Article

NFSA-DTI: A Novel Drug–Target Interaction Prediction Model Using Neural Fingerprint and Self-Attention Mechanism

Feiyang Liu, Huang Xu, Peng Cui * , Shuo Li, Hongbo Wang and Ziye Wu *

School of Information, Guizhou University of Finance and Economics, Guiyang 550025, China;
feiyang@mail.gufe.edu.cn (F.L.)
* Correspondence: pcui@mail.gufe.edu.cn (P.C.); zywu@mail.gufe.edu.cn (Z.W.)

Abstract: Existing deep learning methods have shown outstanding performance in predicting drug–target interactions. However, they still have limitations: (1) the over-reliance on locally extracted features by some single encoders, with insufficient consideration of global features, and (2) the inadequate modeling and learning of local crucial interaction sites in drug–target interaction pairs. In this study, we propose a novel drug–target interaction prediction model called the Neural Fingerprint and Self-Attention Mechanism (NFSA-DTI), which effectively integrates the local information of drug molecules and target sequences with their respective global features. The neural fingerprint method is used in this model to extract global features of drug molecules, while the self-attention mechanism is utilized to enhance CNN's capability in capturing the long-distance dependencies between the subsequences in the target amino acid sequence. In the feature fusion module, we improve the bilinear attention network by incorporating attention pooling, which enhances the model's ability to learn local crucial interaction sites in the drug–target pair. The experimental results on three benchmark datasets demonstrated that NFSA-DTI outperformed all baseline models in predictive performance. Furthermore, case studies illustrated that our model could provide valuable insights for drug discovery. Moreover, our model offers molecular-level interpretations.

Keywords: drug–target interaction prediction; drug discovery; graph neural network; neural fingerprint; self-attention; attention pooling; interpretability

1. Introduction

Despite significant advances in basic life sciences and biotechnology, the drug discovery and development process is still limited by slow time and high costs [1]. The average duration for creating a small-molecule medication is approximately 15 years, costing around USD 2 billion [2]. While clinical studies are widely acknowledged as a crucial component of drug development, the greatest potential for time and cost savings lies in the earlier discovery stage [3,4]. The prediction of drug–target interaction (DTI) plays a pivotal role in guiding research and development efforts, making it an essential part of the drug discovery process [5,6].

Recently, significant advancements in drug–target interaction prediction have been achieved due to rapid progress in deep learning (DL)-based approaches [7–10]. Many of these approaches seamlessly integrate the drug chemical space, protein sequence, and interaction information into a comprehensive end-to-end framework [11,12]. They approach DTI prediction as a binary classification task and utilize various deep encoding and decoding modules, such as graph neural networks [13–15], deep neural networks [16,17], or transformer architectures [18,19], to make predictions. With the advancements in deep learning techniques, these models can automatically learn data-driven representations of drugs and proteins from extensive DTI data instead of solely relying on predefined descriptors.

However, the current DL-based models, despite exhibiting decent performance, are subject to two limitations. One limitation arises from the predominant focus on local feature

extraction, neglecting the acquisition of global representation [13,14,16]. The prediction of drug–target interaction necessitates comprehensive consideration of various aspects pertaining to receptor and ligand information [20]. Therefore, focusing only on local features tends to limit the modeling power and forecasting performance of the model. The second limitation arises from the inability to explicitly model and learn the local key interaction sites between the drug molecule and target protein. This limitation impedes the ability to provide a more intuitive and molecular-level interpretability of the mechanisms underlying drug–target interaction.

In recent years, the combination of molecular fingerprinting and deep learning models has further improved the performance of models in dealing with complex molecular structures [21]. For example, Zhu et al. [17] proposed a fingerprint-embedding framework for drug–target binding affinity prediction (FingerDTA), which uses convolutional neural networks (CNNs) to extract local patterns and utilizes fixed fingerprints to characterize global information. Lee et al. [16] proposed a novel model called DeepConv-DTI, which captures the local residue patterns by convolving amino acid subsequences of various lengths, and it uses a fully connected neural network to encode the fixed ECFP4 [22] drug fingerprint. On the one hand, these traditional fingerprints are usually generated based on the local structure of the molecule (such as specific atoms or chemical substructures) and rely on predefined rules, so they cannot effectively capture the overall topology of the molecule [23]. In contrast, the neural fingerprint [24] based on a graph neural network can gradually integrate local information into global information by recursively aggregating the drug features of all nodes. On the other hand, the limitation of CNN as the protein encoder is that the dependencies between distant amino acids in the protein sequence cannot be captured well [25]. Fortunately, the self-attention mechanism can assign different weights to each position in the sequence, allowing it to capture dependencies between positions more effectively [26]. This offers the potential to enhance CNN's capability in capturing the global information within the sequence.

Building on the advantages of neural fingerprints, Duvenaud et al. [24] integrated them into graph neural networks (GNNs) to propose the neural fingerprint graph neural network (NFGNN). Originally designed for predicting molecular properties, NFGNN has been shown through comparative experiments by Feldmann et al. [27] to enhance the generalization ability of molecular property predictions, demonstrating robust performance even for molecules that differ from the training set. Recently, the application of NFGNN in DTI prediction has begun to make progress. For instance, Joshy et al. [28] employed NFGNN in the search for drugs to treat orphan diseases, validating its effectiveness. However, the application of NFGNN in DTI prediction is quite scarce. The full potential of NFGNN in this area remains to be further explored.

In recent years, the self-attention mechanism has increasingly been incorporated into prediction models in the DTI field due to its advantages [29]. For instance, compared to DrugBAN developed by Bai et al. [14], CAT-DTI proposed by Zeng et al. [15] has introduced further innovations in the protein encoder module. They integrate the transformer architecture with CNN, enhancing the feature extraction capabilities of CNNs through the self-attention mechanism. Although CAT-DTI demonstrates superior performance compared to existing baseline models, the combination of these two architectures can complicate the training process and hyperparameter tuning. Furthermore, CAT-DTI does not adequately capture local site-specific interactions in drug–target pairs, which limits its interpretability. Therefore, further research is needed to explore the scientific integration of self-attention mechanisms into prediction models and to improve model interpretability.

Therefore, in order to address the aforementioned limitations and explore potential areas for improvement, we propose a drug–target interaction prediction model (NFSA-DTI) based on neural fingerprint and a self-attention mechanism. In conclusion, the key contributions of our work can be summarized as follows:

- We propose an enhancing self-attention convolutional module (ESACM) that utilizes the self-attention mechanism to enhance CNN's ability in capturing the long-distance

dependencies between the subsequences in the target amino acid sequence. ESACM enables the protein encoder to comprehensively calculate the impact of subsequences at different positions on the target sequence.

- We have successfully applied the neural fingerprint graph neural network (NFGNN) to the drug–target interaction prediction task and verified its effectiveness. In contrast to fixed fingerprints, which require extensive lookup tables to uniquely encode all possible molecular structures, neural fingerprints can encode all molecular structures using trainable parameters.
- We propose the integration of an attention pooling method into the bilinear attention network and demonstrate its effectiveness. This pooling method can assign weights to each element in the input matrix based on its importance, thereby augmenting the model's learning ability of the local pivotal binding sites within the drug–target pair.

2. Results and Discussion

2.1. Model Performance Comparison

We compared NFSA-DTI with seven baseline models in three benchmark datasets. The results of the comparative experiments are presented in Table 1.

Table 1. Performance comparison with baseline models on the BindingDB, BioSNAP, and Human dataset (**Best**).

