when processing copy number alteration data in the computer field, the process may include noise filtering, GC correction, standardization, etc. The extracted copy number alteration data features can be selected based on different research purposes, such as the frequency and length of CNAs. In the preprocessing section of copy number alteration data in existing studies, gene level copy values are binarized and numbered before being input into the model [5].

Gene expression refers to the process of transcribing genes into RNA molecules and translating RNA molecules into biologically active protein molecules. In the field of medicine, gene expression data is widely used to study the discovery of biomarkers, the pathogenesis of diseases, and disease diagnosis. The current medical field mainly uses RNA sequencing technology to extract gene expression data. Before inputting gene expression data into the model, preprocessing and feature extraction are usually performed. When preprocessing gene expression data, data cleansing is usually performed to filter low expression gene data. Because the expression measurement methods for gene expression data in different datasets are different, the gene expression data is standardized and paired homogenized. At the same time, gene expression data will also be standardized, such as base number standardization, RPKM standardization, etc, which can make gene expression data comparable [6].

## B. Dataset

The gene mutation data, copy number variation and gene expression data used in multiple current research methods mainly come from publicly available datasets such as the Genomics of Drug Sensitivity in Cancer (GDSC) [7], the Cancer Treatment Response Portal (CTRP) [8], the Cancer Cell Line Encyclopedia (CCLE) [8], and The Cancer Genome Atlas (TCGA) [9]. CCLE collects genomic data from over 1000 human tumor cell lines. GDSC contains a large amount of data on drug sensitivity of tumor cell lines and provides relevant genomic data. CTRP contains a large amount of data on the drug response of tumor cell lines. These publicly available datasets provide a wide range of genomic data for existing research, which can help promote the development of medicine. Among them, GDSC, CTRP, and CCLE provide some drug reaction data that can be used for research on drug reaction prediction. TCGA contains clinical information and corresponding genomic data of patients with multiple types of cancer. When conducting experiments for predicting drug reactions, attention should also be paid to the differences between in vivo and in vitro experimental datasets, and the separation of in vivo and in vitro datasets is crucial. In vivo experiments, TCGA datasets are often used to better understand the drug response effects due to the need for data closer to the real biological environment. In vitro experiments, it may be necessary to screen a large number of compounds to test the effectiveness of candidates, and sometimes to verify some drug toxicity or side effects. Therefore, GDSC, CCLE, and CTRP data are usually used.

Existing research methods can explore the relationship between tumor cell genomic data and drug response by processing and analyzing these genomic data. The combination of multi-omics data and deep learning can develop drug response prediction models and promote the development of personalized treatment for cancer.

## III. MODELS

In recent years, more and more research has applied deep learning to the field of drug response prediction, such as convolutional neural networks (CNN) and graph convolutional neural networks (GCN). Deep learning has multi-layer neural networks, which can display the features of input data through intuitive visualization methods when processing high-dimensional data such as gene expression profiles [10]. While capturing complex relationships in the data, it can also learn meaningful feature representations from the original data, reducing the complexity of data processing, increasing the interpretability of the data, and improving prediction accuracy. With the continuous development of deep learning, the effectiveness of deep learning models applied in the field of drug reaction prediction is becoming increasingly accurate. This article mainly divides the data types used in existing drug reaction prediction models into the following two categories: mono-omics data and multi-omics data. Table 1 summarizes

some existing drug reaction prediction models, and we have introduced these methods.

### A. Mono-omics data

In previous studies, due to the limited amount of publicly available omics data and inadequate feature extraction techniques, drug response models only utilized patient mono-omics features and combined with some other information for prediction, such as SMILES encoding and compound fingerprints. A drug response prediction model based on mono-omics data usually inputs patient gene expression data into the model to predict the response value of each patient to the drug. The approximate working process of mono-omics data prediction drug response model we mentioned is shown in Figure 1.

Geeleher et al. [11] proposed a method to predict the patient's response to chemotherapy drugs based on the baseline gene expression data before treatment. The baseline gene expression data in the patient's cell line and the patient's oncogene expression data are used as inputs, and the output is to predict the patient's response to the drug (IC<sub>50</sub>). Through the method proposed in this paper, the author has proved that the number of drug responders can be increased by using only the baseline oncogene expression data. This literature has validated the reliability of this method through different cross experiments, which has a certain driving effect on personalized healthcare.

