International Journal of
Molecular Sciences

MDPI

Article

Cancer Drug Sensitivity Prediction Based on Deep Transfer Learning

Weijun Meng [1], Xinyu Xu [2], Zhichao Xiao [2], Lin Gao [2] and Liang Yu [2,*]

[1] School of Computer Science and Technology, Xi'an University of Posts & Telecommunications, Xi'an 710071, China; mengweijun@xupt.edu.cn
[2] School of Computer Science and Technology, Xidian University, Xi'an 710071, China; 22031212455@stu.xidian.edu.cn (X.X.); x1425906650@163.com (Z.X.); lgao@mail.xidian.edu.cn (L.G.)
* Correspondence: lyu@xidian.edu.cn

Abstract: In recent years, many approved drugs have been discovered using phenotypic screening, which elaborates the exact mechanisms of action or molecular targets of drugs. Drug susceptibility prediction is an important type of phenotypic screening. Large-scale pharmacogenomics studies have provided us with large amounts of drug sensitivity data. By analyzing these data using computational methods, we can effectively build models to predict drug susceptibility. However, due to the differences in data distribution among databases, researchers cannot directly utilize data from multiple sources. In this study, we propose a deep transfer learning model. We integrate the genomic characterization of cancer cell lines with chemical information on compounds, combined with the Encyclopedia of Cancer Cell Lines (CCLE) and the Genomics of Cancer Drug Sensitivity (GDSC) datasets, through a domain-adapted approach and predict the half-maximal inhibitory concentrations (IC50 values). Afterward, the validity of the prediction results of our model is verified. This study effectively addresses the challenge of cross-database distribution discrepancies in drug sensitivity prediction by integrating multi-source heterogeneous data and constructing a deep transfer learning model. This model serves as a reliable computational tool for precision drug development. Its widespread application can facilitate the optimization of therapeutic strategies in personalized medicine while also providing technical support for high-throughput drug screening and the discovery of new drug targets.

Keywords: deep transfer learning; domain-adapted approach; drug sensitivity; multi-source data

Academic Editor: Stergios Boussios

Received: 11 January 2025
Revised: 27 February 2025
Accepted: 6 March 2025
Published: 10 March 2025

Citation: Meng, W.; Xu, X.; Xiao, Z.; Gao, L.; Yu, L. Cancer Drug Sensitivity Prediction Based on Deep Transfer Learning. *Int. J. Mol. Sci.* **2025**, *26*, 2468. https://doi.org/10.3390/ijms26062468

1. Introduction

Pharmaceutical research, as we know it today, began when chemistry reached a threshold level of maturity, guided by pharmacology and clinical sciences, and has contributed to the advancement of medicine more than any other scientific factor [1]. The emergence of molecular biology, especially the development of genomic science, has had a profound impact on drug discovery [2–4]. Genome science, combined with bioinformatics tools, enables us to identify the genetic bases of multifactorial diseases to select the most appropriate treatments [5,6]. Molecular biology enables the understanding of disease processes at the genetic level and the identification of optimal molecular targets for drug intervention [7–9]. Target modification is the highest level of validation for drug discovery, such as the blockade of receptors or the inhibition of enzymes under the action of a drug, resulting in reversal of the disease state. Whereas phenotypic changes in isolated cells caused by compounds that modify a target constitute a minimal validation, if the phenotypic changes

can be reproducibly induced in animal models of certain disease-related mechanisms and the confidence of the target varies with the animal model as the number increases, the modification of that target results in the desired phenotypic change [10–12].

Drug discovery is mainly based on molecular target discovery and phenotypic screening, the former being the main method of drug discovery in recent decades. However, in recent years, most of the large number of approved drugs originated from phenotypic screening, and the exact mechanism of action or molecular target of each was elaborated. High-throughput screening techniques have provided diverse drug susceptibility data for cancer cell lines and hundreds of compounds, and large-scale pharmacogenomics studies of cancer genomes have provided unprecedented insights into the study of anticancer therapies to determine putative predictions of drug susceptibility and to conduct phenotypic screening to develop new anticancer drugs for anticancer therapy [13]. The Cancer Cell Line Encyclopedia (CCLE) [14] and the Genomics of Drug Sensitivity in Cancer (GDSC) [15] datasets are the most popular datasets in this field.

Currently, many research groups have used computational methods for drug sensitivity prediction, most of which are based on machine learning. The simultaneous modeling of compound and cell line characteristics facilitates drug sensitivity prediction, drug side effect analysis, and model extrapolation to new compounds and cell lines [16]. Menden et al. integrated cell line genomic features, including microsatellite sequences, sequence and copy number variations, and one-dimensional (1D) and two-dimensional (2D) features of compounds. A model for predicting the half-maximal inhibitory concentration (IC50) was established by using a neural network and Random Forest (RF) [17]. Another study built a dual-layer integrated cell line-drug network containing a cell line similarity network and a drug similarity network based on the Pearson correlation coefficients of the gene expression profiles of the cell line and information on compounds for predicting IC50 values [18]. Studies have also used multitask learning to integrate various omics data to predict drug responses [19,20]. However, drugs and cell lines are often represented by predefined features, such as structural features and omics features, respectively. Traditional machine learning-based methods often face the "small n, large p" problem. This is because the number of cell lines is much smaller than the number of genes in the gene map, which limits the prediction performance of traditional machine learning-based methods. As a pivotal branch of machine learning, deep learning leverages the construction of multi-layer nonlinear neural network architectures to autonomously learn hierarchical abstract features from data, thereby enabling the identification and prediction of complex patterns through these high-level representations [21–25]. In studying drug responses, various studies have used cell line genomic features and the structural features of drugs to predict drug susceptibility. DeepDR [26] is a model for predicting IC50 values with deep neural networks. In DeepDSC [27], a stacked autoencoder is used to extract genomic features of cell lines, and these features are combined with drug fingerprints to predict IC50 values. Drug fingerprints refer to distinctive chemical profiles or spectral patterns utilized for the identification and differentiation of pharmaceutical compounds, which are extensively applied in drug analysis. For instance, the infrared spectroscopy or mass spectrometry profiles of specific drugs can serve as unique fingerprints to facilitate the precise identification of their chemical constituents. In GraphDRP [28], the molecular structure of a drug is converted into a graph structure and then a graph convolutional neural network is used to obtain a representation of the drug, which is combined with a one-dimensional convolutional representation of the copy number variation feature of the cell line to predict the drug response.

Transfer learning is a machine learning approach that enhances learning efficiency and performance on related but distinct tasks by leveraging knowledge gained from one

task. Transfer learning methods can be applied to datasets that are in the same feature space but have different distributions. However, in transfer learning, we need to assume that there is some consistency between the datasets. According to HaibeKains et al. [29], the GDSC and CCLE databases have a good correlation in terms of gene expression [30]. This implies that the gene expression data in these two databases have a certain level of consistency and similarity within the same feature space. Such consistency provides the foundation for transfer learning, as it suggests that the patterns learned from one database regarding the relationship between gene expression and drug response may be applicable to another database as well. The database researchers noted that the drug response data mostly corresponded to drug-insensitive lines with far fewer outliers and speculated that biological differences between the cell lines contributed to the poorer correlations. By applying the knowledge learned about the relationship between gene expression and drug response from one database to another, the overlapping information can be leveraged to enhance the accuracy of drug response models. There is significant overlap between the GDSC and CCLE databases, and both databases provide information on the same biological processes, making them suitable candidates for transfer learning methods that can be used to further design more accurate drug response models. Saugato et al. [31] proposed two classes of transfer learning solutions for drug response prediction. One is a transfer learning method based on cost optimization, which mainly establishes a relationship between the two latent variables, namely, the GDSC and CCLE database gene expression values, and the AUC (area under the dose—response curve) values. The other is the domain adaptation method, which looks for a mapping relationship between the gene expression values and AUC values. This method considers the mapping of the gene expression values between the two databases to be linear and the mapping of the AUC-predicted values to be nonlinear.

