Retraction Notice

The Editor-in-Chief and the publisher have retracted this article, which was submitted as part of a guest-edited special section. An investigation uncovered evidence of compromised peer review and determined the paper is out of the scope of the journal. The Editor and publisher no longer have confidence in the results and conclusions of the article.

MS disagrees with the retraction. SD either did not respond or could not be reached.

DeepFNN-DTBA: prediction of drug-target binding affinity via feed-forward neural network on drug-protein sequences

Moolchand Sharma* **and Suman Deswal**
Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonipat, Haryana, India

Abstract. Pharmaceutical research essentially depends on drug-target interactions (DTIs). Standard methods of experimentation to uncover DTIs are costly and time-consuming, and thus artificial intelligence and machine learning have become popular. Multimodal imaging also provides significant amounts of anatomical, functional, and molecular information, accelerating drug discovery and development. Imaging technologies help understand disease mechanisms find new pharmacological targets and evaluate new drug candidates and how well they work. In this research, we developed a model based on deep learning (DL) that employs sequence information for targets and medicines to ascertain binding affinities of DTIs named feed-forward neural network (FNN)-DT binding affinity. Existing studies for the prediction of binding affinity of DTs either employ three-dimensional structures of protein–ligand complexes or two-dimensional characteristics of compounds. A novel technique used in this research: a dense network with dropouts were used to show the protein and drug sequences. These findings support the proposed DL-based approach for predicting binding affinity in DTIs, which use 1D representations of targets and medicines. In one of the standard datasets, the proposed FNNs outperformed the Kronecker regularized least squares, gradient boosting machines, deep drug target affinity algorithm, wide drug-target affinity, and similarity-based convolutional neural network model techniques with a 0.89 concordance index and 0.235 mean square error. © *2022 SPIE and IS&T* [DOI: 10.1117/1.JEI.32.5.052304]

Keywords: drug-target; drug-target interactions; deep learning; feed-forward neural network; machine learning; artificial intelligence; drug target affinity; drug target binding affinity; multimodal imaging.

Paper 220966SS received Sep. 15, 2022; accepted for publication Nov. 17, 2022; published online Dec. 6, 2022; retracted Sep. 13, 2023.

1 Introduction

Drug-target interaction (DTI) discovery is an emerging field of study because it is important for the creation of protein-targeted therapeutics and the identification of new drug candidates.[1–3] Targets are protein molecules, such as receptors, enzymes, etc., present in human cells that interact with pharmacological molecules.[4–6] Phenotypic changes are created by the drug's effect on the target's pharmacological aspects.[7,8] The DTIs must be understood by the pharmaceutical mechanism of action.[9] Hence, new target proteins are uncovered with the help of molecular medicine and genomic research. Sometimes drugs target incomplete proteins. In recent years, confirmation of interactions between drugs and their targets has been done using experimental methods. Few FDA-approved drug candidates have satisfactory activity because of toxicity.[10] The DTI testing is time-consuming and expensive.[11] To discover DTIs, new computational technologies are required.

Ligand and target-based strategies have been used to identify DTs.[12] Techniques that leverage ligand similarity are used to arrange the pharmacological properties and relationships between

*Address all correspondence to Moolchand Sharma, sharma.cs06@gmail.com

1017-9909/2022/$28.00 © 2022 SPIE and IS&T

target proteins. Keiser et al.[13] devised an approach to discover protein targets using chemical two-dimensional (2D) structural similarities. Protein domain information is not used in this method. The quality and integrity of target structure information are critical to target-based techniques.[14,15] The use of molecular docking and scoring functions is necessary for making predictions about DTIs. Molecular docking predicts the best way for a ligand to attach to a receptor to make a stable complex. A scoring function shows this complex's affinity. Docking methods cannot be used on protein targets with three-dimensional (3D) structures that are unknown. The ligand-based and target-based prediction methods use similarities in the structure of the ligand and the stereo structure of the target to predict the relationship between the ligand and the target.

Even though new therapeutic targets and ligands are being found, it is still hard to find active ligands among the millions of small molecule compounds. Machine learning (ML) and artificial intelligence (AI) must be used to predict how drugs will interact with their targets. With ML and AI, it is possible to predict DTIs before they happen. This ability to predict DTsI makes it easier to find treatments for the disorders being investigated. Based on the type of prediction challenge, DTI prediction algorithms can be classified into two groups. Binary classification is used to determine if a medicine interacts with a target protein. It is also possible to make predictions about the strength of DT binding using regression analysis (i.e., the binding affinity).

