Heliyon 11 (2025) e42476

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

InceptionDTA: Predicting drug-target binding affinity with biological context features and inception networks

Mahmood Kalemati, Mojtaba Zamani Emani, Somayyeh Koohi *

Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

ARTICLE INFO

Keywords:
Drug-target binding affinity prediction
Deep representation learning
Inception network
Interaction
CharVec encoding

ABSTRACT

Predicting drug-target binding affinity via in silico methods is crucial in drug discovery. Traditional machine learning relies on manually engineered features from limited data, leading to suboptimal performance. In contrast, deep learning excels at extracting features from raw sequences but often overlooks essential biological context features, hindering effective binding prediction. Additionally, these models struggle to capture global and local feature distributions efficiently in protein sequences and drug SMILES. Previous state-of-the-art models, like transformers and graph-based approaches, face scalability and resource efficiency challenges. Transformers struggle with scalability, while graph-based methods have difficulty handling large datasets and complex molecular structures. In this paper, we introduce InceptionDTA, a novel drug-target binding affinity prediction model that leverages CharVec, an enhanced variant of Prot2Vec, to incorporate both biological context and categorical features into protein sequence encoding. InceptionDTA utilizes a multi-scale convolutional architecture based on the Inception network to capture features at various spatial resolutions, enabling the extraction of both local and global features from protein sequences and drug SMILES. We evaluate InceptionDTA across a range of benchmark datasets commonly used in drug-target binding affinity prediction. Our results demonstrate that InceptionDTA outperforms various sequence-based, transformer-based, and graph-based deep learning approaches across warm-start, refined, and cold-start splitting settings. In addition to using CharVec, which demonstrates greater accuracy in absolute predictions, InceptionDTA also includes a version that employs simple label encoding and excels in ranking and predicting relative binding affinities. This versatility highlights how InceptionDTA can effectively adapt to various predictive requirements. These results emphasize the promise of our approach in expediting drug repurposing initiatives, enabling the discovery of new drugs, and contributing to advancements in disease treatment.

1. Introduction

Predicting drug-target binding affinity is a complex challenge in drug discovery. Traditionally, it is approached as a classification task based on drug-target similarity, often neglecting the strength of interactions [1] [2], [3], [4]. Conversely, binding affinity prediction can be formulated as a regression task, incorporating continuous values to represent the affinity strength, thus offering a more realistic and informative model for drug-target interactions [5]. Despite the availability of wet lab experiments and simulation-based

* Corresponding author.
E-mail address: koohi@sharif.edu (S. Koohi).

https://doi.org/10.1016/j.heliyon.2025.e42476
Received 3 July 2024; Received in revised form 23 January 2025; Accepted 4 February 2025
Available online 5 February 2025

methods, they remain prohibitively expensive, time-consuming, and reliant on molecular structures [6] [7].

To address the need for cost-effective prediction methods, machine learning techniques have proven effective in speeding up predictions using available protein and drug sequences. An earlier method, the Kronecker Regularized Least Squares (KronRLS) algorithm, utilizes compound similarity and Smith-Waterman similarity for drugs and proteins, respectively [5] [8], [9]. However, its reliance on linear algorithms limits its ability to capture nonlinear relationships and it employs simplistic feature constructions. In contrast, SimBoost uses gradient boosting with features from drug-drug, protein-protein, and network interactions [8], [10]. Although it captures complex dependencies, SimBoost requires expert knowledge for feature engineering and is computationally intensive, relying on tools like the PubChem clustering server [11].

To address the limitations of similarity-based methods, including manual feature definition and reliance on pre-existing knowledge, sequence-based deep neural networks have emerged as more effective alternatives for predicting drug-target binding affinity. For example, DeepDTA [12] utilizes convolutional neural networks (CNNs) to automatically extract features from raw sequences of proteins and drugs. Although this approach advances beyond traditional methods by eliminating manual feature curation, its effectiveness is limited by its dependence on labeled data and simplistic input encoding, which may overlook crucial chemical and biological context. Another method, WideDTA [13], integrates multiple data sources, such as protein sequences, domains, motifs, ligand SMILES (Simplified Molecular Input Line Entry System), and common substructures. However, despite these enhancements, the reliance on basic label encoding may still be insufficient for achieving precise binding affinity predictions. GANsDTA [14] employs a semi-supervised approach using generative adversarial networks (GANs) [15] to enhance binding affinity predictions by incorporating unlabeled data for feature extraction. However, its reliance on a simple fully connected GANs may not effectively extract local patterns in protein sequences and drug SMILES.

Recent advancements have explored Transformer architectures [16] to capture long-range dependencies in biological and chemical sequences, aiming for more precise embeddings of drug and target sequences. A notable example is MT-DTI [17], which uses a molecule transformer to improve drug representation, complemented by CNN blocks for extracting protein features. However, this approach demands extensive computational resources, requiring a many-core TPU machine for training. Another significant development is TEFDTA [18], which introduces an attention-driven transformer encoder to convert drug SMILES into MACCS (Molecular Access System) fingerprints, capturing detailed substructure information crucial for predicting binding affinities. Additionally, G-K BertDTA [19] combines knowledge-based BERT (KB-BERT) with a graph isomorphism network (GIN) to extract semantic features from SMILES molecules, enhancing the learning of relational features between isomorphic SMILES structures. While these Transformer-based methods effectively capture complex dependencies, they face challenges in managing computational time and space, particularly as datasets grow larger.

To address inefficiencies in encoding and representation, recent studies [20] [21], [22], have turned to graph neural network architectures for feature extraction from drug and protein structures. For instance, GraphDTA [20] uses graph representations to model drugs and applies Graph Neural Networks (GNNs) for feature extraction. Similarly, Dynamic Graph DTA (DGDTA) [23] introduces a dynamic graph attention network to evaluate the importance of drug features, coupled with a bidirectional Long Short-Term Memory (Bi-LSTM) network for contextual feature extraction from protein sequences. GPCNDTA [24] creates drug graphs from 2D topology and physicochemical properties, and protein graphs from contact matrices and residue properties. It uses a residual CensNet and residual EW-GCN to extract features from these graphs. While these methods have demonstrated high accuracy in prediction tasks, they face significant challenges related to scalability, particularly when handling extensive datasets and complex molecular structures.

To overcome the limitations of existing methods—such as the inadequacy of simple label encoding, the need for expert knowledge in feature selection and engineering, dependence on the availability of labeled data, and high computational costs—this study enhances the Prot2Vec [25] representation of amino acid sequences. Originally trained with a Skipgram neural network [26], we introduce a novel protein encoding scheme named CharVec. CharVec encoding captures biological context from amino acid sequences at the character level, providing richer representations that reflect their biological significance. CharVec is specifically designed to address these challenges and improve the accuracy of drug-target binding affinity predictions.

Additionally, we integrate Inception-based modules [27] to effectively capture both local and global features of proteins and drugs. Inception networks are a deep learning architecture that employs multiple convolutional filters of varying sizes in parallel, enabling the model to learn features at different scales within the same layer. Previous research [28], [29] [30], demonstrates the utility of inception networks in drug-target interaction (DTI) and predicting binding affinity. For example, MDeePred [28] uses inception modules to combine multi-channel protein features and drug fingerprints, while DeepConv-DTI [29] employs inception-like modules for protein features and a fully-connected block for drug fingerprints. However, these methods rely on diverse molecular fingerprints and simple feedforward networks, which may limit robust drug representation. FRnet-DTI [30] uses molecular fingerprints for drugs and structural and evolutionary features for proteins with two-dimensional inception modules. However, this approach requires additional feature generation for both proteins and drugs.

Motivated by the limitations of current approaches—including their reliance on diverse features, complex feature construction and selection processes, and costly two-dimensional inception modules—we aim to advance DTA prediction by leveraging inception networks. Our method enhances the ability to learn comprehensive and accurate representations directly from raw sequence data for both proteins and drugs, reducing dependence on pre-processed features and simplifying the prediction process. Our approach comprises five key components: (a) an unsupervised model for generating primary vector representations of individual amino acids, (b) protein sequence embedding based on the presented amino acid encoding (i.e., CharVec), while drug sequences are encoded by label encoding, (c) two separate Inception-based networks for feature extraction and representation learning of both drugs and proteins, (d) a dense layer for concatenating the learned representations of both input sequences, and finally, (e) a fully connected block for predicting binding affinities.

In summary, the primary contributions of InceptionDTA can be outlined as follows.

- Enhanced Protein Representation with CharVec: We introduce CharVec, an improved version of Prot2Vec, for encoding amino acids to better capture biological and categorical features crucial for accurate drug-target binding affinity predictions.
- Advanced Feature Extraction Techniques: InceptionDTA utilizes Inception networks to capture multi-scale features at various spatial resolutions from protein sequences and drug SMILES, ensuring comprehensive and precise representations for binding affinity prediction.
- Efficiency and Adaptability: By design, Inception networks optimize computational resources through the efficient use of varying kernel sizes and pooling operations, balancing reduced computational load with high feature extraction accuracy.