Datasets	Model	AUROC	AUPRC	Accuracy	Sensitivity	Specificity
BindingDB	SVM	0.939	0.928	0.825	0.781	0.866
	RF	0.942	0.921	0.880	0.875	0.892
	DeepConv-DTI	0.945	0.925	0.882	0.873	0.894
	GraphDTA	0.951	0.934	0.888	0.882	0.897
	MolTrans	0.952	0.936	0.887	0.877	0.902
	DrugBAN	0.960	0.948	0.904	0.900	0.908
	CAT-DTI	0.960	0.947	0.896	0.884	**0.913**
	NFSA-DTI	**0.965**	**0.957**	**0.907**	**0.908**	0.906
BioSNAP	SVM	0.862	0.864	0.777	0.711	0.841
	RF	0.860	0.886	0.804	0.823	0.786
	DeepConv-DTI	0.886	0.890	0.805	0.760	0.851
	GraphDTA	0.887	0.890	0.800	0.745	0.854
	MolTrans	0.895	0.897	0.825	0.818	0.831
	DrugBAN	0.903	0.902	0.834	0.820	0.847
	CAT-DTI	**0.909**	0.907	0.836	**0.825**	0.847
	NFSA-DTI	**0.909**	**0.909**	**0.839**	0.819	**0.858**
Human	SVM	0.913	0.905	0.838	0.782	0.830
	RF	0.939	0.927	0.866	0.833	0.893
	DeepConv-DTI	0.978	0.982	0.878	0.830	0.938
	GraphDTA	0.965	0.955	0.908	0.912	0.904
	MolTrans	0.981	0.976	0.941	**0.949**	0.939
	DrugBAN	0.981	0.969	0.940	0.938	0.941
	CAT-DTI	0.983	0.976	0.942	0.929	**0.957**
	NFSA-DTI	**0.988**	**0.984**	**0.945**	0.944	0.955

In terms of AUROC, AUPRC, and accuracy, NFSA-DTI was better than the baseline models according to the results in Table 1. In the BindingDB dataset, compared with the currently best-performing CAT-DTI, AUROC, AUPRC, and accuracy of NFSA-DTI were improved by 0.5, 1.0, and 1.1 percentage points, respectively. We believe that the stronger global feature extraction capability, as well as the introduction of multiple attention mechanisms, motivated NFSA-DTI to perform better than CAT-DTI. However, the performance gap between NFSA-DTI and CAT-DTI in the BioSNAP dataset was not

particularly obvious. This phenomenon may be caused by the fact that on a balanced and data-heavy BioSNAP dataset the model may be more prone to overfitting during training because it must simultaneously process a large number of samples from various classes, and the features between different classes may interfere with each other, thus affecting the overall performance. To address this issue, we modified the number of neurons per layer in both the NFGNN and CNN models, as well as adjusted their dropout rates, to reduce the risk of potential overfitting. These modifications are intended to further narrow the performance gap with the CAT-DTI model in terms of AUROC and AUPRC.

In terms of sensitivity and specificity, NFSA-DTI did not show complete advantages over the baseline models. This may be because the model tended to predict a certain class of samples, thus inevitably sacrificing a part of the sensitivity to the minority class or the specificity to the majority class. However, the model performed well on AUROC and AUPRC, which indicated that the model could still effectively distinguish between positive and negative samples in general. Overall, NFSA-DTI outperformed all baseline models under most evaluation metrics, thus demonstrating the effectiveness of our proposed model.

2.2. Ablation Study

In order to verify the influence of each innovation module in NFSA-DTI on the model's performance, we designed several model variants to conduct ablation experiments. The results are shown in Figure 1.
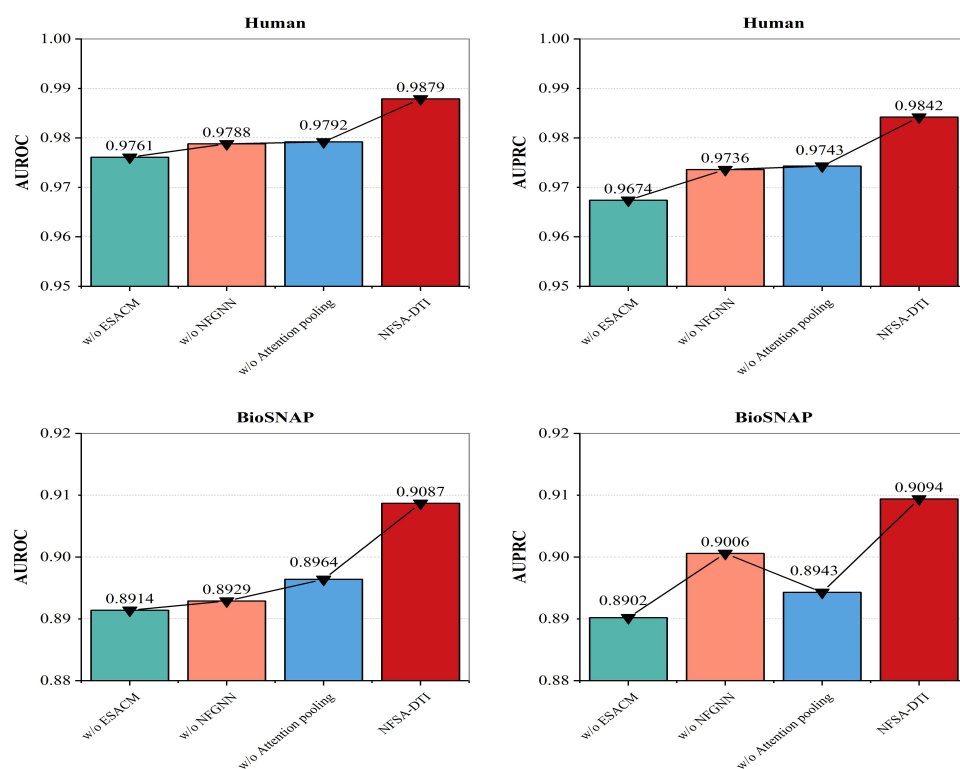

Figure 1. Ablation study on the Human and BioSNAP datasets.

The three ablation models are explained as follows:

(1) w/o ESACM: Protein encoder exclusively employs a three-layer CNN without the addition of the self-attention enhancing unit (others same as NFSA-DTI).

(2) w/o NFGNN: Drug encoder employs a three-layer GCN instead of a three-layer NFGNN (others same as NFSA-DTI).

(3) w/o Attention pooling: Sum pooling is used for the final pooling of the bilinear attention network instead of attention pooling (others same as NFSA-DTI).

According to the results in Figure 1, compared with w/o ESACM, NFSA-DTI achieved a 1.18 percentage point increase in AUROC and a 1.68 percentage point increase in AUPRC on the Human dataset, and a 1.73 percentage point increase in AUROC and a 1.92 percentage point increase in AUPRC on the BioSNAP dataset. This is because, compared with the pure use of CNNs as the protein encoder, ESACM could use the self-attention mechanism to effectively enhance the ability of CNNs to capture the long-distance dependencies between the subsequences in the target amino acid sequence, thus further improving the performance. Compared with w/o NFGNN and w/o Attention pooling, NFSA-DTI also had relatively obvious advantages, which also indicated that compared with using GNN alone as the drug encoder, NFGNN used the neural fingerprint encoding the entire molecular structure as an additional global information supplement. In addition, compared with the fixed fingerprint, the neural fingerprint could be obtained by parameter learning and had more efficient retrieval advantages. Compared with sum pooling as the final pooling operation of the bilinear attention network, attention pooling could enhance the model's ability to learn the local important binding sites in a drug–target pair.

2.3. Case Study

In order to verify our model's ability to discover new drugs and new target proteins, we constructed a new test dataset from the DrugBank online database, consisting of drug–target pairs absent in the BindingDB dataset, and we used the model trained on the BindingDB dataset to conduct the test. Taking Vascular Endothelial Growth Factor Receptor 2 and Nortriptyline as examples, according to the interaction probability score p, the predicted top 10 drug and target protein candidates are shown in Table 2 and Table 3, respectively.

Table 2. The predicted top 10 drug candidates for Vascular Endothelial Growth Factor Receptor 2 (P35968) in the newly constructed dataset.

Rank	Drug Name	DrugBank ID	Evidence
1	**Sorafenib**	**DB00398**	Iyer et al. [30]
2	Regorafenib	DB08896	Southan et al. [31]
3	2-Aminobenzimidazole	DB06938	Unknown
4	1-Naphthalenecarboxamide	DB07274	Unknown
5	Ponatinib	DB08901	Unknown
6	**Sunitinib**	**DB01268**	Schoffski et al. [32]
7	Tyrosine Kinase-IN-1	DB05014	Unknown
8	**Lenvatinib**	**DB09078**	Matsui et al. [33]
9	Fostamatinib	DB12010	Unknown
10	RAF265	DB05984	Southan et al. [31]

Table 3. The predicted top 10 target protein candidates for Nortriptyline (DB00540) in the newly constructed dataset.