Affinity values for how well a protein binds to a ligand are very important, but the basic binary classification problem does not consider these values. K_d, K_i, and the half-maximal inhibitory concentration (IC50) are all ways to consider binding affinity, which is how well a drug interacts with its target (DT). Low IC50 values indicate strong binding. A low K_i means a robust binding affinity. Dissociation and inhibition constants as expressed by their negative logarithms, pK_d or pK_i, are commonly used to express the values of K_d and K_i.[16] Collecting data for DTI prediction research based on binary categorization is challenging because negative (nonbinding) information is uncommon. Negative (nonbinding) samples are those in which the DT pairings are unknown. System accuracy is hindered by the unavailability of actual negative samples that impair prediction systems. Using binding affinity scores instead of synthetic negative samples in DT prediction provides more realistic data sets and eliminates the need for artificial negative samples.

The deep drug target affinity (DeepDTA) model that was previously published has been improved in this study. We trained deep learning (DL) models on sequences (one-dimensional (1D) representations) to estimate drug-protein binding affinities. In place of the exterior characteristics or 3D structures of binding complexes, protein sequences and simplified molecular input line entry system (SMILES) representations of compounds are used. SMILES strings and protein sequences were fed into an explainable Substructure partition fingerprint (ESPF) encoder. From there, a feed-forward neural network (FNN) was trained on the representations. This dataset[17] was utilized to test our model's performance and compare our results with those of other models, such as DeepDTA,[18] the similarity-based convolutional neural network (CNN) model,[19] wide drug-target affinity (WideDTA),[20] gradient boosting machines (SimBoost),[21] and Kronecker regularized least squares (KronRLS).[22] Two different blocks are used to represent proteins and medications, which are further concatenated and then entered into a thick and strongly connected layer. This model outperforms the other techniques listed above on the Davis dataset. In addition, our suggested model obtains the lowest mean squared error (MSE). The following are the paper's primary contributions.

- The novel DeepFNN-DT binding affinity (DTBA) DTA prediction model using an FNN is developed to determine the binding affinity.
- The model performs best against the existing five state-of-the-art methods with a CI score of 0.895 and an MSE of 0.234.
- Advising the most optimal model with the most optimal encoding approaches for DT binding affinities, which can be groundbreaking in the development of novel drugs.

After an introductory section, the study is divided into three sections: a literature review in Sec. 2, a description of the current setup in Sec. 3, and the results of the research in Sec. 4. The next consists of datasets, evaluation parameters, and experimental setup. The

other section includes the experiments and results, followed by a discussion and a list of sources cited.

2 Literature Review

The effectiveness of a candidate ligand in the early stages of drug discovery for a therapeutic target is based on its affinity (e.g., a protein). There are $\sim 10^{60}$ synthetically accessible tiny molecules. Exploring this space is computationally infeasible. In spite of this, the combinatorial explosion draws attention to a major challenge in drug discovery: how to accurately assess the affinity of a large number of small molecules. There is a significant trade-off between time, money, and precision in both experimental and computational drug screening methods. In the pharmaceutical industry, ML is used to improve both the virtual drug screening and physics-based evaluations of small compounds. Over the past decade, DL has been highly effective in several AI research areas. The study of artificial neural networks (ANNs) led to the development of this method because they have been proven to be superior to existing ML algorithms in various fields, including those of image and speech recognition as well as natural language processing. In the last several years, DL has finally been used in the field of pharmaceutical research. The ability to foretell a compound's bioactivity has emerged as a crucial component of modern drug discovery.[23] Over the past two decades, there has been a tremendous rise in the amount of compound activity and biological data thanks to novel experimental techniques including high-throughput screening (HTS),[24,25] "parallel synthesis," and others.[26] There are new approaches for developing quantitative structure-activity relationship (QSAR) models, such as matrix factorization[27] and DL, that have been around for a long time but have recently been employed. DL has taken advantage of the ever-increasing amounts of data and the ever-growing computing capacity. The versatility of the NN architecture in DL sets it apart from most other ML methods. Recurrent neural networks (RNNs) and feed-forward CNNs will be explored in this course. In QSAR modeling, single-layer NNs have been utilized for a long time. Still, as data sizes and computational capacity have grown, multilayer feed-forward networks have become a logical choice for predicting bioactivity. These techniques, such as matrix factorization and DL, are used alongside established methods such as support vector machines (SVMs), neural networks, and random forests (RFs).[27] The volume of data and the ever-increasing processing power of computers have benefitted DL.