2. Methods

2.1. Datasets

We assessed our proposed method using three extensively utilized benchmark datasets for drug-target binding affinity prediction. The first dataset, referred to as the Davis dataset [31], comprises binding affinity values of 68 drugs and 442 proteins, measured by the kinase dissociation constant (Kd). Following the approach outlined in [8] and [12], we employed log-transformed Kd values (pKd), as depicted in Equation (1).

$$pK_d = -\log_{10}\left(\frac{K_d}{1e9}\right) \tag{1}$$

For the second dataset, the Kiba dataset [32] encompasses binding affinities of 2111 drugs and 229 proteins, quantified by a singular bioactivity score known as the Kiba score [32]. These scores are derived from statistical amalgamations of diverse measures, including dissociation constant (Kd), inhibition constant (Ki), and half maximal inhibitory concentration (IC50). For the third dataset, we opted for the refined version of the PDBbind dataset, chosen for its superior quality compared to the general and core sets [33]. To maintain data integrity, we eliminated redundancies stemming from multiple sequences corresponding to the same drugs. This refined set comprises logarithmically transformed Ki and Kd binding values for 1606 protein sequences and 4231 drug SMILES. Table 1 provides a summary of three datasets. Additionally, we conducted an evaluation of the proposed method utilizing an enhanced iteration of the Davis dataset. Specifically, we recognized that redundancy and biases within the data, stemming from multiple annotations for similar proteins and drugs, could potentially impact the performance of DTA prediction methods. In response to this concern, we scrutinized and compared the efficacy of our proposed method against alternative approaches using a refined rendition of the Davis dataset. To achieve this, we excluded redundant protein sequences from the Davis dataset.

2.2. InceptionDTA method

The proposed approach, termed InceptionDTA, comprises four primary steps: a) encoder, b) feature extractor, c) concatenator, and d) predictor.

A) Encoder: In the initial stage, we encode drug and protein sequences utilizing two distinct methodologies. For drug sequences, represented in SMILES format [34], we employ the label encoding technique, commonly utilized in previous studies. Specifically, each character in the sequences is assigned a unique integer value (e.g., 'N' = 1, 'C' = 2, ' = ' = 3, '1' = 4, '2' = 5, etc.). Consequently, each sequence is converted into an integer sequence of the same length (e.g., "C=CC2=NN1″ encodes to [2 3 2 2 5 3 1 1 4]).

For protein representation, we introduce a novel amino acid encoding scheme that leverages a word embedding model trained on monomer units of proteins, known as amino acids [35]. We posit that such a representation encompasses richer biological features compared to simple label encoding. Additionally, we anticipate that this approach retains more contextual dependency, as well as biological and chemical characteristics, compared to traditional one-hot encoding and k-mer-based word embedding techniques [25]. To realize these objectives, we follow a five-step process outlined in Fig. 1 for amino acid encoding using the CharVec method.

Table 1
Datasets' information.

	Proteins	Compounds	Interactions	Affinity measure
Davis	442	68	30056	Dissociation constant(Kd)
Kiba	229	2111	118254	Inhibition constant(Ki) Dissociation constant(Kd) Half-maximal inhibitory concentration(IC50)
PDBbind	1606	4231	5014	Inhibition constant (Ki) Dissociation constant (Kd)

1) Data Acquisition: In the initial phase, we employ the widely recognized protein embedding approach, Prot2Vec [25], trained using the Skipgram neural network architecture on amino acid sequences [26] [35], to generate primary vector representations of amino acids. The model is trained at the character level, focusing on individual amino acids rather than n-gram segments (i.e., k-mers). For character-level training to generate Prot2Vec representation, the entire Swiss-Prot dataset, which contains 600,000 protein sequences is utilized [35].
2) Primary Dictionary (PD) Construction: In the second step, we construct a dictionary comprising amino acid letters and their corresponding pre-trained vectors for each individual amino acid. To achieve this, each amino acid letter $X_i \in$ {A,B,C,D,E,F,G,H, I,K,L,M,N,O,P,Q,R,S,T,U,W,V,Y,X,Z} is assigned as the key, while the corresponding 50-dimensional distributed vector (e.g., "D" is embedded into a 1 × 50 vector as [0.1651, 0.0419, 0.0192, …, 0.01807, 0.1077, 0.0123]) is designated as the value.
3) Maximum Value Extraction: In the third step, the representative context feature for each key is identified by extracting the maximum value from the corresponding vector in the primary dictionary (e.g., 0.1651 for "D"). Essentially, only the largest value within each vector is retained as the representative value for the associated key (i.e., amino acid letters), while the remaining values are discarded.
4) Secondary Dictionary (SD) Construction: In the fourth step, we create a new dictionary comprising amino acid letters as keys and their associated 1 × 20 one-hot vectors as values. To achieve this, we multiply the extracted maximum value for each key by the corresponding one-hot vector for each amino acid, as illustrated in Fig. 1. For instance, the representative vector for "D" is encoded as a 1 × 20 vector [0.1651, 0.0, 0.0 … 0.0, 0.0, 0.0]. This approach ensures that the new vectors incorporate both context features from Prot2Vec and categorical features from one-hot encoding. Furthermore, by reducing the dimensionality from 50 to 20, the new vectors achieve decreased computational complexity and enable a more scalable model compared to the original Prot2Vec.
5) Protein Sequence Embedding. In the fifth and final step, each protein sequence $P = {}'X_1X_2X_3 \ldots X_{L-2}X_{L-1}X_L{}'$, where Xi {i ϵ [1, L]} represent amino acid letters and L denotes the length of P, is represented by the subsequent encoded amino acids from SD, resulting in a matrix of size L × 20.

B) Feature Extractor: In the second step, we utilize two distinct inception blocks for representation learning and feature extraction from the encoded protein and drug sequences. The encoded SMILES sequence undergoes processing via an embedding layer,

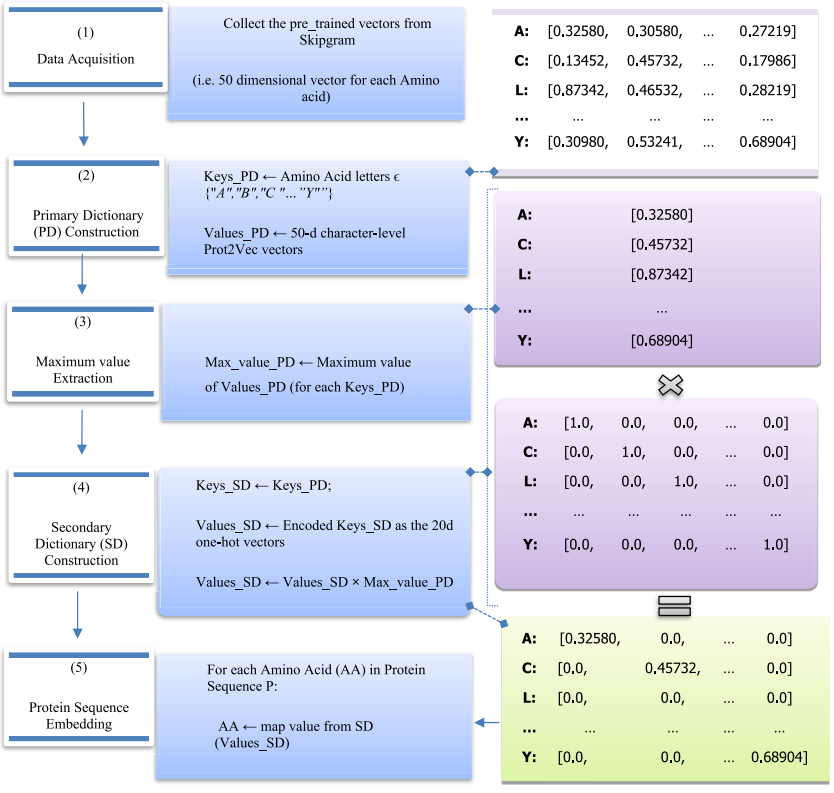

Fig. 1. CharVec Encoding Scheme. Step 1: Utilize Prot2Vec for primary vector representations of amino acids. Step 2: Construct the Primary Dictionary (PD) with amino acid letters and pre-trained vectors. Step 3: Extract maximum values from PD for representative context features. Step 4: Multiply each extracted maximum value by the corresponding one-hot vector for amino acids to form the Secondary Dictionary (SD). Step 5: Embed protein sequences P of length L using encoded amino acids from SD, resulting in an Lx20 matrix.

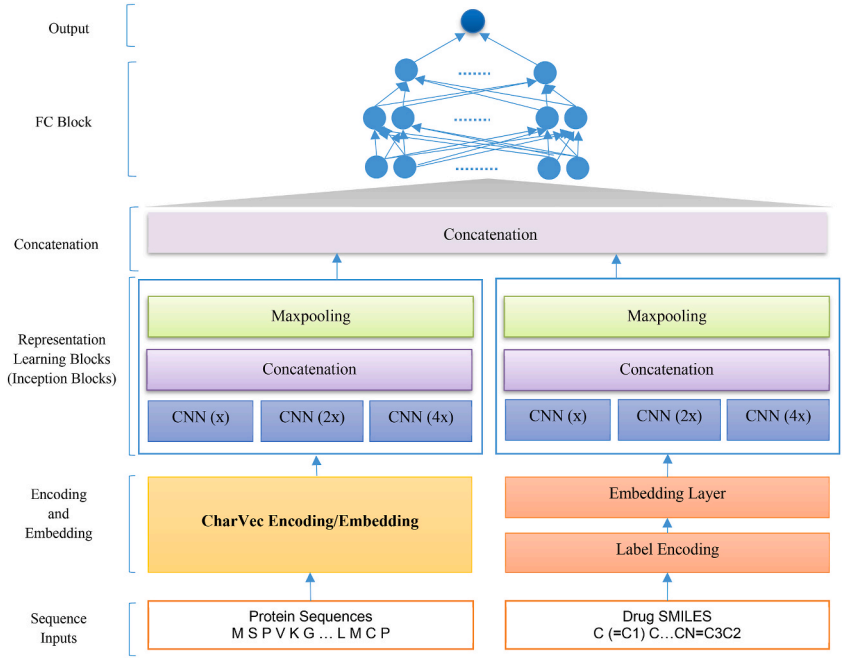

Fig. 2. The proposed deep neural network architecture for binding affinity prediction. Protein sequences are embedded at the character level using CharVec encoding based on a pre-trained Skipgram model, while drug sequences in SMILES format are encoded using simple label encoding. The encoded sequences are then passed through an inception block consisting of three parallel CNN layers with filter sizes denoted by 'x', followed by concatenation and max-pooling layers. The learned representations of both inputs are concatenated and fed into a predictor comprising three fully connected layers followed by an output layer for final prediction.

which is then fed into an inception block comprising three parallel CNN layers with varying filter sizes. This architecture is designed to capture both the global and local distribution of features within the sequences.