DeepDSC [12] is a deep neural network that extracts genomic features from cell line expression data using a deep automatic encoder, combined with compound fingerprints as input, to output the final response data of the cell line to drugs. The cell line expression data and drug sensitivity data in this method were obtained from CCLE and GDSC. The root mean square error (RMSE) results obtained after conducting a 10 fold cross validation experiment indicate that DeepDSC can fill in missing drug sensitivity values and has the best ability, proving that DeepDSC is very helpful for future clinical cancer drug sensitivity prediction research.

PaccMann [13] is a multimodal attention neural network used to predict the sensitivity of anticancer drugs, encoding gene expression profiles using attention mechanisms and encoding SMILES using encoders. The encoded gene expression profile and SMILES constitute a drug cell pair to input into the model, and the output is IC<sub>50</sub> sensitivity value. A 25 fold cross validation experiment was conducted with six models to test the performance of the model. The results showed that the performance of the model can surpass the baseline model, proving that the attention based encoder can more effectively extract drug features for drug reaction prediction.

RefDNN [14] is a deep neural network model used to predict the response of anticancer drugs. The model takes gene expression data and drug molecular structure of cell lines as inputs, and outputs drug responses in a given cell line. The RefDNN model can effectively predict drug characteristics by referencing drugs, and can predict untrained expression data

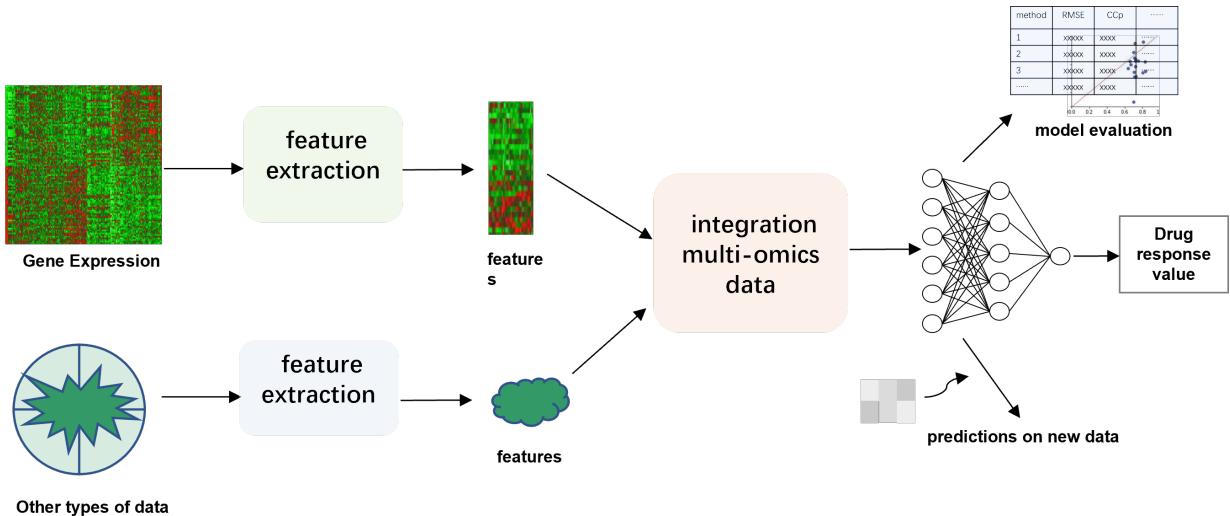


Fig. 1. Mono-omics data model process

and drugs based on similarity. RefDNN model uses ElasticNet regularization to process high-dimensional gene expression data from GDSC and CCLE. Comparative experiments were conducted between RefDNN and five classification models, and RefDNN showed good predictive accuracy. RefDNN has also made contributions in identifying drug resistance biomarkers.

Although some existing studies have proved that mono-omics data has a certain role in drug response prediction, mono-omics data is prone to model overfitting problems due to insufficient data and low data dimensions, which leads to unstable performance and inaccurate results of drug response prediction.