However, the number of cell lines in this task is much smaller than the number of genetic maps, which limits the predictive performance of traditional machine learning-based methods. Deep learning has been widely used to extract features from complex, high-dimensional features and complete target prediction tasks, and its performance is even better. Existing drug sensitivity methods based on deep learning often incorporate numerous features of cancer cell lines, such as sequence variations, copy number variations, etc. However, these features are not all easily available for different cancer cell lines. All physiological manifestations of organisms are derived from the most basic gene expression. Therefore, one of the motivations for this work is to mine the key features of cancer cell line gene expression values for prediction tasks based on deep feedforward networks. Inspired by deep learning and transfer learning, this paper proposes a deep transfer learning framework that integrates the gene expression profiles of two database cell lines and the molecular structures of compounds. This method is called DADSP (Domain Adaptation for Drug Sensitivity Prediction). Our model extracts the features of gene expression maps from stacked self-encodings, obtains low-dimensional representations of these features, and combines them with molecular features of compounds to obtain prediction results through a deep feedforward network. Combined with our validity verification, it is confirmed that our model conforms to the biological mechanism of drug response. Hence, it is not only a supplement to deep learning in drug response prediction but also provides a means for transferring data distribution processing among different databases of the same biological process.

The codes and datasets are available at https://github.com/xzc196/DADSP, (accessed on 5 March 2025).

2. Results

2.1. Performance Comparison with Other Algorithms

To comprehensively evaluate the performance of the DADSP method, this section compares the DADSP method with commonly used domestic and international methods. We refer to the model structure of [10], remove the feature extractor and regression predictor of the model separately, and keep the rest of the parts consistent to form a standard deep feedforward network drug susceptibility prediction model. The comparison is primarily based on ablation experiments to verify the effectiveness of the transfer learning module in the model. In these experiments, the drug features are all represented using the Morgan fingerprint. The following models are introduced in the ablation experiments: DADSPA-: a model without pre-training using a stacked autoencoder. DeepDSC-1 [27]: a deep feedforward network model that does not use a domain adversarial discriminator and is trained and tested only on target domain data. DeepDSC-2 [27]: a model with the same network structure as DeepDSC-1, but the parameters are transferred using a stacked autoencoder trained on both source and target domain data. SLA (Selective Learning Algorithm) [32]: a model that defines an intermediate domain to capture source domain information and transfer it to the target domain. In this model, the data from the GDSC and CCLE are jointly used to pre-train a stacked autoencoder, and the parameters are then transferred to a feedforward neural network. Additionally, different choices of regression predictors were also compared, including Random Forest (RF) [33], Logistic Regression (LR) [34], and Support Vector Regression (SVR) [35], as shown in Table 1.

Table 1. Model comparison results.

Method	*RMSE*	R^2
DADSP-A	0.64	0.43
DADSPA-	0.71	0.31
DADSP-B	0.69	0.35
DeepDSC-1	0.82	0.11
DeepDSC-2	0.72	0.29
SLA	0.82	0.10
RF	0.75	0.27
LR	0.75	0.26
SVR	0.73	0.29

From Table 1, we can see that the DeepDSC-1 model, which only uses target domain data, performs poorly, indicating the necessity of transfer learning in this work. The *RMSE* difference between the DADSPA- model without pre-training and the DADSP-A model is 0.6, demonstrating the important contribution of the stacked autoencoder to the genomic representation. The traditional machine learning methods LR, RF, and SVR do not perform as well as the feedforward neural network, reflecting the training advantage of neural networks on large-scale data. In terms of transfer learning strategies, the adversarial-based DADSP-A method outperforms the domain difference and intermediate domain-based DADSP-B and SLA methods. This paper hypothesizes that for the feature representation of drug sensitivity data, the adversarial-based method, which uses a feedforward neural network binary classifier as the main implementation of transfer learning, is more effective in extracting nonlinear representations between domains, and therefore performs better.

2.2. Blind Test

In a normal test experiment, a drug/cell line appears in both the training and test sets, although its drug response data are not involved in training. However, predicting responses to unseen drugs/cell lines is more challenging. Therefore, in blind testing experiments, the

drugs/cell lines in the testing phase do not exist in the training phase, and we randomly delete a portion of the drugs/cell lines. Instead of removing drugs or cell lines separately for two blinded experiments, we opted to remove both drugs and cell lines simultaneously, making the model predictions more challenging. Our blind test results for different models are presented in Table 2.

Table 2. Blind test results.

Method	*RMSE*	R^2
DADSP-A	0.69	0.32
DADSP-B	0.92	0.01
DeepDSC-1	0.70	0.29
SLA	0.72	0.30

As shown in Table 2, the model performance declined across the board compared to the standard test results, which also confirms that the adversarial DADSP-A still achieved the best performance in the blind test. However, DADSP-B exhibited unexpectedly poor performance in the blind test, which is suspected to be related to the randomly deleted cancer cell lines or drug data, potentially requiring the setting of new hyperparameters to adapt to the training. The results of DeepDSC-1 and SLA did not differ significantly, and both transfer learning strategies achieved certain effects. The blind test experiment examined the model's ability to explore the common hidden feature representations between different drug and cancer cell line characteristics. The adversarial-based method performed the best, which also indirectly confirms its unique advantage in domain adaptation feature distance.

2.3. Comparison of the Characteristics of Different Drugs

In addition to cell line gene expression features, the extraction methods that are used for drug feature representation have a huge impact on the drug sensitivity prediction task. We compare different extraction methods of drug molecular features. We use DADSP-A for experiments, only changing the drug feature extractor and leaving the rest of the model unchanged. We screen four drug feature extraction methods, all from the current state-of-the-art deep drug sensitivity prediction models. They are known as the hashed Morgan fingerprints of drug molecules [36], struct similarity profiles (SSPs) based on user-defined reference drugs [37], two-dimensional matrix information of drug molecules extracted and constructed by convolutional neural networks (CNNs) [38], and drug molecular graph structures that are extracted using graph convolutional neural networks (GNNs) to obtain individual drug representations [28,39]. We compare the results, as presented in Table 3.

Table 3. Comparison results of drug feature extraction methods.

Method	*RMSE*	R^2
DADSP-A	0.64	0.43
DADSP-A + SSP	0.67	0.35
DADSP-A + CNN	0.76	0.29
DADSP-A + GCN	0.74	0.23

According to the results, the hashed Morgan fingerprints of drug molecules perform the best. A Morgan fingerprint hashes the molecular features of each layer and each atom into a bit vector. Each layer considers the atomic fingerprints that are less than the specified maximum distance from the starting atom and then diffuses out layer by layer. This approach can better express the molecular structure characteristics of the drug. Although convolutional neural networks and graph convolutional neural networks have powerful

effects in feature extraction, we must first process the drug molecular structure into a corresponding two-dimensional matrix structure and graph structure defined by the user. The performance on this task may be limited by this.

2.4. Uknown Drug Response Prediction

In this part of the experiment, we use the best performing model in the normal experiment to predict the IC50 values of a missing pair from a dataset of 39,000 missing drug response pairs that we screened. Figure 1 shows the predicted top and bottom 10 drugs. The results show that, in this part of the experiment, we predict the lowest IC50 for Epthilone B, which can organize cell division by interacting with microscopic proteins [40], and the remaining nine drugs all play positive roles in the treatment or inhibition of cancer. Meanwhile, AICA R and phenformin are predicted to have the highest IC50 values, implying that the cancer is not sensitive to these two drugs. AICA R is used clinically to treat and prevent ischemic heart damage [41], while phenformin is used by people to treat diabetes [38]. Our prediction results regarding the two drugs with the highest IC50 values are the same as those in [28], and the 10 drugs with the lowest IC50 values also have multiple overlaps.

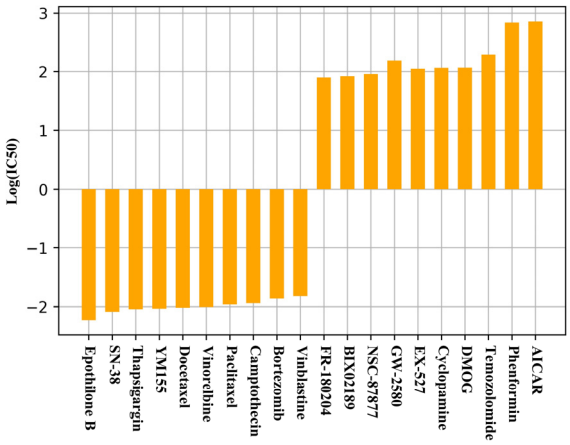

Figure 1. The 10 drugs with the lowest and highest log(IC50) values in the unknown drug response experiment.

This evidence suggests that cancers are less sensitive to drugs with high IC50 values and that our model is effective in predicting missing pairs of potential drug responses.