Rank	Protein Name	Uniprot ID	Evidence
1	**Sodium-dependent noradrenaline transporter**	**P23975**	Kim et al. [34]
2	Alpha-2A adrenergic receptor	P08913	Unknown
3	5-hydroxytryptamine receptor 2A	P28223	Southan et al. [31]
4	5-hydroxytryptamine receptor 1A	P08908	Southan et al. [31]
5	5-hydroxytryptamine receptor 1C	P08909	Southan et al. [31]
6	5-hydroxytryptamine receptor 2C	P28335	Southan et al. [31]
7	Alpha-1B adrenergic receptor	P35368	Unknown
8	**Sodium-dependent serotonin transporter**	**P31645**	Vaishnavi et al. [35]
9	Beta-1 adrenergic receptor	P08588	Unknown
10	Muscarinic acetylcholine receptor M1	P11229	Unknown

As can be seen in Tables 2 and 3, our model successfully predicted three drugs and two proteins (marked in bold) that could be validated in the DrugBank online database and two drugs and four proteins (marked with underline) that could be validated in the TTD [36] online database. The pharmacological effects of other drug–target pairs in the above two tables were currently unknown in the DrugBank and TTD databases and need to be further verified. It is worth mentioning that the drug candidate or target protein candidate with the highest prediction score in the table was consistent with the ranking recorded in the DrugBank database.

Based on the unknown drug–target interaction pairs mentioned above, researchers can discover new drug candidates that may have therapeutic potential, especially for those diseases that have not been fully studied [37]. For example, in Table 2, 2-aminobenzimidazole would potentially act on Vascular Endothelial Growth Factor Receptor 2 as a new drug. Moreover, for already approved drugs, this will help to search for their possible new mechanisms of action and new therapeutic applications, thereby helping researchers to explore the potential of these drugs in the treatment of other diseases [38]. Similarly, in Table 3, Nortriptyline may act as a repositioning drug on a new target Alpha-2A adrenergic receptor. Although these unknown predictions require further experimental validation, they provide new directions and possibilities for drug research.

2.4. Interpretability Analysis

Another advantage of NFSA-DTI is that it provides molecular-level interpretations essential for drug design efforts. For example, in the Human test dataset, we randomly selected a drug–target interaction pair and plotted the 2D visualization result of its drug molecule as shown in Figure 2a, and we further verified our result by using the Protein Data Bank (PDB) [39] online database, as shown in Figure 2b,c.

According to Figure 2a, our model accurately predicted the involvement of the carboxylate group of mycophenolic acid in receptor binding. This prediction was further validated in Figure 2b, where both oxygen atoms of the carboxylate group were shown to act as hydrogen bond acceptors. One oxygen atom interacts with the back-bone of Ser275 and Ser276, while the other forms a specific interaction with the sidechain of Ser276. Additionally, in Figure 2a, our model predicted that the hydroxyl and methyl groups on the phenyl ring, as well as the carbonyl oxygen atom, are involved in receptor binding. This prediction is confirmed in Figure 2b, where the hydroxyl and methyl groups act as hydrogen bond donors, forming interactions with the sidechains of Thr333 and Asp274, respectively. The carbonyl oxygen atom functions as a hydrogen bond acceptor, interacting with the sidechains of Thr333 and Cys331. Unfortunately, the oxygen atom of the epoxy group was mistakenly predicted to be involved in the receptor interaction. Nonetheless, the overall interpretability of our model provides significant reference value.

In Figure 2b, the dark blue ligand exposure area is mostly concentrated near the carboxylate group of mycophenolic acid. In addition, the light blue receptor exposure area is mainly concentrated near Ser276. Combined with the orange area in Figure 2a indicating possible interaction, it can be inferred that the carboxylate group's oxygen negative ion and the oxygen atom of the carboxylate group are the most important in this interaction. The 3D visualization of Figure 2c, as a supplement to the 2D visualization of Figure 2b, will help researchers analyze the 3D structure of mycophenolic acid and inosine monophosphate dehydrogenase acceptor and identify key pharmacophore and binding site.
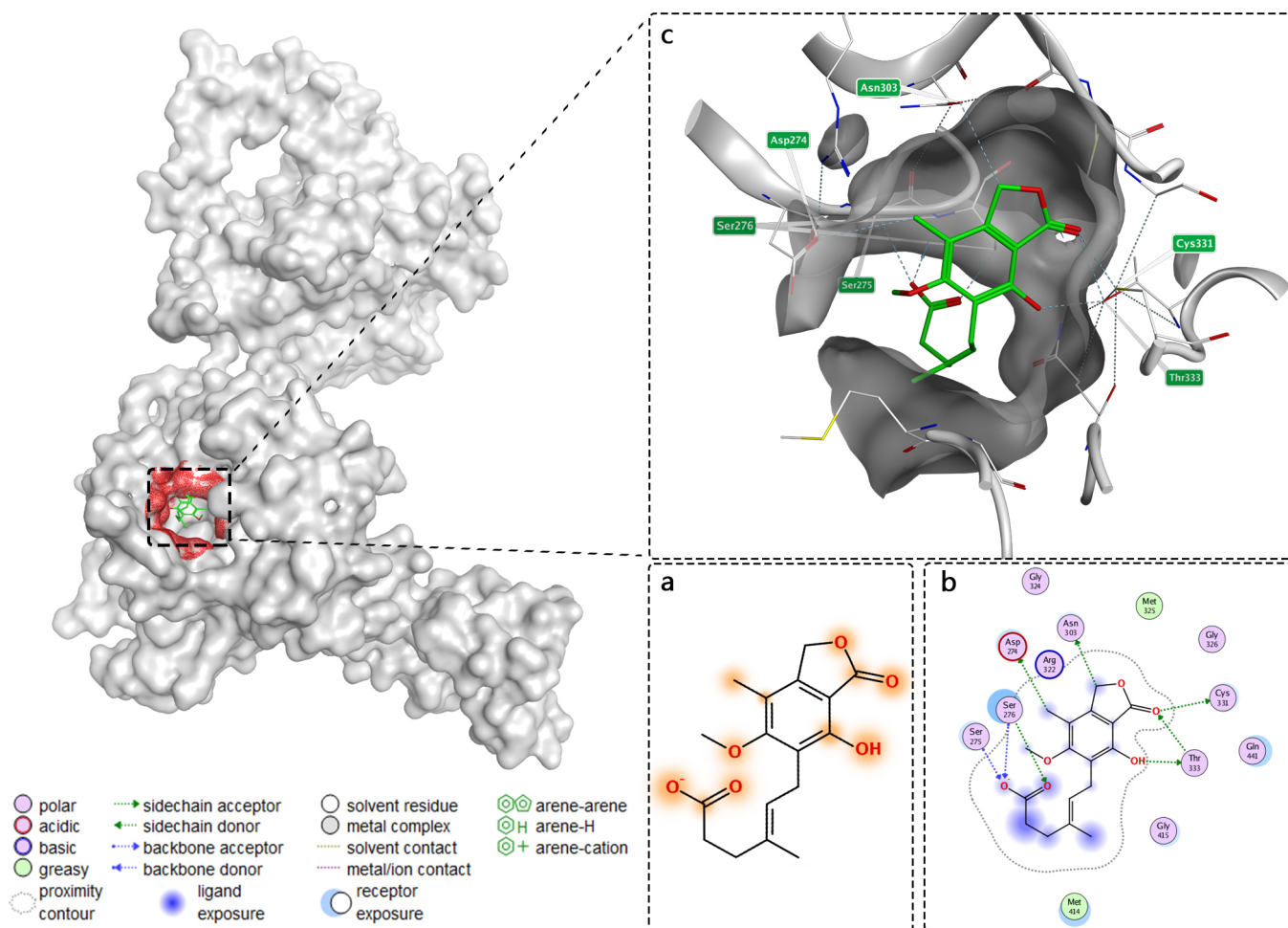
Figure 2. (**a**) The 2D visualization result of mycophenolic acids obtained from NFSA-DTI. The orange highlights indicate possible local binding sites, with darker color and larger area indicating greater likelihood. (**b**,**c**) The 2D and 3D diagrams of the interaction between mycophenolic acid and inosine monophosphate dehydrogenase from the PDB online database, drawn by the software Molecular Operating Environment (MOE 2019.0102) [40].

3. Materials and Methods

3.1. Datasets

We evaluated NFSA-DTI using three publicly available DTI datasets: BindingDB, BioSNAP, and Human. Information about them is shown below.