### B. Multi-omics data

The multi-omics data of cancer patients combines data from different levels and types of patients, which can more comprehensively reflect the patient's tumor progression and drug treatment response. Gene mutation, copy number variation, and gene expression data, which are commonly used omics data in the field of drug response prediction, help to understand the characteristics of tumors, facilitate doctors to provide precise treatment, design effective treatment plans that are more suitable for different patients, and promote the development of personalized cancer treatment. The approximate working process of predicting drug response models using multi-omics data we mentioned is shown in Figure 2.

MOLI [15] is a representative method in the field of drug response prediction, and it is the first method to use combined loss function in deep neural network. MOLI takes gene expression data, somatic mutation and copy number alteration as input after feature coding through encoder, and inputs classifier after integration of multi omics features, and finally outputs drug response value. This method constructs a three tuple loss function based on the multigroup data of a given drug response cell line and the multigroup data

of patients who do not respond to a given drug, and constructs a combined loss function combined with the binary cross entropy classification loss function used by classifier. Through experimental comparison with other methods, the combined loss function used by MOLI can achieve better results than the early integration method of multi-omics data. Meanwhile, compared to other deep neural network-based monoomics methods, MOLI has a higher AUC index and better predictive performance. The multi omics integration method and combined loss function proposed by MOLI have good performance in drug response prediction. This method provides a novel and feasible method for the field of drug response prediction, and provides a promising direction for personalized treatment. Afterwards, many methods were improved on the basis of MOLI and achieved good results in predicting drug reactions.

DeepDR [16] is a drug response prediction model that includes three deep neural networks (DNNs). This model takes tumor mutation and expression data as inputs, and outputs the tumor's response to drugs. The model first extracts high-dimensional mutation data and expression profiles from TCGA through two automatic encoders. Then, a prediction network is used to integrate two omics data and predict drug reactions. This model compared with other models, and the experimental results showed a mean square error of 1.96, which is superior to previous classical methods in terms of prediction error and stability. This model has contributed to the translation of pharmacogenomics characteristics of cell lines and promoted the development of tumor drug response prediction. At the same time, the multi-omics integration method proposed by the model is also applicable to other omics features, which is beneficial for multi-omics data models to predict drug reactions and open up a new path for personalized cancer treatment.