2.5. Predicting Critical Genes for Drug Responsiveness

Deep learning models can extract expression information from input data but suffer from poor interpretability. While our model makes good predictions, an understanding of the genetic signatures that are active in the model is also necessary. Therefore, this experiment selects the most sensitive drug, Epothilone B (EpoB), and the three cell lines with the lowest IC50 values for this drug to further study the contributions of genetic characteristics to the prediction of the cell line response. At the same time, we introduce the integrated gradient method [42], which is a state-of-the-art feature interpretation method for deep neural networks. This method can feed back the contributions of active genetic features to the model to the input layer, and then key genes can be obtained. Specifically, key genes are identified based on the accumulation of neuron gradients along the path of the fully connected layer in the neural network. Table 4 presents the ten genes with the highest scores in the three cell lines. Metallothionein 1M in the MNK7 and ZR-75-30 cell lines is predicted to be the most critical gene. The high expression of this gene in gastric cancer tissue is a promoting factor of gastric cancer invasion and metastasis

and is related to the occurrence of gastric cancer [43,44]. At the same time, it plays an important role in the hormone regulation of breast cancer and the occurrence of breast cancer. We also perform KEGG gene enrichment analysis on the top few hundred key genes of the ZR-75-30 cell line using the R package ClusterProfiler, version 4.14.6, https://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html, (accessed on 5 March 2025) [45], and the results are shown in Figure 2.

Table 4. Critical genes in MKN7, ZR-75-30, and MEL-HO identified by the integrated gradient method.

MKN7		ZR-75-30		MEL-HO	
Critical Gene	**Score**	**Critical Gene**	**Score**	**Critical Gene**	**Score**
ENSG00000205364	0.002803	ENSG00000111700	0.003027	ENSG00000205364	0.003027
ENSG00000187908	0.002338	ENSG00000183032	0.002608	ENSG00000164821	0.002608
ENSG00000164821	0.002276	ENSG00000158023	0.002451	ENSG00000158023	0.002451
ENSG00000174469	0.002231	ENSG00000103316	0.002122	ENSG00000187908	0.002122
ENSG00000158023	0.002206	ENSG00000183668	0.002003	ENSG00000183032	0.002003
ENSG00000111404	0.002066	ENSG00000111249	0.001953	ENSG00000110077	0.001953
ENSG00000183032	0.002022	ENSG00000187908	0.001919	ENSG00000111404	0.001919
ENSG00000183668	0.001981	ENSG00000165168	0.001918	ENSG00000183668	0.001918
ENSG00000167083	0.001887	ENSG00000164821	0.001834	ENSG00000166049	0.001834
ENSG00000120162	0.001884	ENSG00000166049	0.001831	ENSG00000111249	0.001831

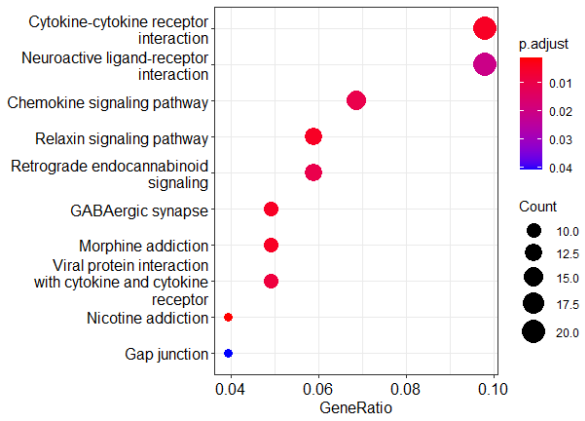

Figure 2. Pathways enriched for key genes in ZR-75-30 cells. Multiple pathways are involved in the mechanism of action or pathogenesis of cancer.

The three pathways with the largest numbers of critical genes, namely, the cytokine–cytokine receptor interaction, neuroactive ligand–receptor interaction [46], and chemokine signaling pathways, are all related to the pathogenesis of breast cancer [47–49]. Epothilone has been reported to be a nontaxane microtubule stabilizer that has a similar mode of action to paclitaxel but is active in paclitaxel-resistant cells. Metallothionein is highly expressed in the response of EpoB to cancer and may play a role in chemoresistance [50].

According to the results, our deep learning model can accurately extract the active genetic features in the drug response, which is in line with the antitumor mechanism of the drug.

2.6. Feature Space Comparison After Domain Adaptation

Next, we conduct a comparison experiment of the feature space before and after model training. We use the t-SNE [51] algorithm to map high-dimensional features into a two-dimensional space. Figure 3 shows the feature space of the initial source and target domains. Figure 4 shows the feature space after we trained the model. As shown in the figure, the data feature distributions of the source domain and the target domain after domain adaptation are significantly closer. A similar data distribution ensures the prediction effect of the test set.

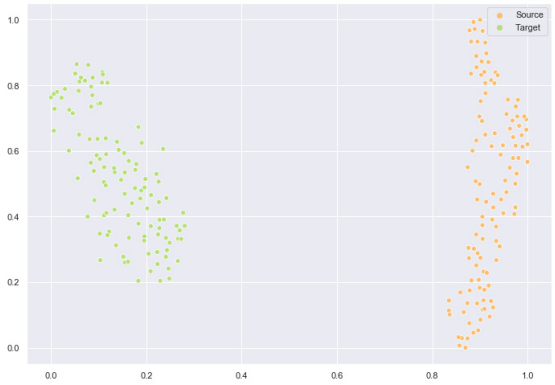

Figure 3. Feature space before domain adaptation.

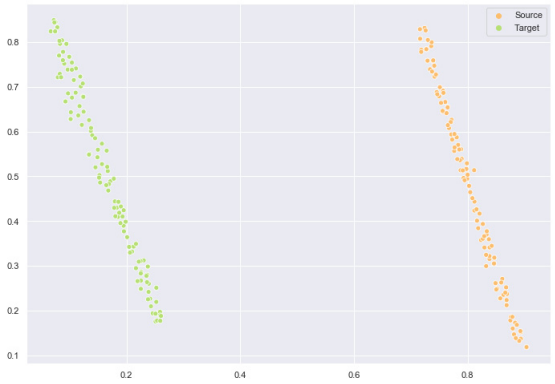

Figure 4. Feature space after domain adaptation.

Overall, our domain adaptation algorithm achieves the expected results and plays an important role in the post-transfer tests.

3. Discussion

This article proposes a domain-adaptive drug sensitivity prediction method called DADSP and provides a detailed description. The DADSP method takes cancer cell line gene expression data, drug feature data, and drug sensitivity response data as the input and outputs the predicted drug sensitivity value IC50. Furthermore, this article provides a detailed explanation of the sources and specific forms of the input data. We elaborate on the specific model framework of the DADSP model, considering domain adaptation for different types of learning methods. We propose adversarial-based and domain discrepancy-based deep learning frameworks for drug sensitivity prediction tasks. In the model framework, we adapt the network parameters trained through unsupervised training of stacked autoencoders, which effectively reduce dimensionality and represent features well for high-dimensional features. We also provide a detailed comparison of DADSP with other methods in terms of model performance and transfer learning strategies. The adversarial transfer learning approach leverages the idea of using GANs to learn the

implicit relationship function between the source domain and the target domain [52]. This implicit relationship is learned through complex nonlinear transformations in the network, making it more adaptable to different tasks. Therefore, the adversarial-based DADSP model in this article achieves the best performance on the target domain test set. Additionally, various methods are analyzed and discussed.

The article conducts effective analysis of the model's prediction results. Firstly, a comparative experiment is conducted on different drug feature extraction methods, and it is found that the molecular fingerprint-based extraction method is more suitable for feedforward neural networks. The active genetic features in drug sensitivity prediction tasks are analyzed using the integral gradient method [42]. Through a series of research investigations, it is discovered that the active genetic features are all related to the mechanisms of cancer. The changes in domain feature distances before and after model training are visualized and analyzed, providing readers with a more intuitive understanding of the effects of domain adaptation. Through a series of analysis experiments on the model, the DADSP model not only completes the drug sensitivity prediction task but also confirms the successful application of domain adaptation methods in the problem of inconsistent data distributions.

4. Materials and Methods

4.1. Data Sources

In this study, we use cell line expression data and drug response data from GDSC and CCLE, as well as compound structure files from the PubChem [53] database. The GDSC and CCLE contain omics data and drug response data for thousands of cell lines. The gene expression (transcriptome data) of a cell line represents the level of activity of a gene when a cell from that line is in a certain state. The omics data also indicate whether a genome has produced mutations and copy number variations. The drug response is an important indicator for measuring whether cells are inhibited under the action of a drug. Specifically, it is the IC50 value or AUC value of the drug required to inhibit half of the biological activity of the cell line.