- The BindingDB Dataset [41] is a large drug–target dataset, containing thousands of small molecule drugs and protein targets. The targets also cover different species, but mainly focus on human targets. It is worth mentioning that BindingDB is unbalanced in terms of dataset distribution.
- The BioSNAP dataset, developed by Huang et al. [18] and Zitnik et al. [42] from the DrugBank database [43], is a balanced dataset. It includes verified positive samples as well as an equal number of randomly paired negative samples that have never been encountered before. This dataset considers interactions between small chemical drugs and target proteins, all of which have been experimentally validated via biological experiments or formal pharmacological studies.
- The Human dataset, constructed by Liu et al. [44], is a balanced dataset, incorporating high-confidence negative samples obtained through silicon screening methods.

Table 4 gives the detailed statistics of these three datasets. In the table, Interactions* are defined as the total number of interacting sample pairs and non-interacting sample

pairs in the dataset. P2N represents the ratio of the number of interacting sample pairs to non-interacting sample pairs in the dataset.

Table 4. Experimental datasets statistics.

Datasets	Drugs	Proteins	Interactions*	P2N
BindingDB	14,643	2623	49,199	0.725
BioSNAP	4510	2181	27,464	1.014
Human	2726	2001	6728	1

The positive samples in the three datasets are based on the literature and experimental verification. However, due to the relative scarcity of experimentally verified negative sample data, researchers usually need to adopt different strategies to construct negative sample sets [9,16]. For the Human dataset, based on the existing biological data and chemical characteristics, the researchers used silicon screening computing technology to efficiently screen out the combinations with low correlation as negative samples, so as to improve the confidence of negative samples. Regarding the BioSNAP dataset, using the biological network construction method, the researchers identified pairs of nodes that were not directly connected in the network and randomly selected negative samples from them to ensure the confidence of the negative samples. For the BindingDB dataset, researchers generated negative samples by excluding known drug–target binding pairs and rigorously screened negative samples to ensure that these combinations did not overlap with positive samples.

3.2. Baselines

In order to comprehensively evaluate the prediction performance of our model, we chose seven representative baseline models, including two machine learning models (SVM [45] and RF [46]), and five deep learning models (DeepConv-DTI [16], GraphDTA [13], MolTrans [18], DrugBAN [14], and CAT-DTI [15]). Information about them is shown below.

- **SVM**: By learning the optimal hyperplane, the interaction between drugs and targets can be effectively distinguished in the high-dimensional feature space, which has a relatively strong classification ability and good generalization performance.
- **RF**: By integrating multiple decision trees, the interaction between drugs and targets is predicted in a voting manner, which has strong anti-noise and robustness. It performs well when dealing with high-dimensional data, but may be vulnerable to uneven feature importance.
- **DeepConv-DTI**: By using a convolutional neural network, amino acid subsequences of various lengths are convolved to capture local residue patterns, and a fully connected neural network is used to encode the fixed ECFP4 drug fingerprint. It outperforms previous models based on protein descriptors.
- **GraphDTA**: By representing drugs as graphs and using the graph neural network to predict the affinity of the drug to the target, it can effectively process the topological structure data of drug molecules and improve the prediction accuracy.
- **MolTrans**: By introducing the self-attention mechanism of the transformer, the drug molecules and protein sequences are embedded into a unified vector space, in order to effectively capture the interaction characteristics between them. It has high flexibility and performance in dealing with complex molecular relationships.
- **DrugBAN**: By introducing a bilinear attention network, the interaction strength between the drug and the substructure of the target will be embedded into the bilinear attention matrix for downstream prediction tasks. It can better capture the local feature correlation and improve the performance of the model.
- **CAT-DTI**: By combining the graph convolutional neural network, a transformer architecture, and the cross-attention mechanism, it can effectively capture the information of drug and target sequences and improve the prediction performance. It has an

advantage when dealing with long target sequences and is better able to model complex interactions.

3.3. Metrics

In our experiment, the AUROC (area under the receiver operating characteristic curve) and AUPRC (area under the precision-recall curve) serve as primary metrics for assessing model classification performance. AUROC measures the model's ability to discriminate between positive and negative samples across different classification thresholds. It considers both the true positive rate (sensitivity) and the false positive rate (1—specificity). A higher AUROC value indicates better discriminative power and overall performance. AUPRC, on the other hand, focuses on the trade-off between precision and recall. It measures the model's ability to correctly identify positive samples while minimizing false positives. In addition, we adopted accuracy, sensitivity, and specificity under the optimal F1 score threshold as additional evaluation metrics.

3.4. Implement Details

To optimize the implementation of NFSA-DTI, we randomly split each dataset into train/validate/test sets at a ratio of 0.7/0.1/0.2. The test set is employed to evaluate the model's final performance and is entirely independent of the training and validation processes, thereby ensuring the objectivity and accuracy of the evaluation results. Compared to the 7:2:1 division, the 7:1:2 division enhances the reliability of the model's final evaluation. We used the Human dataset to train the NFSA-DTI model and determined the initial combination of hyperparameters. Subsequently, we fine-tuned the parameters and trained NFSA-DTI. When NFSA-DTI exhibited superior performance on the validation set of each dataset, we determined the optimal values of the hyperparameters. The partial learning curves for the BindingDB validation set are depicted in Figure 3, illustrating the influence of varying certain hyperparameter values. The figure demonstrates that the optimal performance of the model is achieved when the learning rate of the optimizer reaches 5×10^{-5} and the number of attention heads in the bilinear attention network is set to 2. The main hyperparameters configuration is shown in Table 5. In this study, we conducted our experiments using Python 3.8 as the programming language. The deep learning framework used is PyTorch version 1.12, and we trained the model using an RTX 4090 24 G, accelerating the computation using NVIDIA CUDA version 11.6.

Table 5. Hyperparameters configuration.

| Module	Hyperparameters	Value
ESACM	Initial amino acid embedding dimension	128
	Kernel size	[3, 6, 9]
	Number of filters	[128, 128, 128]
	Heads of self-attention	2
NFGNN	Initial atom embedding dimension	128
	Hidden node dimensions	[128, 128, 128]
Bilinear attention network	Heads of bilinear attention	2
	Bilinear embedding dimension	768
	Attention pooling window size	3
	Attention pooling stride	3
Fully connected decoder	Number of hidden neurons	512
Optimizer	Learning rate	5×10^{-5}
	Epoch	100
Mini-batch	Batch size	64
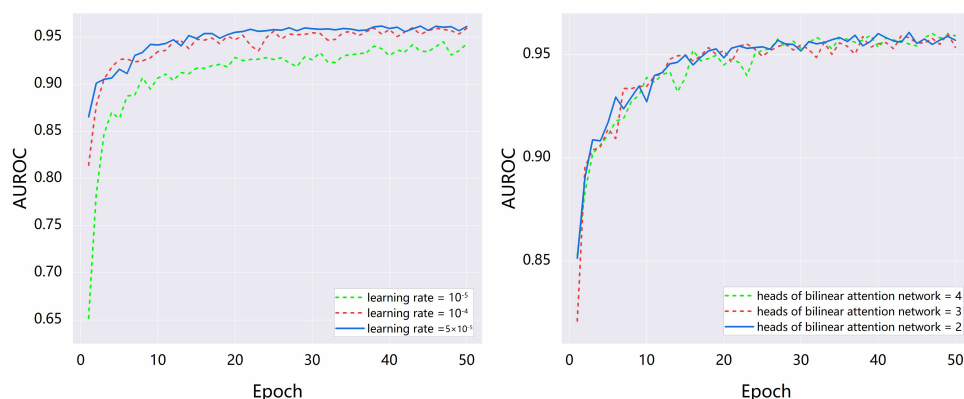
Figure 3. Learning curves of NFSA-DTI when changing some hyperparameters on the validation set of the BindingDB dataset.