The DL's neural network architecture is different from other methods of teaching machines to learn because it can be changed. We consider fully connected (FC) FNNs, RNNs, and CNNs. Multilayer feed-forward networks are the best choice for predicting bioactivity because they have been used for a long time in QSAR modeling and because the amount of data and computing power keeps growing. CNNs are an excellent choice for processing biological images because they use high-throughput imaging technology. CNNs have gained significant success in the field of computer vision. The use of DL in drug discovery is advancing quickly, with new papers being published on the topic almost weekly. Dahl et al.[28] used a large number of 2D topological descriptors to solve the Merck Kaggle challenge dataset. The deep neural network (DNN) performed better than the traditional RF method on 13 out of 15 targets, according to the study's findings.

To summarize: (1) DNNs can handle tens of thousands of descriptors without feature selection; (2) dropout can avoid the overfitting problem that plagues traditional ANNs; (3) hyperparameter (number of layers, number of nodes per layer, type of activation functions, etc.) optimization can maximize DNN performance; and (4) DNN models that execute many tasks outperform those that perform a single task.[29,30] Mayr et al.[31] found that multitask DNN models won the Tox21 challenge on a dataset of 12,000 compounds for 12 high throughput toxicity experiments. Training of the model was done in parallel on GPU processors, the same as Dahl's design.[28,32] The DNN included a dropout network and a rectified linear unit (ReLU) activation function. During training, they used an extensive feature set containing static descriptors (3D, 2D descriptors, and predefined toxicophores) and dynamically produced extended connectivity finger-print (ECFP) descriptors to make DNN self-feature deduction easier. In each hidden layer, a statistical association analysis using only ECFP was carried out for DNN models, allowing the discovery of substructures closely associated with well-known toxicophores. A single-task DNN

Table 1 Comparison of existing approaches on the benchmark DAVIS dataset.

Author, year and references	Model used	DAVIS dataset	Encoding technique
		MSE and CI (concordance index)	Protein-drug
Ozturk et al.[18]	DeepDTA	0.420 and 0.886	S-W- CNN
Ozturk et. al.[18]	DeepDTA	0.261 and 0.878	CNN-CNN
Ozturk et. al.[18]	DeepDTA	0.419 and 0.835	CNN-PS
Ozturk et. al.[18]	DeepDTA	0.608 and 0.790	S-W-PS
Shim et al.[19]	SimCNN-DTA	0.3190 and 0.8501	S-W-S-W
Öztürk et al.[20]	WideDTA	0.262 and 0.886	PS-PDM
He et. al.[21]	SIMBOOST	0.282 and 0.872	S-W-PS
Pahikkala et. al.[22]	KronRLS	0.379 and 0.871	S-W-PS
Abbasi et al.[35]	DeepCDA	0.459 and 0.8396	S-W-PS

Note: S-W: Smith-Waterman, PS-Pubchem Sim,

and more traditional ML approaches were outperformed by a DNN that can perform several tasks at the same time.

In a comprehensive study, Ramsundar et al.[33] investigated the systematic development of multitasking DNNs and the effectiveness of single-task DNN models. Multitask models consistently outperformed single-task and RF models, according to their findings. DNN was tested against other well-known ML methods, such as SVMs, RFs, and many more, on seven datasets selected from ChEMBL by Koutsoukas et al.[34] Other than ML, DNN outperformed them all. A comparative analysis of existing approaches and acceptable encoding models for each method is shown in Table 1.

Table 1 displays standard databases for drugs and proteins (DAVIS datasets). It has been shown that current techniques, including encoding techniques, used by expert systems for sequence presentation learning more precisely estimate binding affinities for DTs.