The GDSC is the largest cell line drug sensitivity database. Through preprocessing, we obtained drug response data for more than 213 drugs on 914 cell lines and obtained the expressions of 24 drugs on 504 cell lines from the CCLE. We downloaded array gene expression data for thousands of cell lines from both databases separately and processed all of these data via robust multi-array average (RMA) normalization. To better evaluate our model, we retrieved 16,017 shared genes from both databases. Therefore, we obtained standardized two-dimensional gene expression matrices from the two databases, and the values in the matrix primarily fell in the range of 3 to 10. To evaluate the drug response data, we selected IC50 as a measurement index, with a lower IC50 value indicating that the cell line was more sensitive to the corresponding drug. We treated IC50 values as -log10IC50 values. Data for more than 200 drugs in canonical SMILES format [54] were downloaded from the PubChem database through the RDKit package, version 4.14.6, https://www.rdkit.org/docs/index.html, (accessed on 5 March 2025) [55] and processed into the desired expression formats, including 256-bit hashed Morgan fingerprints. A drug graph representation structure was designed using the atomic features of Deepchem [39] and the molecular graph construction mode of GraphDRP [28]. In this article, CCLE data is regarded as the target domain dataset, that is, the test set. The size of the source domain dataset is about 10 times the size of the target domain dataset.

4.2. Model Input Data

This section will introduce the data format for input to the DADSP model. The input data mainly consists of three parts:

(1) The gene expression profiles of cancer cell lines, represented as d, where 16,016 is the number of shared genes and N is the number of training samples in a batch.
(2) Drug features, represented as $x_2 \in \mathbb{R}^{N\times 256}$, where 256 is the length of the hashed Morgan fingerprint. Subsequent chapters will conduct experimental analysis on different ways of representing drug features.
(3) Cancer cell line–drug sensitivity data, represented in the format [Cell Line ID, Drug ID, IC50 value].

4.3. Deep Transfer Learning and Autoencoder

Deep learning has been used to process scRNA-seq data to extract an increasing number of abstract features of the original input through a series of nonlinearly transformed hidden layers in deep architecture learning [21]. In recent years, deep learning has achieved widespread application development in bioinformatics [56–64], computer vision [65,66], and natural language processing [67,68]. Deep transfer learning (DTL) is a technology that uses deep learning to build transfer learning models which has made great progress in the imaging field and includes methods such as mapping-based deep transfer methods [69] and adversarial-based transfer methods [70,71]. As illustrated in Figure 5, the data of the source domain and the target domain are mapped to a new data space to obtain new feature representations and then make predictions based on different task regressions or classifications. Domain adaptation [72,73] is an important method in deep transfer learning. It is often used in situations where the source domain data distribution is inconsistent with the target domain data distribution. Domain adaptation enables learners to generalize across different domains with different distributions by matching marginal and conditional distributions.

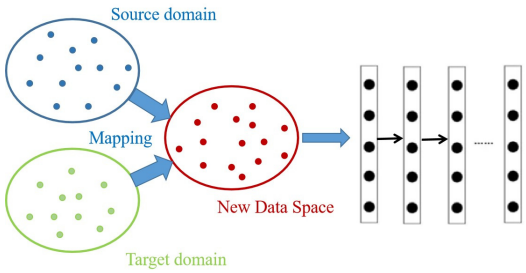

Figure 5. General framework for deep transfer learning. The source domain and the target domain are mapped into a common data space.

In this research, we use the GDSC database as the source domain and the CCLE database as the target domain, and we realize the transfer of knowledge from the GDSC to the CCLE. Before applying the domain adaptation strategy, we first extract high-level low-dimensional feature representations of the data through various methods.

A stacked autoencoder (SAE) is a layer-by-layer unsupervised deep learning model that attempts to reconstruct the input under constraints such as low dimensionality and noise-free nature [74,75]. An SAE is a stack of multiple autoencoders, as illustrated in Figure 6, and the encoder and decoder are represented by Formulas (1) and (2).

$$h = f_\theta(wx + b) \tag{1}$$

$$x\prime = g_{\theta\prime}(h) = g_{\theta\prime}(f(wx + b)) \tag{2}$$

where x is the input layer, h is the middle bottleneck layer, and x' is the output layer after reconstruction.

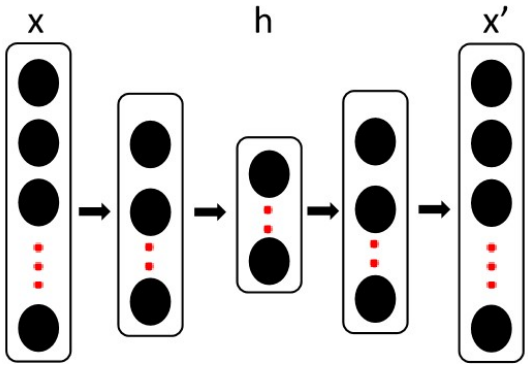

Figure 6. Architecture of a stacked autoencoder. x and x' denote the input and the reconstructed output, respectively, and h denotes the encoded feature representation.

During layer-by-layer training, a single self-encoding layer passes through a three-layer network x->h->x and the reconstruction loss between the original vector and the reconstructed vector is used for back propagation at the end of the training of this layer, the output layer is treated as a new input layer, and an autoencoder is trained for the next layer. Finally, the entire SAE network is fine-tuned. Low-dimensional representations are obtained by stacking the gene expression features of pretrained cell lines from autoencoders.

In this study, the test set is an independent target domain dataset, so in the unsupervised training, only the cancer genomic data from the source domain (GDSC database) is used. Finally, the encoder network parameters of the stacked autoencoder are used as the initial parameters of the DADSP gene feature extractor, which also applies the parameter fine-tuning transfer learning idea, saving the training time of the entire DADSP model.

4.4. Our Method

4.4.1. Adversarial-Based Domain Adaptation Models

Deep feedforward networks are widely used in various fields in both regression tasks and classification tasks [76–81]. In the task of drug response prediction, deep feedforward networks exhibit strong modeling capabilities [26]. In deep learning drug susceptibility research, excellent performance has been realized on regression tasks and classification tasks [27,82–85]. According to [86], discrete drug response data category prediction involves information loss compared with regression tasks, so we propose a deep transfer learning model for predicting continuous drug sensitivity values (IC50 values). A flowchart of our DADSP A is shown in Figure 7. Our model can be divided into four parts, namely, a gene expression feature extractor, drug feature extractor, domain discriminator, and regression predictor. For the gene expression feature extractor, we first use a stacked autoencoder as the gene expression feature extractor to extract features from gene transcriptomes. This feature extraction module is composed of a feedforward neural network with the same structure as the encoder part of the stacked autoencoder mentioned earlier, and the parameters are initialized with the transferred network weights. During the training process, the source domain data and the target domain data are concatenated along the batch dimension and fed into this module jointly. For the drug feature extractor, we choose the hashed Morgan fingerprints of the compounds, which are represented as 256-dimensional vectors. We use a two-layer feedforward neural network to construct this feature extraction module. Next is the domain discriminator module, which is composed of a three-layer feedforward neural network. The input to this module is the feature vector from the last layer of the gene feature extractor. The source and target domain data concatenated along the batch dimension are

then passed through the domain discriminator, which outputs a two-dimensional domain probability distribution. This is jointly optimized with the constructed domain labels of [0, 1] using the cross-entropy loss function. The expression of the cross-entropy loss function is

$$L = -\left[y \log y' + (1 - y) \log(1 - y')\right] \tag{3}$$

where y' is the output of the domain discriminator and y is the true domain label. Based on the adversarial learning principle, we need to maximize this loss function to make the binary classifier unable to distinguish the source and target domain data, thereby achieving the goal of similar data distributions. While ensuring the prediction task, we also need to perform the domain adversarial objective task. Therefore, there is a gradient reversal layer (GRL) [70] in front of the domain discriminator module. The role of the GRL is to automatically reverse the gradients during backpropagation, while keeping them unchanged during forward propagation. The formula for the GRL is as follows:

$$R_\lambda(x) = x \tag{4}$$

$$\frac{dR_\lambda}{dx} = -\lambda I \tag{5}$$

$$\lambda = \frac{2}{1 + \exp(-\gamma \cdot p)} - 1 \tag{6}$$

where γ is a constant, set to 10, and p is the ratio of the current iteration number to the total iteration number.