3.5. Methods

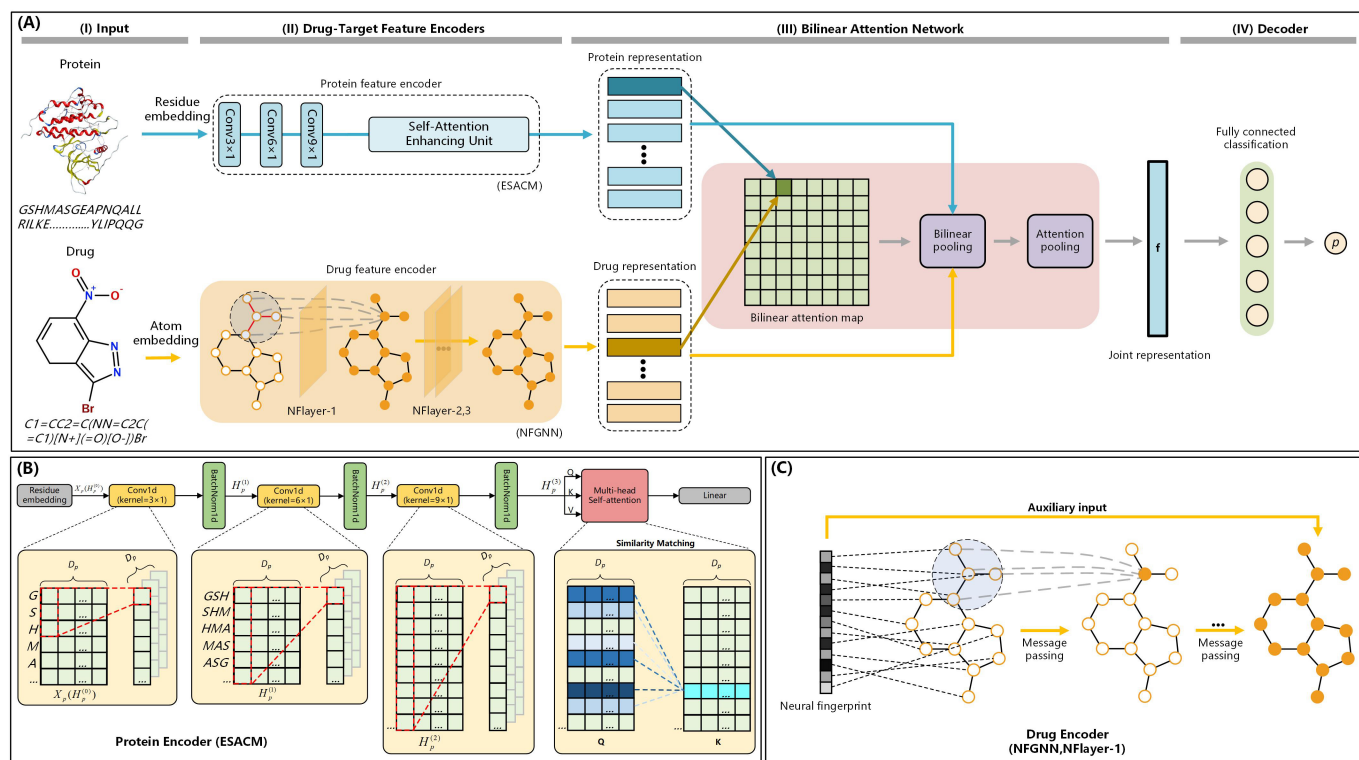Figure 4A shows the framework of our proposed NFSA-DTI model.

Figure 4. (**A**) The framework of NFSA-DTI. It includes four stages. (I) The target protein's amino acid sequence is transformed into a feature matrix, while the drug molecule's SMILES is processed into graph structure data. (II) The 3-layer CNN processes the two-dimensional feature matrix and obtains the protein representation after the self-attention enhancing unit. Correspondingly, the 3-layer NFGNN processes the graph structure data and obtains the drug representation. (III) Interactions between protein and drug representations are computed via a bilinear attention mechanism, thereby producing a bilinear attention map. (IV) The joint representation obtained after pooling will be input to the fully connected layer for computing the prediction score p. (**B**) The framework of ESACM. It comprises three 1D convolutional layers, the corresponding 1D batch normalization layers, a self-attention enhancing unit, and a linear layer. The self-attention enhancing unit consists of two stages. (Stage I) Query, key, and value are computed based on the input matrix, in conjunction with their respective weight matrices. (Stage II) The similarity between query and key is evaluated to

derive the attention weights. Subsequently, the output is generated through a weighted summation of the values, utilizing the computed attention weights. (**C**) The flowchart of NFGNN. Firstly, the target node and neighbor nodes surrounding the target node in the molecular graph are integrated and encoded into numerical features. Similar operations are then performed on subsequent nodes to obtain the final neural fingerprint. Subsequently, after the message passing mechanism has completed one iteration of the graph, the neural fingerprint serves as a fixed auxiliary input for updating the graph after each NFlayer.

3.5.1. Problem Formulation and Prerequisites

The objective of DTI prediction is to develop a model M, which effectively maps the combined feature representation space $P \times G$ of a protein sequence P and a drug molecular graph G, yielding an interaction probability score $p \in [0, 1]$.

The target protein is represented by one amino acid sequence denoted as $P = (a_1, \ldots, a_n)$, where each token corresponds to one of the 23 amino acids. The maximum length of P is set to 1200, which is long enough to cover common amino acid sequences while remaining computationally and storage efficient. Amino acid sequences that exceed this threshold are truncated. For amino acid sequences that are not of length 1200, zeros will be padded at the end of the sequence. The amino acid sequences [47] are a standardized representation of proteins that can be expressed and processed in a computable way.

Major current deep learning methods for predicting drug–target interaction typically use the simplified molecular input line entry system (SMILES) [48] as an input representation. The limitation of SMILES lies in its one-dimensional nature, which hinders its ability to capture the intricate spatial arrangement of molecules, potentially resulting in the loss of crucial chemical information. However, in our model, the input SMILES are converted into the corresponding two-dimensional graph structure data G. It creates actual nodes and edges based on the drug molecule's atoms and chemical bonds, with virtual nodes and self-loop edges filling in the rest. Node count is capped at 290.

3.5.2. Protein Feature Encoder

The conventional CNN architecture exhibits certain limitations in handling long-distance dependencies of sequences due to the restricted receptive field of convolution operations [49]. In contrast, the self-attention mechanism enables each position to interact with all other positions within the tensor, granting access to comprehensive information about the entire tensor [50]. Therefore, we propose ESACM to augment model performance. ESACM integrates a self-attention mechanism that enhances the CNNs' ability to capture the long-distance dependencies between the subsequences in the target amino acid sequence. The framework diagram of ESACM is shown in Figure 4B.

Building upon the concept of word embedding for amino acid sequences by Bai et al. [14], the amino acids are initialized into a learnable embedding matrix $\mathbf{E}_p \in \mathbb{R}^{23 \times D_p}$, where 23 is the number of amino acid types and D_p is the latent space dimension. By looking up $\mathbf{E}_p$, each protein sequence P can be initialized to the corresponding feature matrix $\mathbf{X}_p \in \mathbb{R}^{\Theta_p \times D_p}$, where the parameter Θ_p represents the maximum permissible length of a protein sequence, which is generally 1200 in our experiment, and each row of the matrix denotes a residue representation in the target amino acid sequence. In the initial convolution layer, the feature matrix $\mathbf{X}_p$ will be input into the first CNN for one-dimensional convolution to extract amino acid sequence features, and this process utilizes a convolution kernel sized 3×1 and a stride of 1, such as "GSHMAS...PQQG" $\longrightarrow$ "GSH", "SHM", "HMA", "MAS", ..., "PQQ", and "QQG". The next two layers keep the same stride as the first, but further enlarge the kernel size to acquire more features of local subsequences. The convolutional operations can be expressed by the following formula:

$$\mathbf{H}_p^{(l+1)} = \sigma(\text{CNN}(\mathbf{W}_c^{(l)}, \mathbf{b}_c^{(l)}, \mathbf{H}_p^{(l)})), \tag{1}$$

where $\mathbf{H}_p^{(l+1)}$ is the *l*-th hidden protein representation and $\mathbf{H}_p^{(0)} = \mathbf{X}_p$. $\sigma(\cdot)$ denotes ReLU$(\cdot)$. $\mathbf{W}_c^{(l)}$ and $\mathbf{b}_c^{(l)}$ are learnable weight matrices (filters) and bias vector in the *l*-th CNN layer.

The output of the third convolutional layer will subsequently undergo computation in the self-attention enhancing unit. Specifically, the self-attention enhancing unit consists of two stages, corresponding to the following formulas.