C) Concatenator: In the third step, the learned representations of both inputs are concatenated and forwarded to the predictor.

(D) In the final step, three layers of a fully connected block followed by an output layer are employed to provide the ultimate prediction. Fig. 2 illustrates the proposed architecture for drug-target binding affinity prediction.

It is worth noting that we utilize the widely used Rectified Linear Unit (ReLU) [36] as the activation function, and the mean squared error (MSE) as the loss function. The loss function is defined and applied according to Equation (2), where n represents the number of samples, and P and Y denote the predicted and actual binding affinity values, respectively.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - Y_i)^2 \tag{2}$$

Lastly, the model was trained for 300 epochs with a batch size of 256 and a learning rate of 0.001. Adam [37] was employed as the optimization algorithm for training the model.

3. Results

3.1. Evaluation metrics

In this section, we evaluate the performance of InceptionDTA using four metrics widely employed for the evaluation of drug-target binding affinity prediction in prior studies [5], [8], [12]. These metrics include the mean squared error (MSE), concordance index (CI), r_m^2 and AUPR. The mean squared error (MSE), defined in Equation (2) quantifies the disparity between actual and predicted binding affinity values. It provides a direct measure of prediction accuracy by penalizing larger errors more heavily, making it a crucial metric for evaluating how well the predictions align with true values. On the other hand, the concordance index assesses the predictive accuracy of a regression model [38]. It is especially useful for predicting the relative order of affinity values rather than exact values. It is determined by Equation (3), where f_i and f_j represent the predicted binding affinity values corresponding to actual affinity values y_i and y_j ($y_i > y_j$), respectively.

$$CI = \frac{1}{z}\sum_{y_i > y_j} h\left(f_i - f_j\right) \tag{3}$$

Here, z represents the normalization constant, and h is the step function defined according to Equation (4). The concordance index (c-index) is a ranking metric that assesses whether the predicted and actual binding affinity values of two random drug-target pairs are in the correct order.

$$h(x) = \begin{cases} 1, if\ x > 0 \\ 0.5, if\ x = 0 \\ 0, if\ x < 0 \end{cases} \tag{4}$$

In addition to the aforementioned metrics, we further assessed the prediction performance of InceptionDTA by incorporating two other metrics: r_m^2 and AUPR. The metric r_m^2, commonly employed to evaluate the external prediction performance of a model [8], [12], measures the squared correlation coefficient values with and without intercept, denoted as r^2 and r_0^2, respectively [39], [40]. A model is deemed acceptable if r_m^2 exceeds 0.5, indicating its effectiveness in prediction. A higher r_m^2 signifies better model fit and effectiveness in capturing data patterns.

$$r_m^2 = r^2 \times \left(1 - \sqrt{(r^2 - r_0^2)}\right) \tag{5}$$

As the final metric, we employed the Area Under Precision-Recall curve (AUPR) for binary classification. AUPR assesses performance by summarizing the trade-off between precision and recall across thresholds, reflecting how well the model identifies positive instances like strong binding interactions. To facilitate this evaluation, we transformed the binding affinity values into binary values using predefined thresholds. Threshold values are selected based on the distribution of affinity scores in the datasets to balance positive and negative samples and to highlight biologically relevant interactions [5]. This balance is crucial for enabling the model to accurately differentiate between strong and weak interactions, thus optimizing its performance [5]. Specifically, we selected pK_d values of 7 and 12.1 for the Davis and Kiba datasets, respectively, consistent with prior studies [8], [12]. Similarly, the threshold for the PDBbind dataset was set to 7.

In summary, the selected metrics for evaluating InceptionDTA were chosen for their ability to provide a comprehensive assessment of performance in drug-target binding affinity prediction. These metrics collectively offer a robust evaluation across various aspects, including accuracy, ranking capability, variance explanation, and classification effectiveness. The use of these metrics aligns with prior studies and provides a thorough understanding of how well InceptionDTA predicts drug-target binding affinities.

3.2. Implementation details

We evaluated the performance of InceptionDTA for drug-target binding affinity prediction on three benchmark datasets, following a methodology akin to previous studies. Each dataset was partitioned into multiple training folds and a disjoint test set. We adopted a 5-fold cross-validation technique, similar to that performed in DeepDTA, wherein the training data in each dataset was split into five folds. Specifically, each dataset was randomly divided into six equal parts. One part was reserved as an independent test set, while the remaining five parts were used for training and hyperparameter tuning through 5-fold cross-validation. The model was trained and tuned using the training folds, with each validation set utilized for the corresponding iteration. Subsequently, we evaluated the model using the disjoint test set to assess its capability for making accurate predictions for unseen data. Table 2 outlines the parameter settings for InceptionDTA. Note that we used a customized single-layer Inception network, where each block consists of three parallel CNN layers, excluding the parallel max pooling layer. This design balances prediction accuracy with runtime efficiency. The method was implemented using popular Python libraries, Keras [41] and Tensorflow [42], on NVIDIA GeForce GTX 1080 with 11 GB of available memory.

3.3. Comparative studies

As discussed previously, InceptionDTA utilizes CharVec, a novel amino acid encoding scheme for character-level protein embedding. It employs separate inception networks for representation learning and a fully connected network for binding affinity prediction. In this section, we compared the performance of InceptionDTA with several popular binding affinity prediction methods

Table 2
Parameter settings for two inception and the FC network.

Length of SMILES(drug) sequences	85, 100, 200
Length of protein sequences	1000
Filter length	4, 8, 16
Number of filters	128, 256, 384
Number of neurons	1024,1024,512
Number of epochs	300
Batch size	256
Dropout [43]	0.1
Optimizer function	adam
Learning rate	0.001

using four widely used metrics: CI, MSE, r_m^2, and AUPR. We employed warm-start, refined, and cold-start data splitting settings for evaluation. In the warm-start setting, both the drugs and the targets in the test set have been seen by the model during training, but their specific combinations may not have been encountered. The refined setting reduces redundancies and biases by excluding similar drugs and proteins that have been annotated multiple times. In the cold-start setting, either the drugs, the targets, or both in the test set are entirely new and have not been seen by the model during training. It should be noted that conducting cold-start settings is crucial for assessing the ability of a model to generalize to new and unseen drugs or targets, thereby reflecting real-world scenarios in drug discovery.

It is noteworthy that we evaluated two implemented versions of InceptionDTA based on the employed encoding for protein. Specifically, InceptionDTA (CharVec) utilized CharVec encoding, while InceptionDTA (Seq) utilized label encoding for protein sequence representation. In our method selection process, we aimed to encompass a broad range of approaches for predicting drug-target binding affinity. We included established baseline techniques like DeepDTA for benchmarking, alongside advanced methodologies such as DGDTA, TEFDTA, and G-K BertDTA to assess novel features and techniques in the field. Selection criteria included relevance to our study objectives and availability of implementations or published results. Specifically, DeepDTA was included because it uses simple label encoding and CNNs for both protein and drug feature extraction. Comparing our method with DeepDTA allows us to assess the effectiveness of our novel protein encoding schema and the advantages of multi-scale feature extraction as opposed to traditional CNN approaches. Similarly, TEFDTA employs fingerprints and transformers for drug feature extraction. Evaluating our method against TEFDTA helps us understand the impact of our inception modules on drug feature extraction compared to the transformer-based methods used in TEFDTA. DGDTA, which uses BiLSTM and CNN for proteins and GNNs for drugs, provides a valuable benchmark for assessing the benefits of our inception network, which handles both protein and drug features in a unified manner. This comparison highlights the advantages of using a single, integrated network versus the varied network approaches employed in DGDTA. Finally, G-K BertDTA incorporates transformers and GNNs for feature extraction. By comparing our method with G-K BertDTA, we can reveal the benefits of utilizing a lightweight inception network versus the more complex network combinations used in G-K BertDTA.

In summary, these baseline methods were chosen to provide a comprehensive comparison and to highlight the strengths of our proposed method in terms of encoding and feature extraction techniques across different network architectures. These methods are categorized based on protein and drug representation, detailed in Table 3.

3.3.1. *Comparing InceptionDTA to baselines and state-of-the-arts for warm-start settings*

InceptionDTA underwent comparison against both baseline and state-of-the-art methods across two benchmark datasets: Davis, and Kiba, as detailed in Section Methods. The comparison results, comprising the average CI, MSE, r_m^2, and AUPR over five test iterations, are presented in Tables 4, and Table 5 for the Davis, and Kiba datasets, respectively. Based on the findings from Tables 4 and 5 and it is evident that InceptionDTA surpasses all alternative methods in terms of CI, MSE, and AUPR for the Davis dataset, and achieves the second-best performance for the Kiba dataset across all performance metrics.