DrugCell [17] is a model that combines visible neural

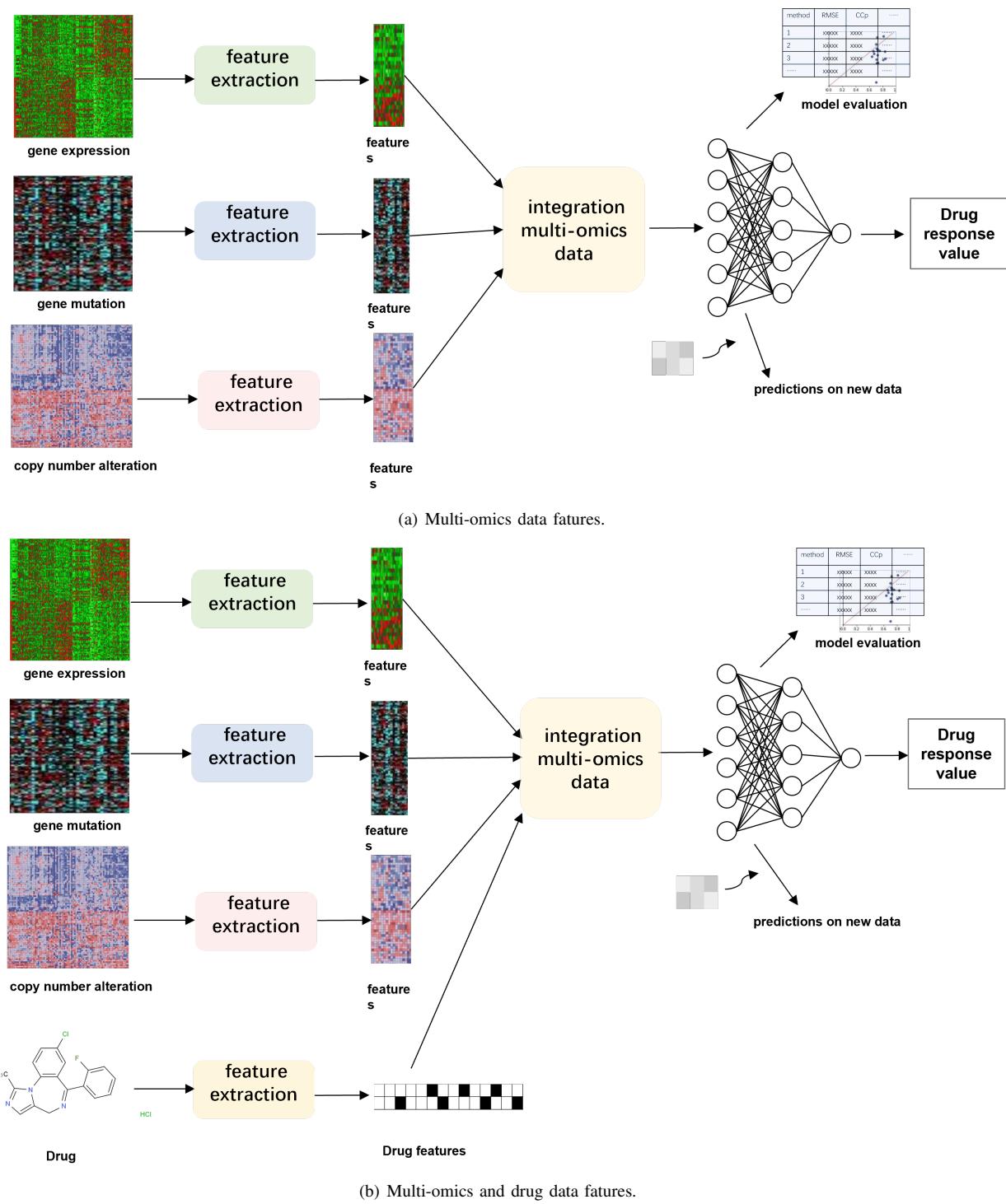


Fig. 2. Multi-omics data model process.

networks (VNN) with artificial neural networks (ANN). VNN encodes genotype data, and ANN embeds the structure of drug compounds. This model takes genotype data and drug compounds as inputs, and puts the processed genotype data and drug compounds into the prediction module, ultimately outputting the drug's response value to cells. In experimental testing, the model successfully predicted the reactions of some drugs to cell lines, demonstrating its application value in predicting drug reactions.

DCNN-DR [18] is a model based on convolutional neural networks and multi omics data to predict drug efficacy. This model uses nearest component analysis (NCA) to select features from CNV, methylation, gene expression, and mutation binary data as input and output drug response values of TCGA-BC patients. The accuracy and reliability of the model were tested by experiments in real breast cancer patient samples, which proved that the performance of the model integrating multi-omics data was better than other models using mono-omics data as input. Combining deep learning with multi omics data feature selection can achieve high-precision classification, which can help doctors develop more reasonable patient treatment plans.

GraTransDRP [19] is a Graph Transformer based on deep learning. This model uses three omics data, namely gene expression (GE), mutation(MUT) and copy number distortion (CNA), methylation (METH), and drug molecular maps, as input to the model. Reduce the dimensionality of GE using KernelPCA, extract features from MUT and CNA, METH using convolutional layers, and extract features from drug molecular maps using Graph Transformer. After splicing the characteristics of three cell lines and drug features, the drug response values are output through two fully connected layers. The experimental section compared the model with GraphDRP and GraOmicDRP, and the experimental results showed that the model had the best results in root mean square error (RMSE) and Pearson correlation coefficient. This model has been improved on the basis of existing research by combining multi-omics data with drug molecular maps, and extracting drug features through Graph Transformer, thereby improving the predictive performance of the model. It is worth noting that introducing new types of data can easily bring redundant information, and it is necessary to choose appropriate dimensionality reduction techniques for feature extraction in order to achieve better data performance after preprocessing.

TCR [20] is a transformer based deep learning model. The model takes drug molecules and multi-Omics data as inputs. After the input, the drug molecular structure uses graph convolution network to extract features. Multi-omics data uses an encoder network to extract features. Use a transformer network to interact omics data information with drug structure, and finally output drug response values through a prediction network. In the experimental section, the performance of TCR was compared with the baseline, and the experimental results showed that TCR had a lower error rate and higher prediction

accuracy. This study demonstrates that the TCR model is a promising method for predicting cancer drug responses.