Stage I:

$$\mathbf{q}_{i,j}^{(l)} = \mathbf{W}_q^{(l)}\mathbf{f}_{i,j}, \mathbf{k}_{i,j}^{(l)} = \mathbf{W}_k^{(l)}\mathbf{f}_{i,j}, \mathbf{v}_{i,j}^{(l)} = \mathbf{W}_v^{(l)}\mathbf{f}_{i,j}. \tag{2}$$

Stage II:

$$p_{i,j} = \prod_{l=1}^{N}\left(\sum_{a,b\in\mathcal{N}_m(i)} A(\mathbf{q}_{i,j}^{(l)}, \mathbf{k}_{a,b}^{(l)})\mathbf{v}_{a,b}^{(l)}\right). \tag{3}$$

The specific formula is presented as follows:

$$p_{i,j} = \prod_{l=1}^{N}\left(\sum_{a,b\in\mathcal{N}_m(i)} A(\mathbf{W}_q^{(l)}\mathbf{f}_{i,j}, \mathbf{W}_k^{(l)}\mathbf{f}_{a,b})\mathbf{W}_v^{(l)}\mathbf{f}_{a,b}\right), \tag{4}$$

where $p_{i,j}$ is the value in row *i* and column *j* of the protein representation. $\prod$ is the concatenation of the outputs of *N* attention heads. $\mathcal{N}_m(i)$ represents one column region whose length is the maximum column length *m* of the feature matrix, and its abscissa is *i*. $A(\mathbf{W}_q^{(l)}\mathbf{f}_{i,j}, \mathbf{W}_k^{(l)}\mathbf{f}_{a,b})$ is the corresponding attention weight with regard to the features within $\mathcal{N}_m(i)$. $\mathbf{W}_q^{(l)}, \mathbf{W}_k^{(l)}$, and $\mathbf{W}_v^{(l)}$ are the projection matrices for queries, keys, and values, respectively. $\mathbf{f}_{i,j} \in \mathbf{H}_p^{(3)}$ and $\mathbf{H}_p^{(3)}$ denote the output matrix of the third CNN. The attention weight is calculated as follows:

$$A(\mathbf{W}_q^{(l)}\mathbf{f}_{i,j}, \mathbf{W}_k^{(l)}\mathbf{f}_{a,b}) = \text{softmax}\left(\frac{(\mathbf{W}_q^{(l)}\mathbf{f}_{i,j})^\top(\mathbf{W}_k^{(l)}\mathbf{f}_{a,b})}{\sqrt{d}}\right), \tag{5}$$

where *d* is the feature dimension of $\mathbf{W}_q^{(l)}\mathbf{f}_{i,j}$.

3.5.3. Drug Feature Encoder

Graph neural network (GCN) [51] primarily updates node representations through local neighborhood information propagation, its ability to model the global structure is relatively limited. NFGNN, a type of neural network proposed by Duvenaud et al. [24], was originally designed for predicting molecular properties and has been proven effective in learning the representation of molecular graphs. Compared with GCN, NFGNN uses the neural fingerprint as an additional global information supplement. Moreover, in contrast to fixed fingerprints that necessitate a substantial number of lookup tables for encoding all possible molecular structures uniquely, neural fingerprints of NFGNN can encode the entirety of molecular structures using trainable parameters. Given the advantages of NFGNN, we therefore introduce NFGNN as the drug encoder of NFSA-DTI.

The flowchart of NFGNN is illustrated in Figure 4C. Specifically, the construction process of the neural fingerprint is represented by the following formulas:

$$\mathbf{r}_a = g(a), \tag{6}$$

$$\mathbf{v}_i = \mathbf{r}_i + \sum_{j=1}^{N}\mathbf{r}_j, j \in \mathcal{N}(i), \tag{7}$$

$$\mathbf{f} = \sum_{i=1}^{M}(\text{softmax}(\sigma(\mathbf{v}_i\mathbf{H}_h)\mathbf{W}_o)), \tag{8}$$

where a represents the node. $g(\cdot)$ is the encoding function of node features. $\mathbf{v}_i$ represents the sum of features about the i-th node and its neighbors. $\mathcal{N}(i)$ is the set of neighbor nodes of the i-th node. $\mathbf{H}_h$ is the hidden weight, and $\mathbf{W}_o$ is the output weight. N denotes the number of neighbor nodes of the i-th node, and M denotes the total number of nodes in the graph. σ represents ReLU$(\cdot)$. The subsequent message passing process can be expressed by the following formulas:

$$\mathbf{m}_{ij}^{(l)} = \rho^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \mathbf{h}_{e_{ij}}^{(l)}), j \in \mathcal{N}(i), \tag{9}$$

$$\mathbf{m}_i^{(l)} = \zeta^{(l)}(\mathbf{m}_{ij}^{(l)} | j \in \mathcal{N}(i)), \tag{10}$$

$$\mathbf{h}_i^{(l+1)} = \phi^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{m}_i^{(l)}, \mathbf{f}), \tag{11}$$

where $\rho^{(\cdot)}$, $\zeta^{(\cdot)}$, and $\phi^{(\cdot)}$ are functions of message construction, message aggregation, and vertex update of the l-th layer NFGNN, respectively. $\mathbf{h}_i^{(l)}$ is the node feature of the i-th node at the l-th layer, $\mathbf{h}_j^{(l)}$ is its neighbor node feature, $\mathbf{h}_{e_{ij}}^{(l)}$ is the corresponding edge feature, and $\mathbf{m}$ is the message.

3.5.4. Bilinear Attention Network

We employ the bilinear attention network proposed by Bai et al. [14] and improve its subsequent pooling operation. Specifically, we replace the summation pooling with an attention pooling. The sum pooling method is relatively straightforward to implement and incurs a low computational cost. However, it lacks the ability to retain the intensity information of features, potentially compromising learning performance. To address this limitation and enhance the model's capacity in capturing important information between drug substructures and target substructures, we employ an attention pooling approach inspired by Er et al.'s [52] investigation on the importance of sentence components. The attention pooling assigns weights based on the significance of each element in the input matrix, generating an enhanced pooled representation. The enhanced pooled representation promotes the model's learning of the drug–target interaction modeling process, thereby improving the overall interpretability of the model. The overall flowchart of the improved bilinear attention network is shown in Figure 5.
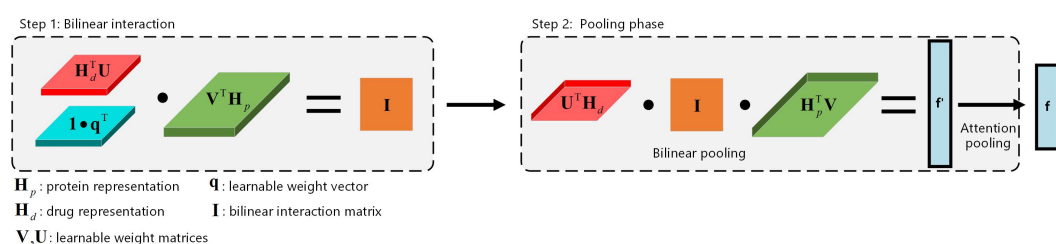

Figure 5. The flowchart of the bilinear attention network. This module consists of two steps. (Step 1) A bilinear interaction matrix is derived through the computation of protein and drug representations. (Step 2) Subsequently, a joint representation is obtained by employing bilinear pooling and attention pooling.

Given the protein representation and drug representation $\mathbf{H}_p^{(3)} = \{\mathbf{h}_p^1, \mathbf{h}_p^2, \ldots, \mathbf{h}_p^M\}$, among them, $\mathbf{h}_p^1 = \{p_{1,1}, p_{1,2}, \ldots, p_{1,e}\}$, ..., $\mathbf{h}_p^M = \{p_{M,1}, p_{M,2}, \ldots, p_{M,e}\}$, and $\mathbf{H}_d^{(3)} = \{\mathbf{h}_1^{(3)}, \mathbf{h}_2^{(3)}, \ldots, \mathbf{h}_N^{(3)}\}$ via the protein and drug encoders, where M represents the maximum number of target protein substructures encoded by ESACM, N represents the maximum number of nodes in the drug molecular graph, and e represents the embedding dimension of the CNN. The bilinear interaction matrix $\mathbf{I} \in \mathbb{R}^{N \times M}$ can be obtained as follows:

$$\mathbf{I} = ((\mathbf{1} \cdot \mathbf{q}^{\top}) \circ \sigma((\mathbf{H}_d^{(3)})^{\top}\mathbf{U})) \cdot \sigma(\mathbf{V}^{\top}\mathbf{H}_p^{(3)}), \tag{12}$$

where $\mathbf{U}$ and $\mathbf{V}$ are learnable weight matrices for drug and protein representations, $\mathbf{q}$ is a learnable weight vector, $\mathbf{1}$ is a fixed all-ones vector, and the symbols $\circ$ and σ represent the Hadamard (element-wise) product and $\text{ReLU}(\cdot)$. The elements in $\mathbf{I}$ represent the intensity of interaction between drug–target substructures, indicating their potential binding sites. In order to intuitively understand bilinear interaction, element $\mathbf{I}_{i,j}$ can be written as follows:

$$\mathbf{I}_{i,j} = \mathbf{q}^{\top}(\sigma(\mathbf{U}^{\top}\mathbf{h}_d^i) \circ \sigma(\mathbf{V}^{\top}\mathbf{h}_p^j)), \tag{13}$$

where $\mathbf{h}_d^i$ is the i-th row of $\mathbf{H}_d^{(3)}$ and $\mathbf{h}_p^j$ is the j-th row of $\mathbf{H}_p^{(3)}$, respectively, denoting the i-th and j-th substructural representations of the drug and protein.