The superior performance on the Davis dataset compared to the Kiba dataset is likely due to the larger number of protein sequences available in the Davis dataset. This indicates that CharVec and multi-scale features extracted using the Inception network perform better when more protein sequence data is available.

This notable performance superiority can be directly attributed to the key advantages of InceptionDTA including incorporation of more detailed biological context features, and utilization of inception modules with variable filter sizes, enabling the consideration of both global and local feature distributions during the representation learning phase of the method.

To further investigate the statistical significance of the binding affinity predictions, we conducted t-tests comparing the best InceptionDTA models to the leading alternative methods based on CI scores. The p-values for each comparison are represented in the plot, with stars indicating significance levels. One star denotes 95 % significance, while additional stars indicate greater significance. An annotation of "ns" indicates that the differences between methods are not statistically significant. Fig. 3 (A, B) displays the distribution of CI scores for DGDTA and InceptionDTA (Seq). The results suggest that InceptionDTA (Seq) outperforms DGDTA for the Davis dataset. However, while DGDTA shows higher CI scores than InceptionDTA (Seq), this difference is not statistically significant.

To evaluate the accuracy of InceptionDTA in predicting binding affinity values, we compared the predicted values with the actual ones, as shown in Fig. 4(A–D). The visual comparison demonstrates that InceptionDTA provides predictions that closely align with the true binding affinities for both the Davis and Kiba datasets. Furthermore, we conducted an analysis of certain outliers to determine if the misclassified drug-target pairs share similar attributes, such as target sequence similarity or chemical similarity of the drugs. We

Table 3
Baselines and state-of-the-art methods key features.

Model	Protein	Drug
DeepDTA	Label encoding, embedding layer, and 1D CNN	Label encoding, embedding layer, and 1D CNN
TEFDTA	Label encoding, embedding layer,and 1D CNN	MACCS fingerprint, embedding layer, and Transformer
DGDTA	Bi-LSTM, and CNN	Dynamic graph attention networks
G-K BertDTA	CNN, and DenseSENet (DenseNet and squeeze-and-excitation (SE) blocks)	Graph isomorphism network (GIN), and KB-BERT
InceptionDTA (CharVec)	CharVec, Inception	Label encoding, embedding layer, and Inception
InceptionDTA(seq)	Label encoding, embedding layer, Inception	Label encoding, embedding layer, Inception

Table 4
The CI, MSE, r_m^2 and AUPR for InceptionDTA compared to the state of the art methods - Davis dataset.

Model	CI (SD)	MSE	rm^2 (SD)	AUPR(SD)
DeepDTA	0.878(0.004)	0.261	0.630 (0.017)	0.714 (0.010)
TEFDTA	0.878(0.002)	0.264	0.635(0.021)	0.702(0.015)
DGDTA	0.880(0.003)	0.273	0.636(0.012)	0.692(0.003)
G-K BertDTA	0.879(0.003)	0.267	0.660(0.003)	0.710(0.01)
InceptionDTA(CharVec)	0.891(0.003)	0.196	0.625(0.020)	0.750(0.009)
InceptionDTA(Seq)	0.897 (0.002)	0.242	0.624(0.028)	0.741(0.016)

Table 5
The CI, MSE, r_m^2 and AUPR for InceptionDTA compared to the state of the art methods - Kiba dataset.

Model	CI (SD)	MSE	rm^2(SD)	AUPR(SD)
DeepDTA	0.863 (0.002)	0.194	0.673 (0.009)	0.788 (0.004)
TEFDTA	0.857(0.002)	0.199	0.703(0.005)	0.791(0.003)
DGDTA	0.892(0.001)	0.145	0.775(0.009)	0.845(0.003)
G-K BertDTA	0.847(0.003)	0.201	0.675(0.01)	0.789(0.010)
InceptionDTA(CharVec)	0.887(0.001)	0.163	0.716(0.018)	0.818 (0.005)
InceptionDTA(Seq)	0.890(0.001)	0.166	0.707(0.007)	0.815(0.004)

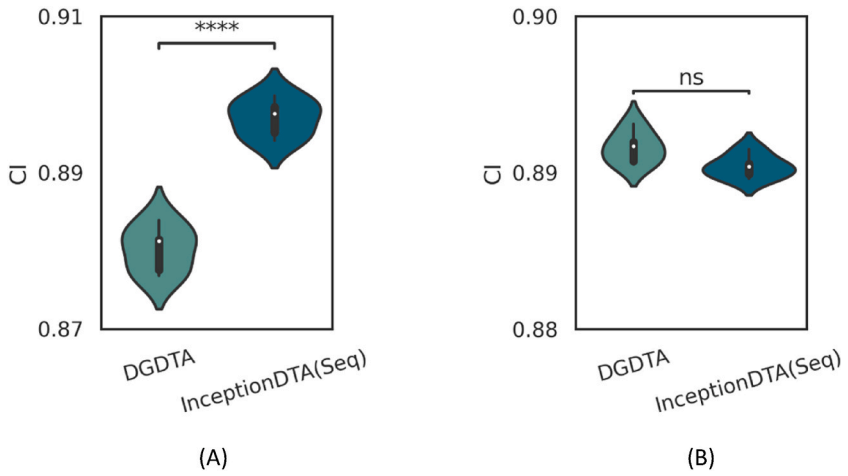

Fig. 3. Distribution of CI scores for InceptionDTA (Seq) and DGDTA across (A) the Davis dataset and (B) the Kiba dataset.

extracted the top outliers, representing approximately 0.1 % of the entire test data from the Davis dataset and, for a comprehensive evaluation, also selected the best drug-target predictions. We calculated Smith-Waterman scores for the protein sequences and Jaccard similarities for the drug SMILES in both the outlier and best prediction sets. Two key observations emerged. For the outliers, the Smith-Waterman scores were higher, and the Jaccard similarities were relatively lower. Conversely, for the best predictions, the Smith-Waterman scores were generally lower, and the Jaccard similarities were relatively higher. This pattern suggests that the outliers have higher sequence similarity but lower structural similarity, whereas the best predictions exhibit lower sequence similarity but higher structural similarity, likely contributing to better model performance on these pairs.

3.3.2. Comparing InceptionDTA to baselines and state-of-the-arts for refined data splitting settings

We evaluated two refined datasets, comprising refined versions of the PDBbind and Davis datasets, where redundancies and biases were mitigated by excluding multiple annotations for similar proteins and drugs. For PDBbind, after removing potential redundancies in drugs with multiple sequences in the SMILES format, the models were trained and evaluated on this refined, smaller dataset, which no longer contained these redundant drug sequences. Similarly, to create the refined Davis dataset, we excluded approximately 15 % of the protein sequences identified as redundant due to unique sequences having multiple protein identifiers. The models were then trained and evaluated on this refined dataset, which no longer included these redundant sequences. Consequently, these refined datasets offer higher-quality binding affinity data.

The comparison results, encompassing the average CI, MSE, r_m^2, and AUPR, are presented in Table 6 for the refined PDBbind dataset and Table 7 for the refined Davis dataset.

Based on the results presented in Table 6, InceptionDTA demonstrates superior performance compared to alternative methods in