3 Existing Framework

In the current framework, CNN, a DL architecture that is often used, is used. CNN is built with both convolutional layers and pooling layers. The pooling layer is used to reduce the number of samples and make the layer's results more general. Some layers are FC in the model. CNN models can take into account local dependencies using filters. The number and size of filters in a CNN are crucial to its ability to acquire new knowledge. The model's pattern-finding prowess improves with the size of its filter set.[18] To learn about the relationship between SMILES strings and protein sequences, two CNN blocks were used. Each CNN block comprises three 1D-convolutional layers with filters that get progressively better. In the second and third convolutional layers, the filters from the first layer were doubled and then tripled. After the convolutional layers came the max-pooling layers. The max-pooling features that came out of this were sent to three DeepDTA FC layers. First, there were 1024 nodes on two FC layers. Then, there were 0.1 dropout layers. Dropout regularizes neural activation to stop overfitting. The output layer came after the third layer, which had 512 nodes. The existing model, which combines two CNN blocks, is shown in Fig. 1.

4 Methodology

This section consists of the proposed architecture followed by datasets and methods used for evaluation against the different state-of-the-art methods.

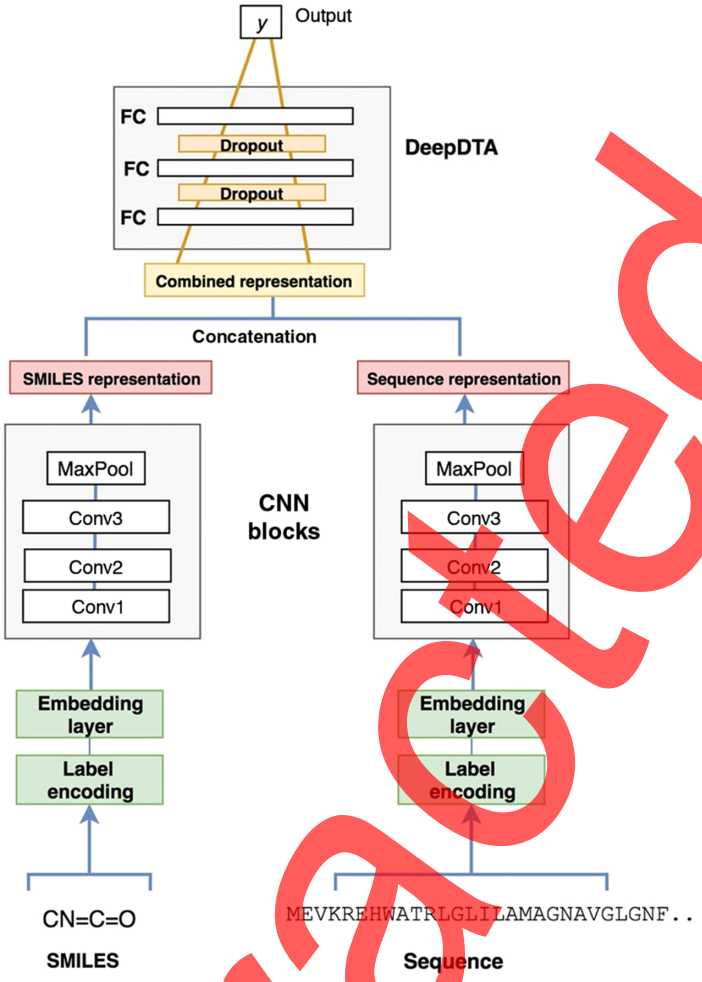

Fig. 1 Existing model with two CNN blocks for SMILES and protein sequences.[18]

4.1 *Proposed Model*

We used a regression analysis to predict protein–ligand interactions to estimate protein–ligand binding affinity. Here, FNNs were employed instead of CNNs to form the basis for our predictive model. The connections between nodes in an FNN do not create a cycle in an ANN. Thus, inputs are processed in a single direction, making the FNN model a basic version of a neural network that can be found. In this model, data move in a single direction and never backward despite passing through multiple hidden nodes. Figure 2 shows the architecture of an FNN.

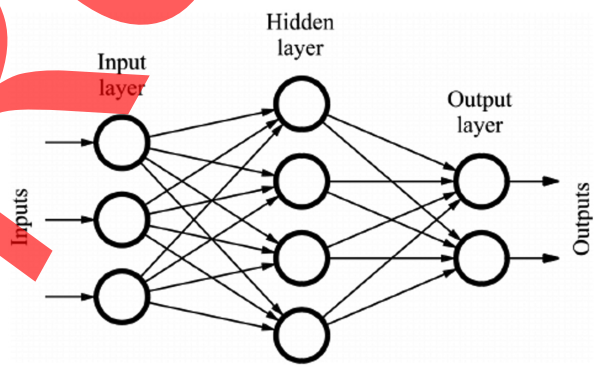

Fig. 2 Architecture of FNN.