In order to obtain the joint representation $\mathbf{f}'$, the interaction matrix $\mathbf{I}$ is processed by a bilinear pooling layer. Specifically, $\mathbf{f}'$ is computed as follows:

$$\mathbf{f}' = \sigma((\mathbf{H}_d^{(3)})^{\top}\mathbf{U})^{\top} \cdot \mathbf{I} \cdot \sigma((\mathbf{H}_p^{(3)})^{\top}\mathbf{V}), \tag{14}$$

where the weight matrices $\mathbf{U}$ and $\mathbf{V}$ are shared with the preceding interaction matrix layer to reduce the number of parameters and mitigate overfitting.

Subsequently, the final joint representation $\mathbf{f}$ will be obtained after the attention pooling, which is represented by the following formulas:

$$\mathbf{W}_{i,j} = \text{Softmax}(\mathbf{b}_a^{\top}\tanh(\mathbf{U}_a\mathbf{I}_{i,j})), \tag{15}$$

$$\mathbf{f} = \sum_{i=1}^{N}\sum_{j=1}^{M}(\mathbf{W}_{i,j} \cdot \mathbf{b}_a')\mathbf{f}', \tag{16}$$

where $\mathbf{b}_a$ and $\mathbf{b}_a'$ represent the bias vectors, $\mathbf{U}_a$ denotes the weight matrix, and $\mathbf{W}_{i,j}$ is the attention weight vector.

3.5.5. Fully Connected Classification

The interaction probability is computed by feeding the joint representation $\mathbf{f}$ into the decoder, which comprises a single fully connected classification layer followed by a sigmoid function, as follows:

$$p = \text{Sigmoid}(\mathbf{W}_o\mathbf{f} + \mathbf{b}_o), \tag{17}$$

where $\mathbf{W}_o$ is the learnable weight matrix and $\mathbf{b}_o$ is the bias vector.

3.5.6. Backpropagation

The backpropagation process aims to minimize the cross-entropy loss [53] in order to optimize the model's performance. The formula is presented as follows:

$$\mathcal{L} = -\sum_i(y_i\log(p_i) + (1 - y_i)\log(1 - p_i)) + \frac{\lambda}{2}\|\Theta\|_2^2, \tag{18}$$

where y_i denotes the ground truth label for the i-th drug–target pair, p_i represents its output probability according to the model, Θ represents all the weight matrices and bias vectors that can be learned, and λ is a hyperparameter used for L2 [54] regularization.

4. Conclusions

In this article, we propose a novel DTI prediction model called NFSA-DTI. In the protein encoder, ESACM captures the long-distance dependencies between the subsequences in the target amino acid sequence while extracting the local features of the sequence. In the drug encoder, NFGNN extracts features of the drug molecular graph via a message-passing

mechanism, supplemented by the neural fingerprint as global information. In the feature fusion module, the bilinear attention network generates an enhanced pooled representation via attention pooling, improving the model's ability to learn key local binding sites in the drug–target pair. The experimental results on three benchmark datasets demonstrated that NFSA-DTI outperformed all baseline models in terms of prediction performance. Further case experiments showed that the model has provided valuable insights for drug discovery efforts. Furthermore, the model provides more intuitive interpretability at the molecular level. In the future, we plan to extend our study by using other benchmark datasets.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: NFSA-DTI and datasets used in this study are available on GitHub at https://github.com/xiaofeinb233/NFSA_DTI (accessed on 20 October 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gholap, A.D.; Uddin, M.J.; Faiyazuddin, M.; Omri, A.; Gowri, S.; Khalid, M. Advances in Artificial Intelligence in Drug Delivery and Development: A Comprehensive Review. *Comput. Biol. Med.* **2024**, *178*, 108702. [CrossRef] [PubMed]
2. Sadybekov, A.V.; Katritch, V. Computational approaches streamlining drug discovery. *Nature* **2023**, *616*, 673–685. [CrossRef] [PubMed]
3. Sarkar, C.; Das, B.; Rawat, V.S.; Wahlang, J.B.; Nongpiur, A.; Tiewsoh, I.; Lyngdoh, N.M.; Das, D.; Bidarolli, M.; Sony, H.T. Artificial intelligence and machine learning technology driven modern drug discovery and development. *Int. J. Mol. Sci.* **2023**, *24*, 2026. [CrossRef] [PubMed]
4. Vora, L.K.; Gholap, A.D.; Jetha, K.; Thakur, R.R.S.; Solanki, H.K.; Chavda, V.P. Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics* **2023**, *15*, 1916. [CrossRef] [PubMed]
5. Blanco-Gonzalez, A.; Cabezon, A.; Seco-Gonzalez, A.; Conde-Torres, D.; Antelo-Riveiro, P.; Pineiro, A.; Garcia-Fandino, R. The role of ai in drug discovery: Challenges, opportunities, and strategies. *Pharmaceuticals* **2023**, *16*, 891. [CrossRef]
6. Singh, S.; Kaur, N.; Gehlot, A. Application of artificial intelligence in drug design: A review. *Comput. Biol. Med.* **2024**, *179*, 108810. [CrossRef]
7. Askr, H.; Elgeldawi, E.; Aboul Ella, H.; Elshaier, Y.A.; Gomaa, M.M.; Hassanien, A.E. Deep learning in drug discovery: An integrative review and future challenges. *Artif. Intell. Rev.* **2023**, *56*, 5975–6037. [CrossRef]
8. Tang, X.; Lei, X.; Zhang, Y. Prediction of Drug-Target Affinity Using Attention Neural Network. *Int. J. Mol. Sci.* **2024**, *25*, 5126. [CrossRef]
9. Huang, Y.; Huang, H.Y.; Chen, Y.; Lin, Y.C.D.; Yao, L.; Lin, T.; Leng, J.; Chang, Y.; Zhang, Y.; Zhu, Z.; et al. A robust drug–target interaction prediction framework with capsule network and transfer learning. *Int. J. Mol. Sci.* **2023**, *24*, 14061. [CrossRef]
10. Wang, X.; Liu, J.; Zhang, C.; Wang, S. SSGraphCPI: A novel model for predicting compound-protein interactions based on deep learning. *Int. J. Mol. Sci.* **2022**, *23*, 3780. [CrossRef]
11. Ekins, S.; Puhl, A.C.; Zorn, K.M.; Lane, T.R.; Russo, D.P.; Klein, J.J.; Hickey, A.J.; Clark, A.M. Exploiting machine learning for end-to-end drug discovery and development. *Nat. Mater.* **2019**, *18*, 435–441. [CrossRef] [PubMed]
12. Wang, L.; Zhou, Y.; Chen, Q. Ammvf-dti: A novel model predicting drug–target interactions based on attention mechanism and multi-view fusion. *Int. J. Mol. Sci.* **2023**, *24*, 14142. [CrossRef] [PubMed]
13. Nguyen, T.; Le, H.; Quinn, T.P.; Nguyen, T.; Le, T.D.; Venkatesh, S. GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics* **2021**, *37*, 1140–1147. [CrossRef] [PubMed]
14. Bai, P.; Miljković, F.; John, B.; Lu, H. Interpretable bilinear attention network with domain adaptation improves drug–target prediction. *Nat. Mach. Intell.* **2023**, *5*, 126–136. [CrossRef]
15. Zeng, X.; Chen, W.; Lei, B. CAT-DTI: Cross-attention and Transformer network with domain adaptation for drug-target interaction prediction. *BMC Bioinform.* **2024**, *25*, 141. [CrossRef]

16. Lee, I.; Keum, J.; Nam, H. DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Comput. Biol.* **2019**, *15*, e1007129. [CrossRef]