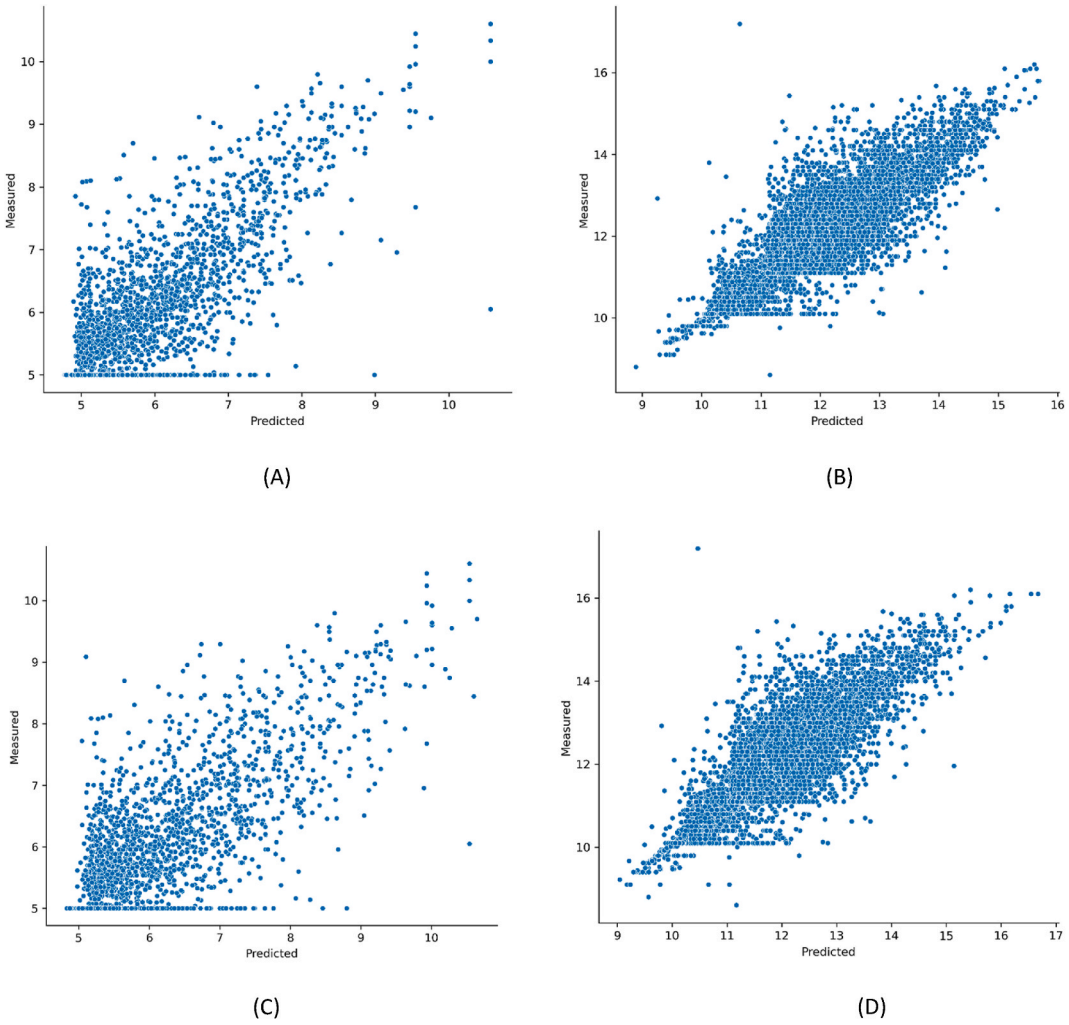

Fig. 4. Predicted versus actual binding affinities: (A) InceptionDTA (CharVec) on the Davis dataset, (B) InceptionDTA (CharVec) on the Kiba dataset, (C) InceptionDTA (Seq) on the Davis dataset, and (D) InceptionDTA (Seq) on the Kiba dataset.

terms of CI, r_m^2 , and AUPR for the refined PDBbind dataset, though the differences are marginal. Similarly, for the refined Davis dataset, InceptionDTA shows the best predictive performance in terms of CI, MSE, and AUPR.

According to Tables 4 and 7, InceptionDTA (Seq) achieves a slightly higher CI on the original Davis dataset compared to the refined dataset. A similar trend is observed with InceptionDTA (CharVec) and other metrics, including MSE and AUPR, where performance is marginally better on the original dataset. These differences likely arise from the smaller size of the refined dataset due to the exclusion of redundant protein sequences. Although the refined dataset offers higher-quality data, the reduced sequence count may limit the ability of the model to generalize, making it more challenging to achieve superior performance. Specifically, the decline in model performance on the refined Davis dataset, may be attributed to several factors. First, the reduction of nearly 15 % of redundant protein sequences has resulted in a smaller dataset compared to the original Davis. This decrease in the number of protein sequences may hinder the ability of the model to learn effective representations. Additionally, as the number of protein sequences decreases, so does

Table 6
The CI, MSE, r_m^2 and AUPR for InceptionDTA compared to the state of the art methods - PDBbind dataset.

Model	CI (SD)	MSE	rm^2(SD)	AUPR(SD)
DeepDTA	0.756(0.005)	2.140	0.471(0.025)	0.777(0.017)
TEFDTA	0.750(0.003)	1.936	0.404(0.02)	0.781(0.006)
DGDTA	0.752(0.005)	1.962	0.405(0.011)	0.750(0.016)
G-K BertDTA	0.743(0.004)	2.032	0.405(0.013)	0.740(0.012)
InceptionDTA(CharVec)	0.752(0.013)	2.090	0.434(0.049)	0.773(0.012)
InceptionDTA(Seq)	0.758(0.004)	2.412	0.475(0.034)	0.784(0.008)

Table 7
The CI, MSE, r_m^2 and AUPR for InceptionDTA compared to the state of the art methods – refined Davis dataset.

Model	CI (SD)	MSE	rm^2(SD)	AUPR(SD)
DeepDTA	0.869(0.004)	0.259	0.616(0.034)	0.695(0.027)
TEFDTA	0.866(0.005)	0.253	0.612(0.021)	0.668(0.022)
DGDTA	0.865(0.003)	0.356	0.616(0.014)	0.671(0.021)
G-K BertDTA	0.872(0.003)	0.240	0.629(0.008)	0.695(0.006)
InceptionDTA(CharVec)	0.886(0.005)	0.219	0.594(0.024)	0.698(0.020)
InceptionDTA(Seq)	0.889(0.003)	0.227	0.587(0.037)	0.703(0.024)

the number of drug-target pairs and interactions, which could further restrict the generalization capability of the model. Consequently, while the refined Davis dataset features high-quality data, these factors may limit the overall predictive performance of the model. Nonetheless, the solid results on the refined dataset underscore the robustness of InceptionDTA across datasets of varying sizes and data quality.

To further assess the statistical significance of the binding affinity predictions, we conducted t-tests comparing the best InceptionDTA models with the leading alternative methods based on CI scores. Fig. 5 (A, B) illustrates the distribution of CI scores for these alternative methods and InceptionDTA (Seq). The results indicate that InceptionDTA (Seq) outperforms both DeepDTA and G-K BertDTA for the PDBbind and refined Davis datasets.

To evaluate the accuracy of InceptionDTA in predicting binding affinity values for the refined datasets, we compared the predicted values with the actual ones, as shown in Fig. 6(A–D). The visual comparison demonstrates that InceptionDTA provides predictions that closely align with the true binding affinities for both the PDBbind and refined Davis datasets.

3.3.3. Comparing InceptionDTA to baselines and state-of-the-arts for cold-start settings

To evaluate the predictive performance of InceptionDTA for unseen drugs and proteins, we implemented various challenging cold-start data splitting configurations, encompassing cold-drug, cold-protein, and cold protein-drug settings. The comparison results, including the average CI and MSE, are visualized in Fig. 7, illustrating the outcomes across three cold-start splitting scenarios on the Davis dataset. As shown in Fig. 7, InceptionDTA (Seq) achieves the highest CI in the cold-protein setting, indicating superior predictive performance for unseen proteins. InceptionDTA (CharVec) also performs well, surpassing all other methods. In the cold-drug setting, G-K BertDTA and DGDTA outperform InceptionDTA, suggesting that advanced graph neural networks (GNNs), such as graph isomorphism networks (GIN) and dynamic graph attention networks, provide more effective drug representations.

In the cold-protein-drug setting, InceptionDTA (CharVec) and InceptionDTA (Seq) achieve the top two CI scores, comparable to G-K BertDTA, demonstrating strong performance among all methods. According to MSE results in Fig. 7, InceptionDTA (Seq) achieves the lowest MSE in the cold-protein setting, indicating the most accurate predictions. Although InceptionDTA (CharVec) has a higher MSE, it still outperforms several other models, including G-K BertDTA. In the cold-drug setting, InceptionDTA (CharVec) achieves the lowest MSE, highlighting its effectiveness in predicting affinities for unseen drugs, with G-K BertDTA being the next best performer. For the cold-protein-drug setting, InceptionDTA (CharVec) achieves the lowest MSE, followed by InceptionDTA (Seq), showcasing its strong performance in this challenging scenario.

Overall, InceptionDTA consistently demonstrates strong performance across diverse cold-start settings. The Seq variant excels in the cold-protein setting for both CI and MSE, while the CharVec variant shows superior performance in MSE for the cold-drug and cold-protein-drug settings. This versatility makes InceptionDTA a robust choice for drug-target affinity prediction tasks, with its different

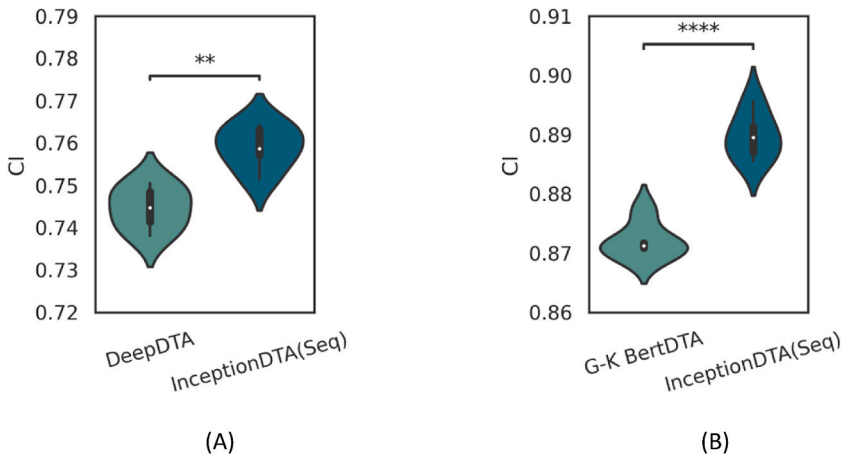

Fig. 5. Distribution of CI scores for (A) InceptionDTA (Seq) and DeepDTA across the PDBbind dataset, and (B) InceptionDTA (Seq) and G-K BertDTA across the refined Davis dataset.