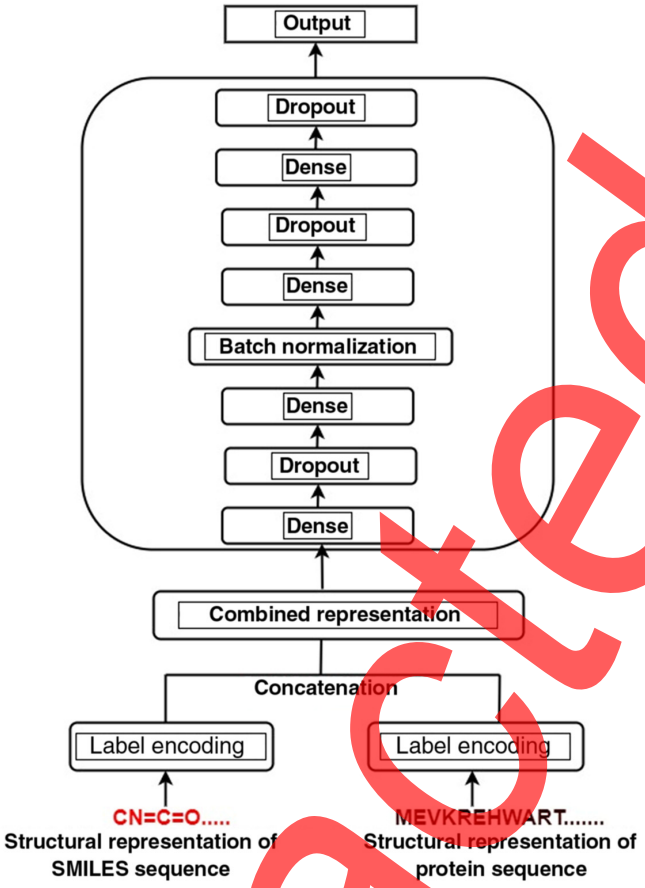

Fig. 3 DeepFNN-DTBA model for drug-protein affinity prediction.

To learn representations from SMILES strings and protein sequences, our proposed approach, DeepFNN-DTBA, employs an ESPF encoding technique. The dense neural network's final input dimension was determined by flattening and concatenating the data from both encoders (none, 6700). A dropout layer with a rate of 0.2 follows the first dense layer with 2048 neurons. To avoid overfitting, a dropout is a regularization approach in which certain neurons are set to zero. A batch normalization layer follows a linear activation function in the second dense layer, which has 1024 nodes. A dropout layer follows the linear activation function of 1024 neurons with a dropout rate of 0.1 in the third dense layer. 1024 neurons made up the fourth dense layer, followed by a dropout layer of 0.01. There are 512 neurons in the last dense layer and another dropout layer of 0.01 neurons before we get to the output layer. Figure 3 depicts the proposed model.

The activation function used in this investigation was the regularized sigmoid activation function, or ReLU, $g(x) = \max(0, x)$, which has seen extensive application in DL studies. It is the goal of any model that uses trial and error learning to achieve the best possible accuracy in its predictions. Because we are working on a regression problem, we chose MSE as the loss function, where f_i is the prediction vector and f_i' is the vector of actual outputs. The sample size is denoted by n. It is a common way to quantify uncertainty in continuous prediction, as shown in the following equation:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(f_i - f_i')^2. \tag{1}$$

The learning process was finished within 200 epochs, and a mini-batch size of 32 was utilized to update the network's weights. Adam was the optimization technique used in the training of the networks, and the default learning rate was 0.001. For the Davis data set, the input dimensional

Table 2 Parameter settings for DeepFNN-DTBA prediction.

Kayer (type)	Output shape	Parameters	Connected to
input_1 (InputLayer)	(None, 2586)	0	
input_2 (InputLayer)	(None, 4114)	0	
Concatenate (concatenate)	(None, 6700)	0	input_1 [0][0] input_2 [0][0]
Dense (dense)	(None, 2048)	13,723,648	concatenate [0][0]
Dropout (Dropout)	(None, 2048)	0	dense [0] [0]
dense_1 (Dense)	(None, 1024)	2,098,176	dropout [0] [0]
batch_normalization (BatchNormalization)	(None, 1024)	4096	dense_1 [0] [0]
dense_2 (Dense)	(None, 1024)	1,049,600	batch_normalization [0] [0]
dropout_1 (Dropout)	(None, 1024)	0	dense_2 [0] [0]
dense_3 (Dense)	(None, 512)	52,4800	dropout_1 [0] [0]
dropout_2 (Dropout)	(None, 512)	0	dense_3 [0] [0]
dense_4 (Dense)	(None, 1)	513	dropout_2 [0] [0]

Total parameters: 17,400,833.
Trainable parameters: 17,398,785.
Nontrainable parameters: 2048.