17. Zhu, X.; Liu, J.; Zhang, J.; Yang, Z.; Yang, F.; Zhang, X. FingerDTA: A fingerprint-embedding framework for drug-target binding affinity prediction. *Big Data Min. Anal.* **2022**, *6*, 1–10. [CrossRef]

18. Huang, K.; Xiao, C.; Glass, L.M.; Sun, J. MolTrans: Molecular interaction transformer for drug–target interaction prediction. *Bioinformatics* **2021**, *37*, 830–836. [CrossRef]

19. Liu, S.; Wang, Y.; Deng, Y.; He, L.; Shao, B.; Yin, J.; Zheng, N.; Liu, T.Y.; Wang, T. Improved drug–target interaction prediction with intermolecular graph transformer. *Briefings Bioinform.* **2022**, *23*, bbac162. [CrossRef]

20. Ding, Y.; Tang, J.; Guo, F. Identification of drug-target interactions via multiple information integration. *Inf. Sci.* **2017**, *418*, 546–560. [CrossRef]

21. Li, Y.; Huang, Y.A.; You, Z.H.; Li, L.P.; Wang, Z. Drug-target interaction prediction based on drug fingerprint information and protein sequence. *Molecules* **2019**, *24*, 2999. [CrossRef] [PubMed]

22. Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [CrossRef] [PubMed]

23. Yang, J.; Cai, Y.; Zhao, K.; Xie, H.; Chen, X. Concepts and applications of chemical fingerprint for hit and lead screening. *Drug Discov. Today* **2022**, *27*, 103356. [CrossRef] [PubMed]

24. Duvenaud, D.K.; Maclaurin, D.; Iparraguirre, J.; Bombarell, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R.P. Convolutional networks on graphs for learning molecular fingerprints. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 2224–2232.

25. Cui, F.; Zhang, Z.; Zou, Q. Sequence representation approaches for sequence-based protein prediction tasks that use deep learning. *Briefings Funct. Genom.* **2021**, *20*, 61–73. [CrossRef]

26. Wei, W.; Wang, Z.; Mao, X.; Zhou, G.; Zhou, P.; Jiang, S. Position-aware self-attention based neural sequence labeling. *Pattern Recognit.* **2021**, *110*, 107636. [CrossRef]

27. Feldmann, C.W.; Sieg, J.; Mathea, M. Analysis of uncertainty of neural fingerprint-based models. *Faraday Discuss* **2024**. [CrossRef]

28. Joshy, A.; Kasyap, G.C.; Reddy, P.D.; Anjusha, I.T.; Nazeer, K.A.A. Drug Target Interaction Prediction using Graph Convo-lution based Neural Fingerprinting. In Proceedings of the 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 24–26 November 2022; pp. 1–6.

29. Bian, J.; Lu, H.; Dong, G.; Wang, G. Hierarchical multimodal self-attention-based graph neural network for DTI predic-tion. *Briefings Bioinform.* **2024**, *25*, bbae293. [CrossRef]

30. Iyer, R.; Fetterly, G.; Lugade, A.; Thanavala, Y. Sorafenib: A clinical and pharmacologic review. *Expert Opin. Pharmacother.* **2010**, *11*, 1943–1955. [CrossRef]

31. Southan, C.; Sharman, J.L.; Benson, H.E.; Faccenda, E.; Pawson, A.J.; Alexander, S.P.; Buneman, O.P.; Davenport, A.P.; McGrath, J.C.; Peters, J.A.; et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: Towards curated quantitative interactions between 1300 protein targets and 6000 ligands. *Nucleic Acids Res.* **2016**, *44*, D1054–D1068. [CrossRef]

32. Schöffski, P.; Dumez, H.; Clement, P.; Hoeben, A.; Prenen, H.; Wolter, P.; Joniau, S.; Roskams, T.; Van Poppel, H. Emerging role of tyrosine kinase inhibitors in the treatment of advanced renal cell cancer: A review. *Ann. Oncol.* **2006**, *17*, 1185–1196. [CrossRef] [PubMed]

33. Matsui, J.; Yamamoto, Y.; Funahashi, Y.; Tsuruoka, A.; Watanabe, T.; Wakabayashi, T.; Uenaka, T.; Asada, M. E7080, a novel inhibitor that targets multiple kinases, has potent antitumor activities against stem cell factor producing human small cell lung cancer H146, based on angiogenesis inhibition. *Int. J. Cancer* **2008**, *122*, 664–671. [CrossRef] [PubMed]

34. Kim, H.; Lim, S.W.; Kim, S.; Kim, J.W.; Chang, Y.H.; Carroll, B.J.; Kim, D.K. Monoamine transporter gene polymorphisms and antidepressant response in Koreans with late-life depression. *JAMA* **2006**, *296*, 1609–1618. [CrossRef] [PubMed]

35. Vaishnavi, S.N.; Nemeroff, C.B.; Plott, S.J.; Rao, S.G.; Kranzler, J.; Owens, M.J. Milnacipran: A comparative analysis of human monoamine uptake and transporter binding affinity. *Biol. Psychiatry* **2004**, *55*, 320–322. [CrossRef]

36. Chen, X.; Ji, Z.L.; Chen, Y.Z. TTD: Therapeutic target database. *Nucleic Acids Res.* **2002**, *30*, 412–415. [CrossRef]

37. Knox, C.; Wilson, M.; Klinger, C.M.; Franklin, M.; Oler, E.; Wilson, A.; Pon, A.; Cox, J.; Chin, N.E.; Strawbridge, S.A.; et al. DrugBank 6.0: The DrugBank Knowledgebase for 2024. *Nucleic Acids Res.* **2023**, *52*, D1265–D1275. [CrossRef]

38. Huang, Y.; Dong, D.; Zhang, W.; Wang, R.; Lin, Y.-C.; Zuo, H.; Huang, H.-Y.; Huang, H.-D. DrugRepoBank: A comprehensive database and discovery platform for accelerating drug repositioning. *Database* **2024**, *2024*, baae051. [CrossRef]

39. Bank, P.D. Protein data bank. *Nat. New Biol.* **1971**, *233*, 10–1038.

40. Vilar, S.; Cozza, G.; Moro, S. Medicinal chemistry and the molecular operating environment (MOE): Application of QSAR and molecular docking to drug discovery. *Curr. Top. Med. Chem.* **2008**, *8*, 1555–1572. [CrossRef]

41. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045–D1053. [CrossRef]

42. Zitnik, M.; Sosic, R.; Leskovec, J. BioSNAP Datasets: Stanford Biomedical Network Dataset Collection. 2018. Available online: http://snap.stanford.edu/biodata (accessed on 20 October 2024).

43. Wishart, D.S.; Knox, C.; Guo, A.C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36*, D901–D906. [CrossRef] [PubMed]

44. Liu, H.; Sun, J.; Guan, J.; Zheng, J.; Zhou, S. Improving compound–protein interaction prediction by building up highly credible negative samples. *Bioinformatics* **2015**, *31*, i221–i229. [CrossRef] [PubMed]

45. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

46. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; Volume 1, pp. 278–282.

47. Chen, L.; Fan, Z.; Chang, J.; Yang, R.; Hou, H.; Guo, H.; Zhang, Y.; Yang, T.; Zhou, C.; Sui, Q.; et al. Sequence-based drug design as a concept in computational drug design. *Nat. Commun.* **2023**, *14*, 4217. [CrossRef]

48. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36. [CrossRef]

49. Younesi, A.; Ansari, M.; Fazli, M.; Ejlali, A.; Shafique, M.; Henkel, J. A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends. *IEEE Access* **2024**, *12*, 41180–41218. [CrossRef]

50. Pan, X.; Ge, C.; Lu, R.; Song, S.; Chen, G.; Huang, Z.; Huang, G. On the integration of self-attention and convolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 815–825.

51. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.

52. Er, M.J.; Zhang, Y.; Wang, N.; Pratama, M. Attention pooling-based convolutional neural network for sentence modelling. *Inf. Sci.* **2016**, *373*, 388–403. [CrossRef]

53. Mao, A.; Mohri, M.; Zhong, Y. Cross-entropy loss functions: Theoretical analysis and applications. In Proceedings of the International Conference on Machine Learning, PMLR, Honolulu, HI, USA, 23–29 July 2023; pp. 23803–23828.

54. Demir-Kavuk, O.; Kamada, M.; Akutsu, T.; Knapp, E.W. Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features. *BMC Bioinform.* **2011**, *12*, 1–10. [CrossRef]