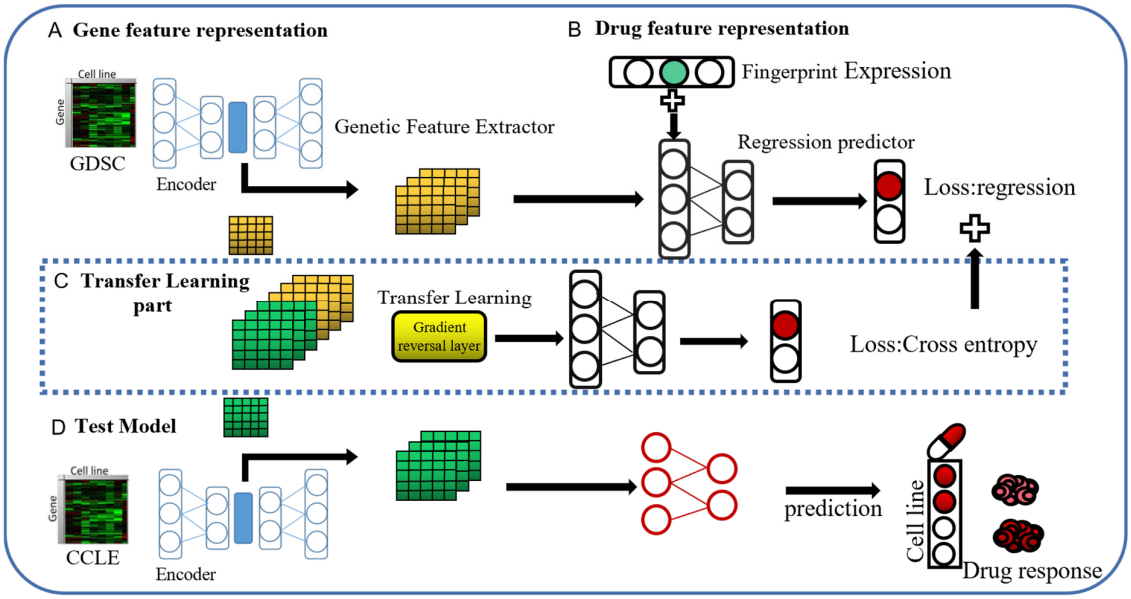

Figure 7. Flowchart of our DADSP-A, which consists of two feature extractors, a regression predictor, and a domain discriminator. (**A**): Gene feature representation, which uses an SAE to map high-dimensional gene expression to low-dimensional representations. (**B**): Drug feature representation. (**C**): Transfer learning part, which achieves the goal of feature proximity by maximizing the domain classification error. (**D**): Test module.

Finally, we need to complete the regression prediction. The module consists of a four-layer feedforward neural network. During the training process, the hidden layer features from the gene feature extractor are split along the batch dimension to extract the source domain features, which are then concatenated with the drug features and fed into the network. The reason for not concatenating the drug features and gene expression

features at the very beginning and inputting them into the deep feedforward network is that the scale of the drug features differs greatly from that of the gene features, and their feature contributions would be diminished.

The last layer of this module uses the Sigmoid activation function to scale the regression prediction values between [0, 1]. The loss optimization function of this module is the Mean Squared Error (MSE). The formula for MSE is

$$L_{MSE} = \sum_{i=0}^{m} (y_i - \hat{y}_i)^2 \quad (7)$$

Among them, y is the true IC50 value of the i-th data in the training set and y is the predicted value corresponding to the sample. The calculation formula of the total loss function of the entire model is

$$L_{Total} = L_{MSE} + L \quad (8)$$

Generally, our deep transfer model consists of two parts. The first part is feature extraction, which reduces the dimensionality of gene expression data using stacking autoencoders and extracts the genetic features. The second part is domain adaptation, which approximates the feature distribution by maximizing the error between the source domain and the target domain. Finally, a regression task is used to predict continuous drug sensitivity values.

Next is model training and hyperparameter settings. Before model training, the input data of the two datasets are min-max normalization processing. The activation function of all layers of the feedforward neural network is RELU and its calculation formula is $R(x) = \max(0, x)$. The experiment uses version 1.15 of the deep learning tool Tensorflow. The training optimization algorithm is Adam and the learning rate is 0.0001. The training batch size batch size is set to 128 and the total number of training rounds epoch is set to 15. Early stopping is set to 3; that is, if the model's loss does not decrease for 3 consecutive epochs, the training will stop.

4.4.2. Deep Transfer Models Based on Autoencoders and Difference Metrics

We also propose a novel structural deep transfer learning model based on autoencoders mapped with the mean maximum discrepancy (MMD), which is also applied to the drug sensitivity prediction task. The MMD loss [87] is widely used in deep transfer learning [69,88]. DAN [69] is applied to minimize the difference between the fully connected layers of source and target domain features after the feature extractor. Our DADSP-B is illustrated in Figure 8. The feature extractor is consistent with that in DADSP-A. The stacked autoencoder is selected and the entire network is constructed by many feedforward networks.

$$MMD\left(X^S, X^T\right) = \| \frac{1}{n^S} \sum_{i=1}^{n^S} \Phi\left(X_i^S\right) - \frac{1}{n^T} \sum_{j=1}^{n^T} \Phi\left(X_j^T\right) \|^2 \quad (9)$$

where n is the dimension of the data sample distribution; X^S and X^T are the sample distributions from the source domain and the target domain, respectively; and $\Phi(\cdot)$ maps the distribution of the original space to a regenerated Hilbert space and then measures the distance between the two sample distributions in the high-dimensional space.

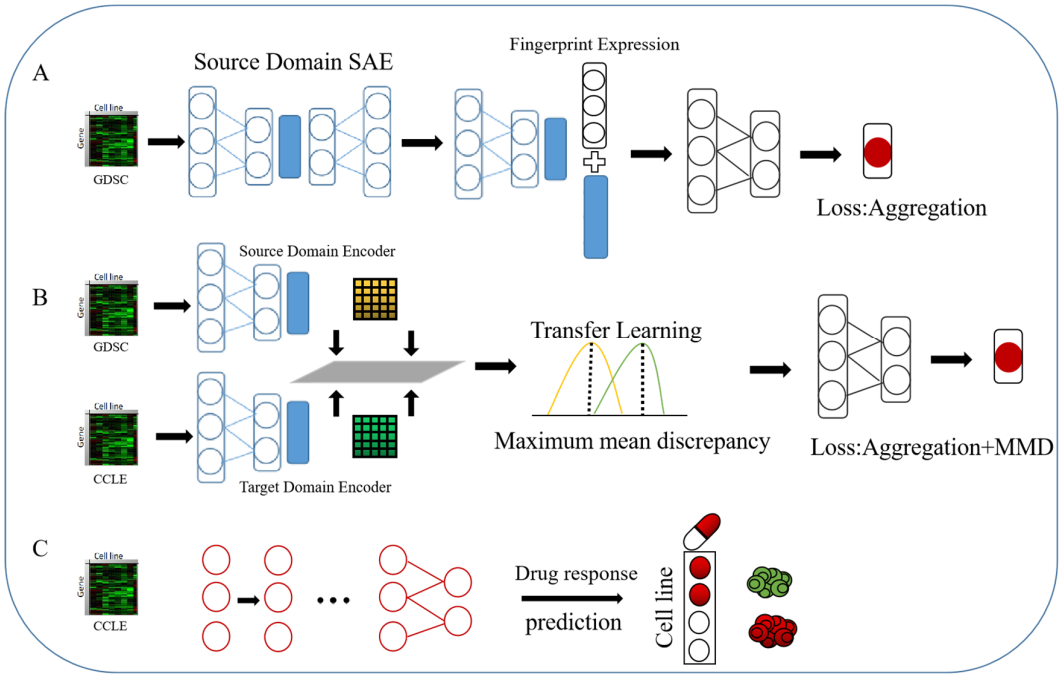

Figure 8. Flowchart of DADSP-B, which consists of two stacked autoencoders and a regression predictor. Part (**A**) uses only the source domain dataset GDSC through the feature extractor and regressor. Part (**B**) is the module of transfer learning, which minimizes the MMD loss of the gene features from the source domain and the target domain through the hidden layer features after the feature extractor to achieve a close feature distance between the two domains. The loss of the overall training is the regression loss of the MSE and MMD losses of the common training. Part (**C**) is the test part, which uses only the target domain dataset CCLE to predict drug sensitivity on the multitrained model.