Super.FELT [21] first selected the features of gene expression data, somatic mutation and CNA data, and then used the multi-omics data as the input of the model after feature coding through the supervised encoder. Next, after integrating the multi-omics features, it was input into the classifier based on neural network, and finally output the drug response value. Super.FELT uses triple loss function (SET) in the supervised encoder to extract features, which is conducive to achieve high translatability. This model has been compared with other 8 different methods in experiments, proving that it has better predictive performance, and the model has performed well in both internal testing and external validation experiments.

Swnet [22] takes gene mutation, gene expression and medicinal chemistry structure as inputs. Use convolutional neural networks to extract important features of gene mutations and gene expression, and use graph neural networks to encode the chemical structure of drugs. Update weight parameters using self attention through the similarity matrix between drugs. Finally, concatenate multiple omics vectors and drug vectors and predict the IC50 value through CNN. The omics data used in this model are from GDSC and CCLE. Compared with the four previously proposed methods, Swnet demonstrated better model performance in terms of R2 and mean square error (MSE). The existing methods still have some shortcomings. Swnet can more effectively extract effective features from a large number of gene expression data, and prove that the combination of gene mutation, gene expression and medicinal chemistry structure features can effectively improve the performance of drug response prediction models.

In the field of drug reaction prediction, models based on multi omics data are currently the most commonly used. In addition to the models listed above, there are many other monoomics based drug reaction prediction models. These models are constantly improving feature extraction and data integration methods, and the effectiveness of the model has been demonstrated through experiments.

TABLE I  
SOME EXISTING DRUG REACTION PREDICTION MODELS

Method terms	Dataset Source	Data type
Geeleher et al. [11]	ArrayExpress	Mono-omics
DeepDSC [12]	GDSC,CCLE	Mono-omics
PaccMann [13]	GDSC	Mono-omics
RefDNN [14]	GDSC,CCLE	Mono-omics
MOLI [15]	GDSC,PDX,TCGA	Multi-omics
DeepDR [16]	CCLE,TCGA	Multi-omics
DrugCell [17]	GDSC,CTRP,CCLE	Multi-omics
DCNN-DR [18]	TCGA	Multi-omics
GraTransDRP [19]	GDSC,CCLE	Multi-omics
TCR [20]	GDSC,CCLE,PRISM,TCGA	Multi-omics
Super.FELT [21]	GDSC,CCLE,CTRP,PDX,TCGA	Multi-omics
Swnet [22]	GDSC,CCLE	Multi-omics

#### IV. DISCUSSIONS

At present, most drug reaction prediction models combine omics data with drug data or other data through advanced deep learning methods, extract important data information through feature extraction methods, and integrate different types of feature data before drug prediction. These methods can greatly improve the accuracy of model prediction and ensure the reliability of the results to a certain extent. However, there are still some problems with existing methods that need to be continuously improved.

Omics data is usually high-dimensional and has a certain degree of complexity. At present, the amount of open omics data is limited. However, deep learning methods usually need to use a large number of omics data to train the model to get a better effect. Using a small amount of omics data for training can easily lead to overfitting and other problems. The current drug reaction prediction models mainly adopt the following methods to solve the problem of insufficient data volume: data augmentation, data conversion, and data integration. Malik et al. [23] proposed a method using random forest to generate new data samples by sampling from the original multi group data set, realizing the purpose of expanding the number of multi group data. Liu et al. [24] proposed a method to convert the interaction relationship between drug compounds and cancer cell line omics features into images. Then, convolutional neural networks were used to extract different features of the image and concatenate them into a representation matrix, converting the original data into different forms of data, thereby increasing the amount of data. DRIM [25] integrates multiple omics data, expanding the number of data while reducing data complexity, enabling better analysis of the interactions between drug omics data.

In drug response research, it is necessary to explain the predictive results of the model clearly in order to determine the impact of certain biological characteristics on the disease and avoid risks in clinical practice. The current proposed drug reaction prediction models mainly use explanatory algorithms to improve the interpretability of the model. Deep-Resp-Forest [26] places different data into multiple decision trees and converts them into molecular feature vectors. The decision tree generates easily understandable classification rules, improving the interpretability of the model. Ammad et al. [27] proposed a method for modeling sample data from different perspectives using linear regression, which can more accurately represent the relationships between different features. The application of linear regression in this model makes it more interpretable.