Table 3 Kinase dataset.

	Proteins	Compounds	Interactions
Davis (K_d)	442	68	30,056

matrix for the drug is (30056, 2586) and for a target is (30056, 4114). Therefore, each of the test findings in Table 4 employed the same substructure summarized in Table 2 to make the compound-protein affinity score prediction.

4.2 *Dataset*

Our proposed model was tested using the Kinase dataset Davis,[17] which had previously been used as standard data sets for evaluating the DTBA prediction models. The Davis dataset includes Kinase protein family selectivity experiments and relevant inhibitors' K_d values. Four hundred and forty-two proteins and 688 ligands are involved. Table 3 is a summary of the data used in our research.

For the Davis dataset,[36] the standard value is K_d in nM. We used the values transformed into log space, pK_d, similar to a prior source[37] as explained in the following equation:

$$pk_d \Leftrightarrow -\log 10\left(\frac{k_{\mathrm{d}}}{1^{\mathrm{e}^9}}\right). \tag{2}$$

The left panel of Fig. 4(a) shows how the binding affinity values are spread out in a pK_d form. More than half of the data (20,931 out of 30,056) is taken up by the peak at pK_d value 5(10,000 nM). So, pairs like these have weak binding affinities ($K_d > 10{,}000$ nM) and are not found in primary screening,[22] making them "true negatives." The PubChem compound identifier were used to get the compound SMILES strings from the PubChem compound database for the

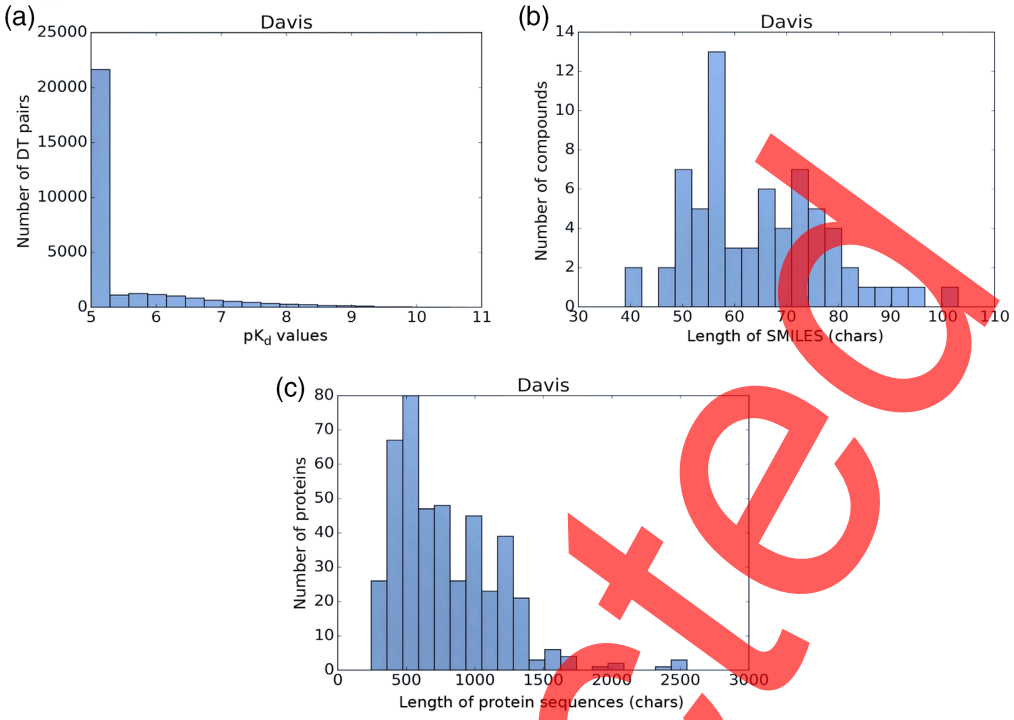

Fig. 4 Davis datasets are summarized in the left and right panels: (a) binding affinity values distribution, (b) the SMILES string length distribution, and (c) protein sequence length distribution.[18]

Davis dataset.[38] Figure 4(b) shows the length distribution of SMILES strings for the Davis compounds. SMILES compounds can be 103 or 64 bases long. The Davis protein sequences were retrieved from UniProt using gene names/RefSeq accession numbers. Figure 4(c) illustrates protein sequence lengths. The maximum protein sequence length is 2549 characters, and the average is 788.