During training, we adopt a freezing strategy [89]. The process is mainly divided into the following steps:

1. The feature extractor and regressor are trained using the source domain data to achieve the best possible performance for the source domain data on the regression task;
2. The target domain data are input into another stacked autoencoder [90]. We share the encoder parameters of the source domain data under the condition of freezing the regressor parameters, the reconstruction loss of the training target domain data layer by layer and the source domain data of the MMD loss.
3. Overall fine-tuning, freezing of the feedforward parameters of all feature extraction layers, and training of the regressor are performed. The loss at this time is only the MSE loss.

Our proposed DADSP-B differs from the adversarial-based DADSP-A in its use of the MMD loss function to measure the difference between the source and target domains. The MMD loss is incorporated into the reconstruction loss training of the stacked autoencoder. The purpose is to make the target domain data approximate the abstract low-dimensional features of the source domain data while reducing the dimensionality to obtain a low-dimensional representation. Through the freezing method, the parameters of the model are focused on the training of each part of the task rather than on the joint training of multiple loss functions, as in multitask learning [91]. We also compare and discuss the two models. The selection of hyperparameters is carried out in two steps. Firstly, we perform pre-training on the model using the original domain dataset, and then we apply grid search to select the hyperparameters after transferring to the target domain. Based on the performance observed in the target domain, we adjust the hyperparameters used in the

pre-training on the original domain. For specific parameter configurations, please refer to the accompanying code.

In general, we propose two deep transfer learning models. One is a domain adaptation method based on domain adversarial learning, which achieves the goal of reducing the distance between the features of the source and target domains by maximizing the cross-entropy loss of the features of the two domains. The other is a transfer learning method based on a difference metric, which directly makes the feature representations of the hidden layers of the two domains close together through the MMD loss. Both methods pass the data through a deep feedforward network after a feature extractor to predict the drug sensitivity score.

4.5. Performance Metrics

We use two metrics, namely, the root mean square error (*RMSE*) and coefficient of determination (R^2), to measure the performance of the model. *RMSE* is a commonly used indicator to measure regression prediction models and is expressed in (4). It is more sensitive to abnormally large errors and is more suitable for drug sensitivity prediction tasks than the mean square error (MSE). The coefficient of determination R^2, also known as the determination coefficient or goodness of fit, is a statistical measure used to evaluate the degree of fit of a regression model. It represents the proportion of the variance in the dependent variable that can be explained by the model. The values of R^2 range from 0 to 1, where a value closer to 1 indicates a better fit of the model to the observed data. The calculation formula, as shown in (5), is often used in regression analysis to judge the degree of fit of a regression equation and is used as a standard to measure model quality.

$$RMSE = \sqrt{\sum (y_{\mathrm{i}} - \widetilde{y}_i)^2 / N} \tag{10}$$

$$R^2 = 1 - \sum (y_i - \widetilde{y}_i^{r0})^2 / \sum (y_i - \widetilde{y}_i)^2 \tag{11}$$

where N is the size of the data; y_i and $\widetilde{y}_i$ are the label data for drug sensitivity prediction and the predicted value corresponding to the i-th input data, respectively; and $\overline{y}$ is the mean value of the target drug $\widetilde{y}_i^{r0} = k\overline{y}_i$, where k is the slope, as defined in (6).

$$k = \sum y_i \widetilde{y}_i / \sum \overline{y}_i^2 \tag{12}$$

5. Conclusions

In this study, we proposed a deep transfer learning model for drug sensitivity prediction which not only predicts drug responses in cancer cell lines but also uses transfer learning to provide solutions for data distribution differences between genomics databases. Our model was trained on the GDSC and CCLE datasets and was shown to achieve the goal of knowledge transfer from the source domain to the target domain. We combined the gene expression signature of the cell line and the chemical structural signature of the drug as input to the model and successfully predicted the IC50 values through a deep neural network. From the results, through the deep transfer learning method, the information of the two databases was combined because only information from a single database was used. At the same time, the model showed excellent results in predicting unknown missing drug response pairs. In conclusion, our model can not only successfully perform information transfer between the two databases but also be used for real drug sensitivity prediction research.

From the performance comparison test, it can be seen that the feature extractor in the DADSP model is the cornerstone of the entire model performance results, so exploring better feature extraction methods for gene expression data and drug data will definitely

improve the basic performance of the model. For data, source domain data and target domain data also have inconsistent label spaces. In this work, the labels are normalized. Therefore, in future work, we plan to add a label distribution optimization scheme based on DADSP to achieve more reliable domain adaptation.

Author Contributions: Methodology, W.M., X.X. and L.Y.; Validation, L.Y.; Formal analysis, Z.X.; Investigation, L.G.; Data curation, W.M. and L.G.; Writing–original draft, X.X.; Writing–review & editing, Z.X.; Supervision, W.M., L.G. and L.Y.; Project administration, L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China [grant nos. 62472344, 62072353 and 62272065] and Xidian University Specially Funded Project for Interdisciplinary Exploration (No. TZJH2024027).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: Thanks to all those who maintain excellent databases and to all experimentalists who enabled this work by making their data publicly available.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Yang, Y.; Gao, D.; Xie, X.; Qin, J.; Li, J.; Lin, H.; Yan, D.; Deng, K. DeepIDC: A Prediction Framework of Injectable Drug Combination Based on Heterogeneous Information and Deep Learning. *Clin. Pharmacokinet.* **2022**, *61*, 1749–1759. [CrossRef]
2. Shaker, B.; Tran, K.M.; Jung, C.; Na, D. Introduction of Advanced Methods for Structure-based Drug Discovery. *Curr. Bioinform.* **2021**, *16*, 351–363. [CrossRef]
3. Shaker, B.; Ahmad, S.; Lee, J.; Jung, C.; Na, D. In silico methods and tools for drug discovery. *Comput. Biol. Med.* **2021**, *137*, 104851. [CrossRef]
4. Zeng, X.; Wang, F.; Luo, Y.; Kang, S.-G.; Tang, J.; Lightstone, F.C.; Fang, E.F.; Cornell, W.; Nussinov, R.; Cheng, F. Deep generative molecular design reshapes drug discovery. *Cell Rep. Med.* **2022**, *3*, 100794. [CrossRef]
5. Long, J.; Yang, H.; Yang, Z.; Jia, Q.; Liu, L.; Kong, L.; Cui, H.; Ding, S.; Qin, Q.; Zhang, N.; et al. Integrated biomarker profiling of the metabolome associated with impaired fasting glucose and type 2 diabetes mellitus in large-scale Chinese patients. *Clin. Transl. Med.* **2021**, *11*, e432. (In English) [CrossRef]
6. Cao, C.; Wang, J.; Kwok, D.; Cui, F.; Zhang, Z.; Zhao, D.; Li, M.J.; Zou, Q. webTWAS: A resource for disease candidate susceptibility genes identified by transcriptome-wide association study. *Nucleic Acids Res.* **2021**, *50*, D1123–D1130. [CrossRef]
7. Ding, Y.; Tang, J.; Guo, F.; Zou, Q. Identification of drug–target interactions via multiple kernel-based triple collaborative matrix factorization. *Brief. Bioinform.* **2022**, *23*, bbab582. [CrossRef]
8. Wang, Y.; Pang, C.; Wang, Y.; Jin, J.; Zhang, J.; Zeng, X.; Su, R.; Zou, Q.; Wei, L. Retrosynthesis prediction with an interpretable deep-learning framework based on molecular assembly tasks. *Nat. Commun.* **2023**, *14*, 6155. [CrossRef] [PubMed]
9. Tang, W.; Wan, S.; Yang, Z.; Teschendorff, A.E.; Zou, Q. Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics* **2018**, *34*, 398–406. [CrossRef]
10. Drews, J. Drug Discovery: A Historical Perspective. *Science* **2000**, *287*, 1960–1964. [CrossRef]
11. Zeng, X.; Xiang, H.; Yu, L.; Wang, J.; Li, K.; Nussinov, R.; Cheng, F. Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. *Nat. Mach. Intell.* **2022**, *4*, 1004–1016. [CrossRef]
12. Ru, X.Q.; Ye, X.C.; Sakurai, T.; Zou, Q. NerLTR-DTA: Drug-target binding affinity prediction based on neighbor relationship and learning to rank. *Bioinformatics* **2022**, *38*, 1964–1971. [CrossRef] [PubMed]
13. Andrade, R.C.; Boroni, M.; Amazonas, M.K.; Vargas, F.R. New drug candidates for osteosarcoma: Drug repurposing based on gene expression signature. *Comput. Biol. Med.* **2021**, *134*, 104470. [CrossRef]
14. Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Al, E. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **2012**, *483*, 603. [CrossRef] [PubMed]