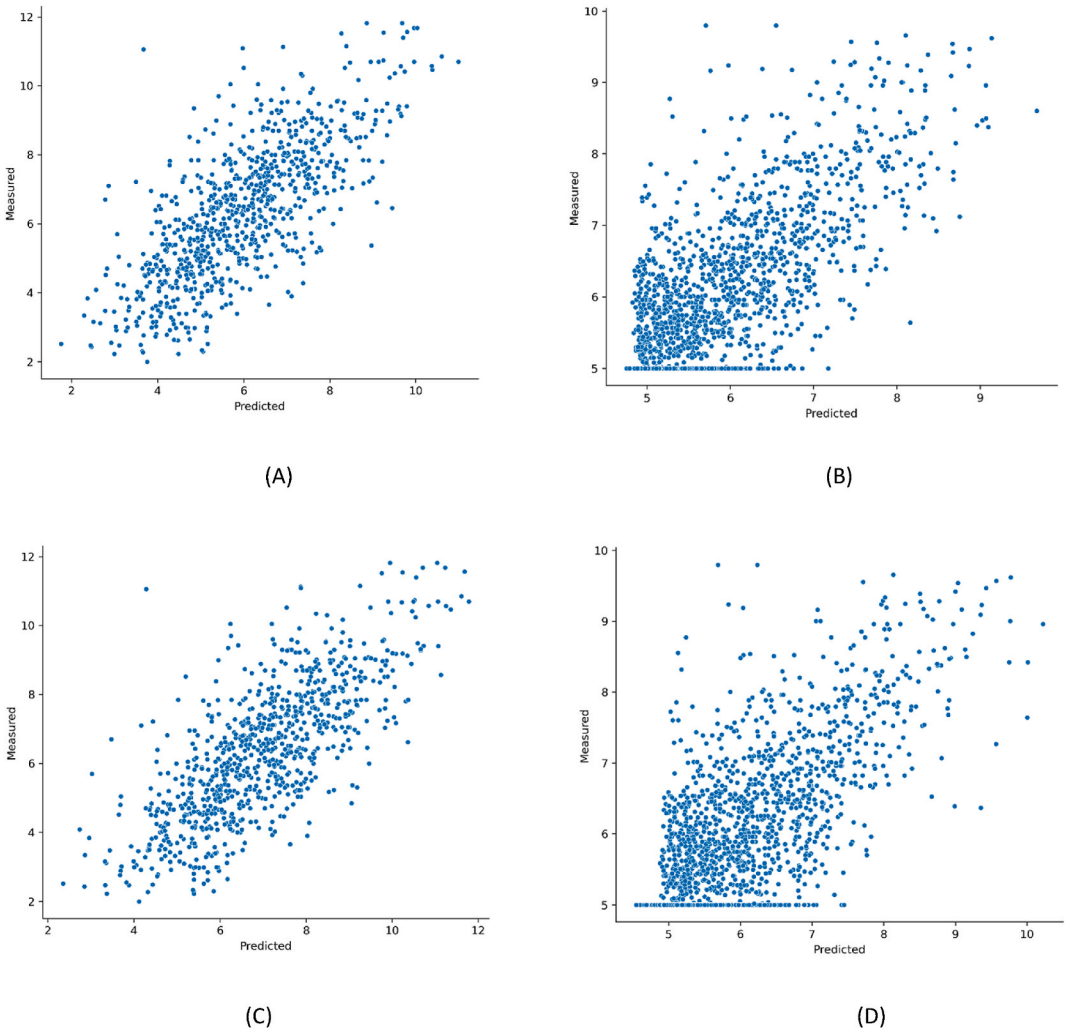

Fig. 6. Predicted versus actual binding affinities: (A) InceptionDTA (CharVec) on the PDBbind dataset, (B) InceptionDTA (CharVec) on the refined Davis dataset, (C) InceptionDTA (Seq) on the PDBbind dataset, and (D) InceptionDTA (Seq) on the refined Davis dataset.

variants complementing each other to handle various predictive challenges effectively.

Moreover, we implemented two rigorous cold-start data splitting strategies based on the physicochemical properties of drug molecules. Specifically, we excluded drug SMILES during model training based on the logP, which characterizes the lipophilicity of corresponding drugs. These excluded drugs were subsequently utilized as test data for evaluation purposes. For this assessment, we employed the PDBbind dataset, which offers essential physicochemical information for the drugs, including logP values calculated using both Open Babel logP [44] and XLOGP3 [45] tools. The comparison results, comprising the CI and MSE, are illustrated in Fig. 8, delineating the outcomes for Open Babel logP and XLOGP3 splitting strategies on the PDBbind dataset.

As observed in Fig. 8, InceptionDTA showcased superior performance compared to alternative methods in terms of CI for both data splitting settings. Additionally, it exhibited the best MSE for the XPLOG3 splitting strategy and the second-best MSE for the Open Babel logP setting.

3.3.4. Control experiments for InceptionDTA performance

To further validate InceptionDTA, we conducted control experiments to demonstrate that overfitting is not occurring. Specifically, we applied a data-scrambling technique, where the binding affinity values for the benchmark datasets were randomized. The models, referred to as Scrambled InceptionDTA (CharVec) and Scrambled InceptionDTA (Seq), were then trained and tested on this scrambled data. As shown in Table 8, the CI values for both scrambled models remained around 0.5, indicating predictions no better than random, while MSE values increased significantly. These results highlight a complete loss of predictive performance, confirming that the InceptionDTA models are not overfitting but are instead learning meaningful patterns from the data.

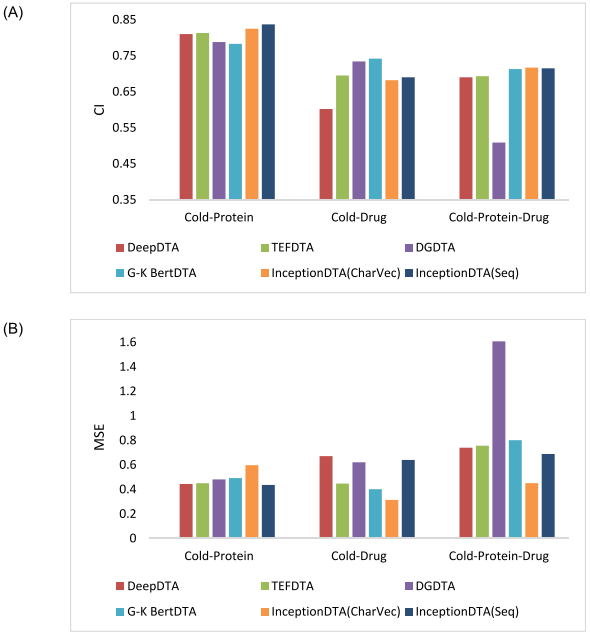

Fig. 7. performance comparisons InceptionDTA to the baselines and state of the art methods –Davis dataset for Cold-start settings. (A) CI, and (B) MSE.

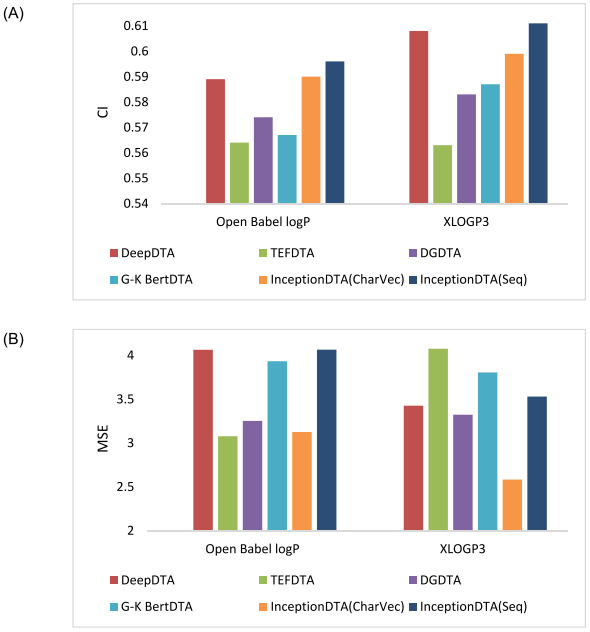

Fig. 8. performance comparisons InceptionDTA to the baselines and state of the art methods –PDBbind dataset for Cold-physio-chemical properties splitting settings. (A) CI, and (B) MSE.

3.3.5. Comparing InceptionDTA to baselines and state-of-the-arts for time complexity

The time complexity of InceptionDTA, in terms of runtime for training and testing, is compared to several alternative methods. Table 9 provides runtime data for both versions of InceptionDTA and four existing methods across different datasets. According to Table 9, DeepDTA demonstrates relatively low runtime, reflecting its use of simple label encoding combined with 1D CNNs for both proteins and drugs. This simplicity contributes to its faster training and testing times. TEFDTA shows significantly higher runtimes, particularly in training, due to its complex architecture involving MACCS fingerprints and Transformers for drug representations, which require extensive computational resources. DGDTA also exhibits higher runtime compared to DeepDTA, attributable to its use of

Table 8
Control experiments for InceptionDTA performance.

Model	Davis		Kiba		PDBbind		Refined Davis	
	CI(SD)	MSE	CI(SD)	MSE	CI(SD)	MSE	CI(SD)	MSE
Scrambled InceptionDTA(CharVec)	0.497(0.004)	0.836	0.504(0.002)	0.744	0.517(0.009)	5.240	0.504(0.007)	0.763
Scrambled InceptionDTA(Seq)	0.497(0.001)	0.825	0.508(0.001)	0.723	0.518(0.010)	6.223	0.511(0.003)	0.808
InceptionDTA(CharVec)	0.891(0.003)	0.196	0.887(0.001)	0.163	0.752(0.013)	2.090	0.886(0.005)	0.219
InceptionDTA(Seq)	0.897(0.002)	0.242	0.890(0.001)	0.166	0.758(0.004)	2.412	0.889(0.003)	0.227

Table 9
Runtime for training and testing of InceptionDTA and alternative methods.