An integer/label encoding was used to represent categories in the inputs. For both representations, the encoding technique/encoders used is ESPF. To figure out which functional groups of drugs cause a certain property, we need an encoding method that breaks drugs and proteins into a discrete set of medium-sized substructures customized to the data that we have and that have good predictive values. ESPF is based on the byte pair encoding algorithm and the subword units[35] in natural language processing. A database of sequences of entities is fed into ESPF (e.g., the amino acid sequence for the protein and SMILES for the drug). This database identifies and substitutes the original sequence (word) in a database with the most likely combination of subsequences (subword units). The output is the subsequences vocabulary set and their frequencies. With these two pieces of information, it can decompose any new unseen sequence into a sequence of frequent subsequences. This sequence can then be turned into a bit vector in which each bit corresponds to one item in the discovered subsequences set.

Empirically, we find ESPF outputs a suitably-sized substructure ordered partition. It successfully identifies essential functional groups for drugs and motifs for proteins. The suitably-sized nonoverlapping outputs provide a tractable path to see which substructures contribute to ML predictive outcomes. SMILES and protein sequences were capped at 85 and 1200 bytes, respectively, so that Davis could create an effective representation form. Proteins might have up to 1000 characters in their sequences, whereas SMILES were limited to 100. We selected these extreme lengths to cover at least 80% of proteins and 90% of compounds based on the trends in Figs. 4(b) and 4(c). Shorter sequences are padded with zeros, and longer ones are trimmed.

4.3 *Performance Evaluation Parameters*

DTIs are not binary. Therefore, the CI is used to make predictions about the model's accuracy. Furthermore, the order of predictions for randomly selected drug target pairings is represented

formally by the correlation coefficient.[22] This indicates that the y_i prediction for the more significant affinity value is higher than that for f_i, the less significant y_j, as demonstrated in the following equation:

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

The step function is denoted by $h(u)$, which provides the values 1.0, 0.5, and 0.0 depending on whether u is greater than or equal to 0. These values are set as $u > 0$, $u = 0$, and $u < 0$, respectively. Z is a normalization constant equal to the number of data pairings with values for the label that are distinct from one another. The concordance index values can range from 0.5 to 1.0, as shown in equation 4, with 0.5 indicating a random predictor and 1.0 indicating total prediction accuracy based on the test data. The values of the CI range from 0.5 to 1.0. The step function is given as

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

Another metric used and explained in Sec. 4.1 is the MSE.

4.4 *Experimental Setup*

As can be seen in Table 2, the learned parameters were used to train the model, which then provided a more accurate performance measure on the independent test set. We have done preprocessing with the DeepPurpose package and modeling in Tensorflow. Our tests were conducted on a Windows 11 machine equipped with an Intel Core i7-11800H processor running at 2.3 GHz and an NVIDIA GeForce RTX 3050 Ti graphics card (4 GB). Using GPUs in tandem with cuDNN allowed the task to be done faster.

5 Experiment and Results

This research proposes a DL model for learning representations for medications and targets based on their sequences. This model employs an FNN combining dense and dropout layers to learn drug and target representations. DeepDTA, WideDTA, the similarity-based CNN model, the KronRLS algorithm, and SimBoost are all comparative analysis techniques that take input similarity matrices for proteins and compounds. These techniques were used to establish a baseline for further examination. The approaches mentioned above used the Smith-Waterman (S-W), Pubchem Sim (PS), and CNN algorithms to compute the pairwise similarities of the proteins and ligands. After that, we evaluated the model by feeding ESPF and ESPF as encoding strategies into our FNN model to see how well it performed. In the CNN DeepDTA model, the amount of time that passed since the beginning of one epoch was 29 s, but in the proposed FNN model, the amount of time that passed since the beginning of one epoch was just 4 s. As a result, we also accomplished a score of 0.895 for the CI and a score of 0.234 for the MSE. Table 4 present the average MSE and CI scores obtained from the independent test set for each of the five models trained on the Davis dataset using the parameters given in Table 2. The comparison analysis with various state-of-the-art models is depicted in Fig. 5, along with the proposed model.