15. Yang, W.; Soares, J.; Greninger, P.; Edelman, E.J.; Lightfoot, H.; Forbes, S.; Bindal, N.; Beare, D.; Smith, J.A.; Thompson, I.R.; et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **2013**, *41*, D955. [CrossRef]
16. Cortes-Ciriano, I.; Mervin, L.; Bender, A. Current Trends in Drug Sensitivity Prediction. *Curr. Pharm. Des.* **2016**, *22*, 6918–6927. [CrossRef]
17. Menden, M.P.; Iorio, F.; Garnett, M.; McDermott, U.; Benes, C.H.; Ballester, P.J.; Saez-Rodriguez, J. Machine Learning Prediction of Cancer Cell Sensitivity to Drugs Based on Genomic and Chemical Properties. *PLoS ONE* **2013**, *8*, e61318. [CrossRef]
18. Zhang, N.; Wang, H.; Fang, Y.; Wang, J.; Zheng, X.; Liu, X.S. Predicting Anticancer Drug Responses Using a Dual-Layer Integrated Cell Line-Drug Network Model. *PLoS Comput. Biol.* **2015**, *11*, e1004498. [CrossRef]
19. Mehmet, G.; Margolin, A.A. Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning. *Bioinformatics* **2014**, *30*, 556–563.
20. Su, R.; Liu, X.; Wei, L.; Zou, Q. Deep-Resp-Forest: A deep forest model to predict anti-cancer drug response. *Methods* **2019**, *166*, 91–102. (In English) [CrossRef]
21. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef]
22. Madugula, S.S.; John, L.; Nagamani, S.; Gaur, A.S.; Poroikov, V.V.; Sastry, G.N. Molecular descriptor analysis of approved drugs using unsupervised learning for drug repurposing. *Comput. Biol. Med.* **2021**, *138*, 104856. [CrossRef]
23. Jin, J.; Yu, Y.; Wang, R.; Zeng, X.; Pang, C.; Jiang, Y.; Li, Z.; Dai, Y.; Su, R.; Zou, Q.; et al. iDNA-ABF: Multi-scale deep biological language learning model for the interpretable prediction of DNA methylations. *Genome Biol.* **2022**, *23*, 219. [CrossRef]
24. Li, H.-L.; Pang, Y.-H.; Liu, B. BioSeq-BLM: A platform for analyzing DNA, RNA and protein sequences based on biological language models. *Nucleic Acids Res.* **2021**, *49*, e129. [CrossRef]
25. Li, H.; Liu, B. BioSeq-Diabolo: Biological sequence similarity analysis using Diabolo. *PLoS Comput. Biol.* **2023**, *19*, e1011214. [CrossRef]
26. Chiu, Y.C.; Chen, H.I.H.; Zhang, T.; Zhang, S.; Gorthi, A.; Wang, L.J.; Huang, Y.; Chen, Y. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med. Genom.* **2019**, *12*, 143–155.
27. Li, M.; Wang, Y.; Zheng, R.; Shi, X.; Li, Y.; Wu, F.-X.; Wang, J. DeepDSC: A Deep Learning Method to Predict Drug Sensitivity of Cancer Cell Lines. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *18*, 575–582. [CrossRef]
28. Nguyen, T.; Nguyen, G.; Nguyen, T.V.; Le, D.H. Graph Convolutional Networks for Drug Response Prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 146–154. [CrossRef]
29. Haibe-Kains, B.; El-Hachem, N.; Birkbak, N.J.; Jin, A.C.; Beck, A.H.; Aerts, H.J.; Quackenbush, J. Inconsistency in large pharmacogenomic studies. *Nature* **2019**, *504*, 389–393. [CrossRef]
30. Stransky, N.; Ghandi, M.; Kryukov, G.V.; Garraway, L.A.; Saez-Rodriguez, J. Pharmacogenomic agreement between two cancer cell line data sets. *Nature* **2015**, *528*, 84–87.
31. Dhruba, S.R.; Rahman, R.; Matlock, K.; Ghosh, S.; Pal, R. Application of transfer learning for cancer drug sensitivity prediction. *BMC Bioinform.* **2018**, *19*, 51–63. [CrossRef] [PubMed]
32. Tan, B.; Song, Y.; Zhong, E.; Qiang, Y. Transitive Transfer Learning. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Sydney, Australia, 10–13 August 2015.
33. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
34. Kleinbaum, D.G.; Dietz, K.; Gail, M.; Klein, M.; Klein, M. *Logistic Regression*; Springer: New York, NY, USA, 2002.
35. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
36. Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [CrossRef]
37. Eckert, H.; Bajorath, J. Molecular similarity analysis in virtual screening: Foundations, limitations and novel approaches. *Drug Discov. Today* **2007**, *12*, 225–233. [CrossRef]
38. Liu, P.; Li, H.; Li, S.; Leung, K.-S. Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. *BMC Bioinform.* **2019**, *20*, 408. [CrossRef]
39. Ramsundar, B.; Eastman, P.; Walters, P.; Pande, V. *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More*; O'Reilly Media: Sebastopol, CA, USA, 2019.
40. Vita, V.T.D.; Hellman, S.; Rosenberg, S.A. Cancer: Principles & practice of oncology. *Eur. J. Cancer Care* **2005**.
41. Corton, J.M.; Gillespie, J.G.; Hawley, S.A.; Hardie, D.G. 5-aminoimidazole-4-carboxamide ribonucleoside. A specific method for activating AMP-activated protein kinase in intact cells? *FEBS J.* **2010**, *229*, 558–565. [CrossRef]
42. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic attribution for deep networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 3319–3328.
43. Xu, G.; Fan, L.; Zhao, S.; OuYang, C. MT1G inhibits the growth and epithelial-mesenchymal transition of gastric cancer cells by regulating the PI3K/AKT signaling pathway. *Genet. Mol. Biol.* **2022**, *45*, e20210067. [CrossRef]