Model	Davis (seconds/epoch)		Kiba (seconds/epoch)		PDBbind (seconds/epoch)		Refined Davis (seconds/epoch)	
	Train	Test	Train	Test	Train	Test	Train	Test
DeepDTA	5	0.75	15	2.3	1.4	0.56	4	0.71
TEFDTA	179	39.17	700	179.9	22.8	4.61	148	31.28
DGDTA	9	1.63	37	6.29	43.1	0.25	7.65	1.41
G-K BertDTA	36	4.61	134	18.41	7	0.79	31	4.18
InceptionDTA (CharVec)	59.7	12.49	242.8	48.88	10.7	2.8	53.78	11.51
InceptionDTA (Seq)	39.5	3.67	155.5	11.63	7.5	0.94	33.5	3.28

Bi-LSTM and CNNs for proteins and dynamic graph attention networks for drugs. This advanced graph-based approach impacts its training and testing times. G-K BertDTA has moderate runtimes, with notable training and testing times. Its use of graph isomorphism networks and KB-BERT for drugs contributes to its computational requirements. InceptionDTA (CharVec) has higher runtimes compared to methods like DeepDTA and DGDTA. This is due to the CharVec encoding for proteins and the use of Inception networks, which introduce additional complexity in feature extraction. It should be noted that since feature vectors derived from Skipgram are used only once to create CharVec and do not need to be re-extracted during training and testing, their impact on the overall runtime is negligible. InceptionDTA (Seq), utilizing simple label encoding and Inception networks, exhibits intermediate runtimes between the other methods. Its use of straightforward encoding schemes combined with Inception networks results in efficient performance.

Overall, these differences in runtime underscore the trade-offs between the complexity of the model architectures and their computational requirements.

3.3.6. A comparison of CharVec and baseline encoding schemes

The prediction performance of CharVec is compared to alternative encoding methods, specifically label encoding and one-hot encoding. We conducted experiments using InceptionDTA with one-hot encoding, referred to as InceptionDTA (One-hot), to evaluate the performance of CharVec. For a more comprehensive assessment, we included InceptionDTA (Seq), which utilizes label encoding.

Table 10 presents the CI and MSE for the models on the Davis dataset. According to the results, InceptionDTA (CharVec) demonstrates better MSE compared to InceptionDTA (One-hot) and InceptionDTA (Seq), while InceptionDTA (Seq) shows superior CI over both InceptionDTA (One-hot) and InceptionDTA (CharVec). Thus, the performance of CharVec and label encoding yields better prediction outcomes in terms of MSE and CI, respectively, when compared to one-hot encoding. It should be noted that, due to the equal length of CharVec and one-hot encoding, the runtime for models using these encodings is comparable. However, as shown in Table 9, the runtime for label encoding is shorter.

4. Discussion

In this section, we delve into a comprehensive analysis of our proposed method and compare its neural network architecture against popular methods for drug-target binding affinity prediction. Our aim is to elucidate the underlying benefits of our approach in terms of architectural design and accuracy enhancement for the prediction task.

As previously discussed, InceptionDTA leverages distributed representation vectors generated using a Skipgram neural network trained on available protein sequences. These vectors enable the direct encoding of each protein sequence via a novel encoding scheme dubbed CharVec. Subsequently, these encoded protein sequences, along with label-encoded drug sequences, are fed into another deep neural network for precise representation learning and binding affinity prediction. Notably, unlike prior approaches, the architecture of this latter model is constructed upon inception modules for both protein and drug sequences.

Our platform offers significant cost advantages over traditional wet lab techniques. By minimizing the need for expensive materials and lengthy experiments, our computational approach delivers faster and more scalable predictions at a lower cost. This efficiency not only accelerates the drug discovery process but also supports high-throughput screening, making it a more economical and effective solution for predicting drug-target binding affinity.

A quantitative comparison of InceptionDTA against well-known architectures for drug-target binding affinity prediction is presented in Section Results. Across three categories of baseline methods (as discussed in Sections Introduction and Results), our

Table 10
performance comparisons CharVec to the baselines encoding methods –Davis dataset.

Model	CI(SD)	MSE
InceptionDTA(One-hot)	0.884(0.004)	0.269
InceptionDTA(CharVec)	0.891(0.003)	0.196
InceptionDTA(Seq)	0.897(0.002)	0.242

comparative study of InceptionDTA yields the following achievements:

In comparison to the first category of DTA methods (i.e., sequence-based), InceptionDTA takes advantage of more data for the representation learning of protein sequences, resulting in more accurate input vectors for feature extraction and binding affinity prediction. Additionally, by utilizing inception modules with multiple filter sizes, InceptionDTA offers more comprehensive and detailed representation learning of protein sequences. Our method surpasses the sequence-based method, DeepDTA, across four datasets in terms of accuracy metrics for warm-start, refined, and cold-start settings. This indicates that incorporating biological context features captured by CharVec and multi-scale features extracted with inception modules contributes to performance improvement. However, the use of an additional model for word embedding may introduce a drawback in terms of time complexity compared to other sequence-based methods.

Similar to the second category of DTA methods (i.e., transformer-based), InceptionDTA employs a semi-supervised approach, leveraging more data for accurate distributed representation of proteins. For instance, TEFDTA utilizes a transformer network for drugs to improve representation and, consequently, offers accurate binding affinity prediction. In comparison to TEFDTA, our method demonstrates efficient performance on benchmark datasets, indicating that InceptionDTA capitalizes more on additional data for representation learning and biological context feature extraction compared to well-known transformer-based methods.

In contrast to graph neural network-based methods for DTA prediction, such as DGDTA and G-K BertDTA, InceptionDTA adopts a sequence-based inception architecture for representation learning of both drugs and proteins. Through comparison studies, our method showcases more accurate distributed representation vectors for protein sequences and drug SMILES compared to DGDTA, which utilizes dynamic graph attention networks for drugs and a combination of Bi-LSTM and CNNs for proteins. Similarly, compared to G-K BertDTA, which employs graph isomorphism networks for drugs and CNNs alongside DenseSENet for protein representation learning, InceptionDTA demonstrates superior performance across most comparison scenarios. For more challenging cold-start settings, our method outperforms graph-based methods in cold-protein and cold-protein-drug scenarios but does not achieve superior performance in the cold-drug setting. This may be due to the advanced capabilities of graph neural networks in extracting more critical features from drugs in cold-drug scenarios.

To further evaluate the performance of our method, we examined both outliers and top predictions. Our analysis reveals a clear distinction between these two sets. Outliers typically exhibit higher sequence similarity but lower structural similarity for proteins and drugs, respectively, while the best predictions show the opposite pattern with lower sequence similarity and higher structural similarity. This contrast suggests that the model performs more effectively on pairs with lower sequence similarity but higher structural similarity. Conversely, it struggles with pairs that have high sequence similarity but low structural similarity, which aligns with the observed performance issues.

In comparing InceptionDTA (CharVec) and InceptionDTA (Seq), where CharVec encoding and simple label encoding are used respectively, InceptionDTA (Seq) generally outperforms InceptionDTA (CharVec) in terms of CI, indicating its superior ability to rank and predict relative binding affinities. Conversely, InceptionDTA (CharVec) usually achieves lower MSE values, suggesting it may provide more accurate absolute predictions. The differences in performance metrics between InceptionDTA (Seq) and InceptionDTA (CharVec) can be attributed to how each encoding scheme captures information and how the models interpret that information. InceptionDTA (Seq) focuses on sequential patterns, enhancing its ability to rank predictions effectively, which contributes to a higher CI. In contrast, InceptionDTA (CharVec) captures more detailed biological context at the character level, resulting in lower MSE due to improved absolute predictions. However, this complexity may hinder its ranking ability. These trade-offs illustrate how each method prioritizes different aspects of model performance, reflecting their strengths and limitations. Therefore, InceptionDTA (Seq) is preferable for applications where CI is the primary performance metric, such as identifying promising drug candidates or repurposing existing drugs based on their relative efficacy, while InceptionDTA (CharVec) may be better suited for scenarios where minimizing MSE is critical. This distinction in performance highlights the flexibility of InceptionDTA, allowing it to address a range of predictive needs effectively.

In summary, our method outperforms several baseline and state-of-the-art sequence-based, transformer-based, and graph-based methods for drug-target binding affinity prediction. Furthermore, given its simple architecture, InceptionDTA can be extended by incorporating additional networks, particularly graph neural networks, to further enhance prediction accuracy.

5. Conclusions

Drug-target binding affinity prediction plays a pivotal role in drug discovery and development. In this paper, we introduced a novel encoding scheme for protein sequences, named CharVec, and a deep learning-based method for drug-target binding affinity prediction, named InceptionDTA. Our approach capitalizes on a pre-trained neural network, Skipgram, to accurately represent targets, complemented by widely used label encoding for drug sequences. Specifically, we leverage character-level modeling (i.e., amino acids) to incorporate biological context features, resulting in more efficient amino acid distributed representation and protein sequence

encoding. Moreover, we employ an inception-based architecture for feature extraction and representation learning from the encoded sequences, culminating in binding affinity prediction. Through comparative studies, InceptionDTA demonstrates superior performance over various deep learning-based methods. These findings underscore InceptionDTA as a high-performance method for this task, offering both prediction accuracy and architectural simplicity.