In future work on drug reaction prediction, the first step should be to improve the quality of multi-omics data as much as possible, and to avoid issues such as data missing and duplication when splicing or fusing multi-omics data. Choosing an appropriate feature extraction method is also crucial, as the stability and representativeness of the extracted multi-omics data features will directly affect the predictive performance of the model [28]. In addition, the entire drug

reaction prediction process can also be demonstrated through interactive interpretation, visualization and other technologies, which can make the model more reliable and provide a basis for clinical drug reaction decision-making.

#### V. CONCLUSION

In this article, we have provided a brief introduction to the most widely used multi omics data in the field of drug reaction prediction, as well as the publicly available datasets commonly used in experiments. In the model section, we first briefly summarized several methods for predicting drug reactions using mono-omics data. Although existing mono-omics data models can play a certain predictive role, the drug response results of patients are usually influenced by multiple factors, and using only mono-omics data has limitations. The subsequent series of drug reaction prediction models based on multi-omics data have been improved in terms of accuracy, interpretability, and reliability. Many existing methods also combine omics data with other types of data, taking into account multiple factors that affect drug response results more comprehensively, and can obtain more accurate results.

Although there are still many areas for improvement in existing drug reaction prediction models, with the continuous progress of technology, there is still great room for development in the field of drug reaction prediction. Accurate drug response prediction results are conducive to the promotion of personalized treatment for cancer patients, and can propose the most suitable quality plan for the individual differences of cancer patients. These research methods for predicting drug reactions have contributed to the practice of personalized treatment in clinical practice.

#### ACKNOWLEDGMENT

This work is supported by two grants with No. 62162023 and No. KYQD(ZR)-21079.

#### REFERENCES

- [1] A. Majumdar, Y. Liu, Y. Lu, S. Wu, and L. Cheng, "kesvr: An ensemble model for drug response prediction in precision medicine using cancer cell lines gene expression," *Genes*, vol. 12, no. 6, p. 844, 2021.
- [2] H. Huang, M. Lin, L. T. Yang, and Q. Zhang, "Autonomous power management with double-q reinforcement learning method," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1938–1946, 2019.
- [3] A. Emdadi and C. Eslahchi, "Auto-hmm-lmf: feature selection based method for prediction of drug response via autoencoder and hidden markov model," *BMC bioinformatics*, vol. 22, no. 1, pp. 1–22, 2021.
- [4] G. Adam, L. Rampášek, Z. Safikhani, P. Smirnov, B. Haibe-Kains, and A. Goldenberg, "Machine learning approaches to drug response prediction: challenges and recent progress," *NPJ precision oncology*, vol. 4, no. 1, p. 19, 2020.
- [5] Y. Zhang, J. W. Martens, J. X. Yu, J. Jiang, A. M. Sieuwerts, M. Smid, J. G. Klijn, Y. Wang, and J. A. Foekens, "Copy number alterations that predict metastatic capability of human breast cancer," *Cancer research*, vol. 69, no. 9, pp. 3795–3801, 2009.
- [6] J. Lovén, D. A. Orlando, A. A. Sigova, C. Y. Lin, P. B. Rahl, C. B. Burge, D. L. Levens, T. I. Lee, and R. A. Young, "Revisiting global gene expression analysis," *Cell*, vol. 151, no. 3, pp. 476–482, 2012.
- [7] W. Yang, J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J. A. Smith, I. R. Thompson *et al.*, "Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells," *Nucleic acids research*, vol. 41, no. D1, pp. D955–D961, 2012.