44. Jadhav, R.R.; Ye, Z.; Huang, R.-L.; Liu, J.; Hsu, P.-Y.; Huang, Y.-W.; Rangel, L.B.; Lai, H.-C.; Roa, J.C.; Kirma, N.B.; et al. Genome-wide DNA methylation analysis reveals estrogen-mediated epigenetic repression of metallothionein-1 gene cluster in breast cancer. *Clin. Epigenetics* **2015**, *7*, 13. [CrossRef]
45. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **2012**, *16*, 284–287. [CrossRef]
46. Ru, X.; Wang, L.; Li, L.; Ding, H.; Ye, X.; Zou, Q. Exploration of the correlation between GPCRs and drugs based on a learning to rank algorithm. *Comput. Biol. Med.* **2020**, *119*, 103660. [CrossRef]
47. Wang, L.; Li, J.; Liu, E.; Kinnebrew, G.; Zhang, X.; Stover, D.; Huo, Y.; Zeng, Z.; Jiang, W.; Cheng, L.; et al. Identification of Alternatively-Activated Pathways between Primary Breast Cancer and Liver Metastatic Cancer Using Microarray Data. *Genes* **2019**, *10*, 753. [CrossRef]
48. Dang, Y.-W.; Lin, P.; Liu, L.-M.; He, R.-Q.; Zhang, L.-J.; Peng, Z.-G.; Li, X.-J.; Chen, G. In silico analysis of the potential mechanism of telocinobufagin on breast cancer MCF-7 cells. *Pathol.-Res. Pract.* **2018**, *214*, 631–643. [CrossRef]
49. Karnoub, A.E.; Weinberg, R.A.; Wakefield, L.; Hunter, K. Chemokine networks and breast cancer metastasis. *Breast Dis.* **2007**, *26*, 75. [CrossRef]
50. Khabele, D.; Lopez-Jones, M.; Yang, W.; Arango, D.; Gross, S.J.; Augenlicht, L.H.; Goldberg, G.L. Tumor necrosis factor-α related gene response to Epothilone B in ovarian cancer. *Gynecol. Oncol.* **2004**, *93*, 19–26. [CrossRef]
51. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
52. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [CrossRef]
53. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [CrossRef]
54. O'boyle, N.M. Towards a Universal SMILES representation—A standard method to generate canonical SMILES based on the InChI. *J. Cheminforma.* **2012**, *4*, 22. [CrossRef]
55. Bento, A.P.; Hersey, A.; Felix, E.; Landrum, G.; Leach, A.R. An Open Source Chemical Structure Curation Pipeline using RDKit. *J. Cheminform.* **2020**, *12*, 51. [CrossRef]
56. Min, S.; Lee, B.; Yoon, S. Deep learning in bioinformatics. *Brief. Bioinform.* **2016**, *18*, 851–869. [CrossRef] [PubMed]
57. Li, Y.; Huang, C.; Ding, L.; Li, Z.; Pan, Y.; Gao, X. Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. *Methods* **2019**, *166*, 4–21. [CrossRef] [PubMed]
58. Zulfiqar, H.; Huang, Q.-L.; Lv, H.; Sun, Z.-J.; Dao, F.-Y.; Lin, H. Deep-4mCGP: A Deep Learning Approach to Predict 4mC Sites in *Geobacter pickeringii* by Using Correlation-Based Feature Selection Technique. *Int. J. Mol. Sci.* **2022**, *23*, 1251. [CrossRef] [PubMed]
59. Lv, H.; Dao, F.; Lin, H. DeepKla: An attention mechanism-based deep neural network for protein lysine lactylation site prediction. *iMeta* **2022**, *1*, e11. [CrossRef]
60. Liu, M.; Li, C.; Chen, R.; Cao, D.; Zeng, X. Geometric Deep Learning for Drug Discovery. *Expert Syst. Appl.* **2023**, *240*, 122498. [CrossRef]
61. Xu, J.; Xu, J.; Meng, Y.; Lu, C.; Cai, L.; Zeng, X.; Nussinov, R.; Cheng, F. Graph embedding and Gaussian mixture variational autoencoder network for end-to-end analysis of single-cell RNA sequencing data. *Cell Rep. Methods* **2023**, *3*, 100382. [CrossRef]
62. Wang, R.; Jiang, Y.; Jin, J.; Yin, C.; Yu, H.; Wang, F.; Feng, J.; Su, R.; Nakai, K.; Zou, Q.; et al. DeepBIO: An automated and interpretable deep-learning platform for high-throughput biological sequence prediction, functional annotation and visualization analysis. *Nucleic Acids Res.* **2023**, *51*, 3017–3029. [CrossRef]
63. Tang, Y.-J.; Pang, Y.-H.; Liu, B. IDP-Seq2Seq: Identification of intrinsically disordered regions based on sequence to sequence learning. *Bioinformatics* **2020**, *36*, 5177–5186. [CrossRef]
64. Yan, K.; Lv, H.; Guo, Y.; Peng, W.; Liu, B. sAMPpred-GAT: Prediction of Antimicrobial Peptide by Graph Attention Network and Predicted Peptide Structure. *Bioinformatics* **2023**, *39*, btac715. [CrossRef]
65. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*. [CrossRef]
66. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
67. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
68. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. *Adv. Neural Inf. Process. Syst.* **2014**.
69. Long, M.; Cao, Y.; Cao, Z.; Wang, J.; Jordan, M.I. Transferable Representation Learning with Deep Adaptation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 3071–3085. [CrossRef]
70. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.

71. Joshi, P.; Vedhanayagam, M.; Ramesh, R. An Ensembled SVM Based Approach for Predicting Adverse Drug Reactions. *Curr. Bioinform.* **2021**, *16*, 422–432. [CrossRef]
72. You, K.; Long, M.; Cao, Z.; Wang, J.; Jordan, M.I. Universal domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 2720–2729.
73. Wang, M.; Deng, W. Deep visual domain adaptation: A survey. *Neurocomputing* **2018**, *312*, 135–153. [CrossRef]
74. Baldi, P.; Guyon, G.; Dror, V.; Lemaire, G.; Taylor, G.; Silver, D. Autoencoders, unsupervised learning and deep architectures. In Proceedings of the UTLW'11 2011 International Conference on Unsupervised and Transfer Learning Workshop, Washington, DC, USA, 2 July 2011; Volume 27.
75. Gehring, J.; Miao, Y.; Metze, F.; Waibel, A. Extracting deep bottleneck features using stacked auto-encoders. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 3377–3381.
76. An, N.; Zhao, W.; Wang, J.; Shang, D.; Zhao, E. Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. *Energy* **2013**, *49*, 279–288. [CrossRef]
77. Bhaskar, K.; Singh, S.N. AWNN-Assisted Wind Power Forecasting Using Feed-Forward Neural Network. *IEEE Trans. Sustain. Energy* **2012**, *3*, 306–315. [CrossRef]
78. Tran, D.; Tan, Y.K. Sensorless Illumination Control of a Networked LED-Lighting System Using Feedforward Neural Network. *IEEE Trans. Ind. Electron.* **2013**, *61*, 2113–2121. [CrossRef]
79. Luo, H.; Wang, J.; Li, M.; Luo, J.; Ni, P.; Zhao, K.; Wu, F.-X.; Pan, Y. Computational Drug Repositioning with Random Walk on a Heterogeneous Network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**, *16*, 1890–1900. [CrossRef] [PubMed]
80. Liu, X.; Yang, H.; Ai, C.; Ding, Y.; Guo, F.; Tang, J. MVML-MPI: Multi-View Multi-Label Learning for Metabolic Pathway Inference. *Brief. Bioinform.* **2023**, *24*, bbad393. [CrossRef] [PubMed]
81. Dou, M.; Ding, J.; Chen, G.; Duan, J.; Guo, F.; Tang, J. IK-DDI: A novel framework based on instance position embedding and key external text for DDI extraction. *Brief. Bioinform.* **2023**, *24*, bbad099. [CrossRef] [PubMed]
82. Choi, J.; Park, S.; Ahn, J. RefDNN: A reference drug based neural network for more accurate prediction of anticancer drug resistance. *Sci. Rep.* **2020**, *10*, 1861. [CrossRef]
83. Yang, H.; Luo, Y.-M.; Ma, C.-Y.; Zhang, T.-Y.; Zhou, T.; Ren, X.-L.; He, X.-L.; Deng, K.-J.; Yan, D.; Tang, H.; et al. A gender specific risk assessment of coronary heart disease based on physical examination data. *Npj Digit. Med.* **2023**, *6*, 136. [CrossRef]
84. Zhang, Z.Y.; Ning, L.; Ye, X.; Yang, Y.H.; Futamura, Y.; Sakurai, T.; Lin, H. iLoc-miRNA: Extracellular/intracellular miRNA prediction using deep BiLSTM with attention mechanism. *Brief. Bioinform.* **2022**, *23*, bbac395. [CrossRef]
85. Wang, Y.; Zhai, Y.; Ding, Y.; Zou, Q. SBSM-Pro: Support Bio-sequence Machine for Proteins. *arXiv* **2023**, arXiv:2308.10275. [CrossRef]
86. Jang, I.S.; Neto, E.C.; Guinney, J.; Friend, S.H.; Margolin, A.A. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. *Biocomputing* **2013**, *19*, 63–74.
87. Borgwardt, K.M.; Gretton, A.; Rasch, M.J.; Kriegel, H.-P.; Schölkopf, B.; Smola, A.J. Integrating structured biological data by Kernel Maximum Mean Discrepancy. *Bioinformatics* **2006**, *22*, e49–e57. [CrossRef]
88. Tzeng, E.; Hoffman, J.; Zhang, N.; Saenko, K.; Darrell, T. Deep Domain Confusion: Maximizing for Domain Invariance. Computer Science. *arXiv* **2014**, arXiv:1412.3474.
89. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. *How Transferable Are Features in Deep Neural Networks?* MIT Press: Cambridge, MA, USA, 2014.
90. Zhuang, F.; Cheng, X.; Luo, P.; Pan, S.J.; He, Q. Supervised Representation Learning with Double Encoding-Layer Autoencoder for Transfer Learning. *ACM Trans. Intell. Syst. Technol.* **2018**, *9*, 1–17. [CrossRef]
91. Caruana, R. Multitask Learning. *Mach. Learn.* **1997**, *28*, 41–75. [CrossRef]