Looking ahead, as part of future work, we plan to explore different lengths of embedded vectors for amino acid sequence encoding. Indeed, excluding dimensionality reduction of Prot2Vec extracted from Skipgram to CharVec may enhance prediction performance by avoiding information loss. However, further dimensionality reduction, while potentially affecting prediction accuracy, can improve the runtime efficiency of the model. Additionally, we aim to investigate the direct passing of encoded vectors into the fully connected network, bypassing the inception layers, to enhance model training efficiency. To achieve this, we can remove inception modules from the protein or drug feature extraction step to create lightweight models. In addition to examining the impact of inception modules and the proposed encoding, this approach offers a trade-off between prediction performance and runtime efficiency. Furthermore, while our current study focuses on CharVec adoption for drug-target binding affinity prediction, ongoing research involves exploring the sequence encoding capabilities of CharVec for other applications, such as protein classification and protein-protein interaction prediction. Moreover, InceptionDTA is designed to predict binding affinity and interactions between drugs and proteins as a regression and classification task, without focusing on specific binding sites or factors like stoichiometry. Predicting multi-site binding or synergistic effects involves a different problem formulation, often requiring techniques such as attention mechanisms. In future work, we aim to explore the integration of attention mechanisms within the InceptionDTA framework to enhance its capacity for modeling complex interactions at multiple binding sites, as well as to improve model interpretability. This integration could enable the model to focus on relevant regions of the protein or drug sequences, potentially boosting prediction accuracy. However, challenges may arise in effectively adapting the architecture to incorporate these mechanisms, such as increased computational complexity and the necessity for additional training data to capture multi-site interactions. A future work could explore incorporating these methods to address such aspects.

CRediT authorship contribution statement

Mahmood Kalemati: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Mojtaba Zamani Emani:** Software, Formal analysis, Data curation. **Somayyeh Koohi:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data and code availability statement

The datasets and code generated and/or used during this study are available in the GitHub repository: https://github.com/mz76m/InceptionDTA.

Funding

Not applicable.

Declaration of competing 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.

Acknowledgements

Not applicable.

References

[1] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, Prediction of drug–target interaction networks from the integration of chemical and genomic spaces, Bioinformatics 24 (13) (2008 Jul 1) i232–i240.
[2] K. Bleakley, Y. Yamanishi, Supervised prediction of drug–target interactions using bipartite local models, Bioinformatics 25 (18) (2009 Sep 15) 2397–2403.
[3] T. Van Laarhoven, S.B. Nabuurs, E. Marchiori, Gaussian interaction profile kernels for predicting drug–target interaction, Bioinformatics 27 (21) (2011 Nov 1) 3036–3043.

[4] D.S. Cao, L.X. Zhang, G.S. Tan, Z. Xiang, W.B. Zeng, Q.S. Xu, A.F. Chen, Computational prediction of drug-target interactions using chemical, biological, and network features, Molecular informatics 33 (10) (2014 Oct) 669–681.
[5] T. Pahikkala, A. Airola, S. Pietilä, S. Shakyawar, A. Szwajda, J. Tang, T. Aittokallio, Toward more realistic drug–target interaction predictions, Briefings Bioinf. 16 (2) (2015 Mar 1) 325–337.
[6] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, Nat. Rev. Drug Discov. 3 (11) (2004 Nov) 935–949.
[7] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2) (2010 Jan 30) 455–461.
[8] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, M. Ester, SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines, J. Cheminf. 9 (1) (2017 Dec) 1–4.
[9] T.F. Smith, M.S. Waterman, Identification of common molecular subsequences, Journal of molecular biology 147 (1) (1981 Mar 25) 195–197.
[10] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 13 (2016 Aug) 785–794.
[11] E.E. Bolton, Y. Wang, P.A. Thiessen, S.H. Bryant, PubChem: integrated platform of small molecules and biological activities, Annu. Rep. Comput. Chem. 4 (2008 Jan 1) 217–241. Elsevier.
[12] H. Öztürk, A. Özgür, E. Ozkirimli, DeepDTA: deep drug–target binding affinity prediction, Bioinformatics 34 (17) (2018 Sep 1) i821–i829.
[13] H. Öztürk, E. Ozkirimli, A. Özgür, WideDTA: prediction of drug-target binding affinity, arXiv preprint arXiv: 1902.04166 (2019 Feb 4).
[14] L. Zhao, J. Wang, L. Pang, Y. Liu, J. Zhang, GANsDTA: predicting drug-target binding affinity using GANs, Front. Genet. 10 (2020 Jan 9) 1243.
[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020 Oct 22) 139–144.
[16] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, Bert: pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv: 1810.04805 (2018 Oct 11).
[17] B. Shin, S. Park, K. Kang, J.C. Ho, Self-attention based molecule representation for predicting drug-target interaction, Machine Learning for Healthcare Conference 28 (2019 Oct) 230–248. PMLR.
[18] Z. Li, P. Ren, H. Yang, J. Zheng, F. Bai, TEFDTA: a transformer encoder and fingerprint representation combined prediction method for bonded and non-bonded drug–target affinities, Bioinformatics 40 (1) (2024 Jan 1) btad778.
[19] X. Qiu, H. Wang, X. Tan, Z. Fang, GK BertDTA: a graph representation learning and semantic embedding-based framework for drug-target affinity prediction, Comput. Biol. Med. 173 (2024 May 1) 108376.
[20] T. Nguyen, H. Le, T.P. Quinn, T. Nguyen, T.D. Le, S. Venkatesh, GraphDTA: predicting drug–target binding affinity with graph neural networks, Bioinformatics 37 (8) (2021 Apr 15) 1140–1147.
[21] M. Jiang, Z. Li, S. Zhang, S. Wang, X. Wang, Q. Yuan, Z. Wei, Drug–target affinity prediction using graph neural network and contact maps, RSC Adv. 10 (35) (2020) 20701–20712.
[22] M. Abdel-Basset, H. Hawash, M. Elhoseny, R.K. Chakrabortty, M. Ryan, DeepH-DTA: deep learning for predicting drug-target interactions: a case study of COVID-19 drug repurposing, IEEE Access 8 (2020 Sep 15) 170433–170451.
[23] H. Zhai, H. Hou, J. Luo, X. Liu, Z. Wu, J. Wang, DGDTA: dynamic graph attention network for predicting drug–target binding affinity, BMC Bioinf. 24 (1) (2023 Sep 30) 367.
[24] L. Zhang, C.C. Wang, Y. Zhang, X. Chen, GPCNDTA: prediction of drug-target binding affinity through cross-attention networks augmented with graph features and pharmacophores, Comput. Biol. Med. 166 (2023 Nov 1) 107512.
[25] E. Asgari, M.R. Mofrad, Continuous distributed representation of biological sequences for deep proteomics and genomics, PLoS One 10 (11) (2015 Nov 10) e0141287.
[26] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013 Jan 16).
[27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[28] A.S. Rifaioglu, R. Cetin Atalay, D. Cansen Kahraman, T. Doğan, M. Martin, V. Atalay, MDeePred: novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery, Bioinformatics 37 (5) (2021 Mar 1) 693–704.
[29] I. Lee, J. Keum, H. Nam, DeepConv-DTI: prediction of drug-target interactions via deep learning with convolution on protein sequences, PLoS Comput. Biol. 15 (6) (2019 Jun 14) e1007129.
[30] F. Rayhan, S. Ahmed, Z. Mousavian, D.M. Farid, S. Shatabda, FRnet-DTI: deep convolutional neural network for drug-target interaction prediction, Heliyon 6 (3) (2020 Mar 1) e00938.
[31] M.I. Davis, J.P. Hunt, S. Herrgard, P. Ciceri, L.M. Wodicka, G. Pallares, M. Hocker, D.K. Treiber, P.P. Zarrinkar, Comprehensive analysis of kinase inhibitor selectivity, Nat. Biotechnol. 29 (11) (2011 Nov) 1046–1051.
[32] J. Tang, A. Szwajda, S. Shakyawar, T. Xu, P. Hintsanen, K. Wennerberg, T. Aittokallio, Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis, J. Chem. Inf. Model. 54 (3) (2014 Mar 24) 735–743.
[33] R. Wang, X. Fang, Y. Lu, S. Wang, The PDBbind database: collection of binding affinities for protein− ligand complexes with known three-dimensional structures, J. Med. Chem. 47 (12) (2004 Jun 3) 2977–2980.
[34] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1) (1988 Feb 1) 31–36.
[35] E. Asgari, N. Poerner, A.C. McHardy, M.R. Mofrad, DeepPrime2Sec: deep learning for protein secondary structure prediction from the primary sequences, bioRxiv (2019 Jan 1) 705426.
[36] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines. Proceedings of the 27th international Conference on Machine Learning (ICML-10), 2010 Jan 1, pp. 807–814.
[37] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014 Dec 22).
[38] M. Gönen, G. Heller, Concordance probability and discriminatory power in proportional hazards regression, Biometrika 92 (4) (2005 Dec 1) 965–970.
[39] K. Roy, P. Chakraborty, I. Mitra, P.K. Ojha, S. Kar, R.N. Das, Some case studies on application of "rm2" metrics for judging quality of quantitative structure–activity relationship predictions: emphasis on scaling of response data, J. Comput. Chem. 34 (12) (2013 May 5) 1071–1082.
[40] P. Pratim Roy, S. Paul, I. Mitra, K. Roy, On two novel parameters for validation of predictive QSAR models, Molecules 14 (5) (2009 May) 1660–1701.
[41] F. Chollet, et al., Keras. https://github.com/fchollet/keras. (Accessed 15 September 2021).
[42] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, Tensorflow: a system for large-scale machine learning. In12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), 2016, pp. 265–283.
[43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research 15 (1) (2014 Jan 1) 1929–1958.
[44] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: an open chemical toolbox, J. Cheminf. 3 (1) (2011 Dec) 1–4.
[45] T. Cheng, Y. Zhao, X. Li, F. Lin, Y. Xu, X. Zhang, Y. Li, R. Wang, L. Lai, Computation of octanol− water partition coefficients by guiding an additive model with knowledge, J. Chem. Inf. Model. 47 (6) (2007 Nov 26) 2140–2148.