- [8] P. Smirnov, Z. Safikhani, N. El-Hachem, D. Wang, A. She, C. Olsen, M. Freeman, H. Selby, D. M. Gendoo, P. Grossmann *et al.*, "Pharmacogx: an r package for analysis of large pharmacogenomic datasets," *Bioinformatics*, vol. 32, no. 8, pp. 1244–1246, 2016.
- [9] J. Liu, T. Lichtenberg, K. A. Hoadley, L. M. Poisson, A. J. Lazar, A. D. Cherniack, A. J. Kovatich, C. C. Benz, D. A. Levine, A. V. Lee *et al.*, "An integrated tcga pan-cancer clinical data resource to drive high-quality survival outcome analytics," *Cell*, vol. 173, no. 2, pp. 400–416, 2018.
- [10] T. Sakellaropoulos, K. Vougas, S. Narang, F. Koinis, A. Kotsinas, A. Polyzos, T. J. Moss, S. Piha-Paul, H. Zhou, E. Kardala *et al.*, "A deep learning framework for predicting response to therapy in cancer," *Cell reports*, vol. 29, no. 11, pp. 3367–3373, 2019.
- [11] P. Geeleher, N. J. Cox, and R. S. Huang, "Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines," *Genome biology*, vol. 15, pp. 1–12, 2014.
- [12] M. Li, Y. Wang, R. Zheng, X. Shi, Y. Li, F.-X. Wu, and J. Wang, "Deepdsc: a deep learning method to predict drug sensitivity of cancer cell lines," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 18, no. 2, pp. 575–582, 2019.
- [13] A. Oskooei, J. Born, M. Manica, V. Subramanian, J. Sáez-Rodríguez, and M. R. Martínez, "Paccmann: prediction of anticancer compound sensitivity with multi-modal attention-based neural networks," *arXiv preprint arXiv:1811.06802*, 2018.
- [14] J. Choi, S. Park, and J. Ahn, "Refdnn: a reference drug based neural network for more accurate prediction of anticancer drug resistance," *Scientific reports*, vol. 10, no. 1, p. 1861, 2020.
- [15] 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, 2019.
- [16] Y.-C. Chiu, H.-I. H. Chen, T. Zhang, S. Zhang, A. Gorthi, L.-J. Wang, Y. Huang, and Y. Chen, "Predicting drug response of tumors from integrated genomic profiles by deep neural networks," *BMC medical genomics*, vol. 12, no. 1, pp. 143–155, 2019.
- [17] B. M. Kuenzi, J. Park, S. H. Fong, K. S. Sanchez, J. Lee, J. F. Kreisberg, J. Ma, and T. Ideker, "Predicting drug response and synergy using a deep learning model of human cancer cells," *Cancer cell*, vol. 38, no. 5, pp. 672–684, 2020.
- [18] D. Khan and S. Shedole, "Leveraging deep learning techniques and integrated omics data for tailored treatment of breast cancer," *Journal of Personalized Medicine*, vol. 12, no. 5, p. 674, 2022.
- [19] T. Chu, T. T. Nguyen, B. D. Hai, Q. H. Nguyen, and T. Nguyen, "Graph transformer for drug response prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2022.
- [20] J. Gao, J. Hu, W. Sun, Y. Shen, X. Zhang, X. Fang, F. Wang, and G. Zhao, "Tcr: A transformer based deep network for predicting cancer drugs response," *arXiv preprint arXiv:2207.04457*, 2022.
- [21] S. Park, J. Soh, and H. Lee, "Super. felt: supervised feature extraction learning using triplet loss for drug response prediction with multi-omics data," *BMC bioinformatics*, vol. 22, no. 1, pp. 1–23, 2021.
- [22] Z. Zuo, P. Wang, X. Chen, L. Tian, H. Ge, and D. Qian, "Swnet: a deep learning model for drug response prediction from cancer genomic signatures and compound chemical structures," *BMC bioinformatics*, vol. 22, no. 1, pp. 1–16, 2021.
- [23] 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.
- [24] P. Liu, H. Li, S. Li, and K.-S. Leung, "Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network," *BMC bioinformatics*, vol. 20, no. 1, pp. 1–14, 2019.
- [25] M. Oh, S. Park, S. Lee, D. Lee, S. Lim, D. Jeong, K. Jo, I. Jung, and S. Kim, "Drim: a web-based system for investigating drug response at the molecular level by condition-specific multi-omics data integration," *Frontiers in Genetics*, vol. 11, p. 564792, 2020.
- [26] 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, 2019.
- [27] M. Ammad-Ud-Din, S. A. Khan, K. Wennerberg, and T. Aittokallio, "Systematic identification of feature combinations for predicting drug response with bayesian multi-view multi-task linear regression," *Bioinformatics*, vol. 33, no. 14, pp. i359–i368, 2017.
- [28] Q. Zhang, Z. Chen, A. Lv, L. Zhao, F. Liu, and J. Zou, "A universal storage architecture for big data in cloud environment," in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*. IEEE, 2013, pp. 476–480.