Hence, in the proposed model, the combination of dense and dropout layers worked successfully to find the bonding nature between the drug and its target. The dense layer is the layer that is most deeply related to the layer that comes before it, and it is responsible for working to change the dimension of the output by performing matrix-vector multiplication. The dropout layer is a mask that, when applied, prevents specific neurons from contributing information to the layer below it while leaving all other neurons unaltered. We can use a dropout layer in the input vector to nullify some of the features of the input vector. Alternatively, we can apply it to a hidden layer, which will nullify some of the neurons buried behind the hidden layer. Dropout layers are

Table 4 The average CI and MSE scores for the Davis dataset with the existing and proposed models.

Models	Proteins	Compounds	CI and MSE
DeepDTA[18]	S-W	CNN	0.886
			0.420
DeepDTA[18]	CNN	CNN	0.878
			0.261
DeepDTA[18]	CNN	PS	0.835
			0.419
DeepDTA[18]	S-W	PS	0.790
			0.608
SimCNN-DTA[19]	S-W	S-W	0.850
			0.319
WideDTA[20]	PS	PDM	0.886
			0.262
SimBoost[21]	S-W	PS	0.872
			0.282
KronRLS[22]	S-W	PS	0.871
			0.379
DeepCDA[35]	S-W	PS	0.839
			0.459
Proposed DeepFNN-DTBA model	ESPF	ESPF	**0.895**
			0.234

Note: S-W: Smith-Waterman, PS-Pubchem Sim, and PDM-Protein Domains and Motifs.

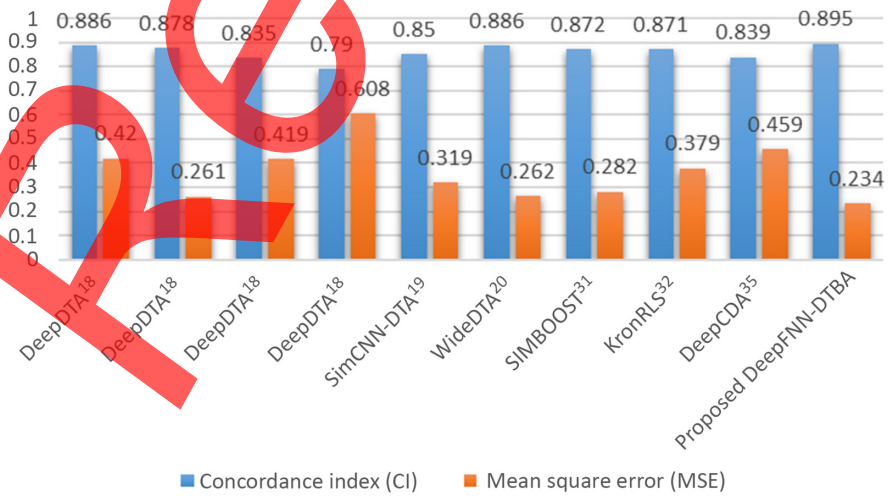

Fig. 5 Comparative analysis with the state-of-the-art models showing CI score and MSE score.

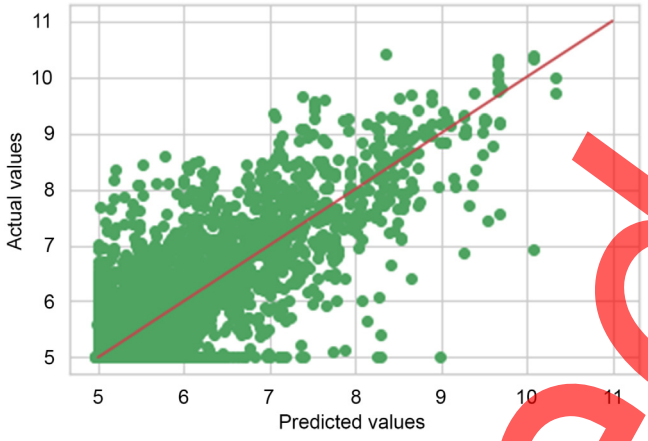

Fig. 6 Comparison between predicted and actual DTBA values.

significant in training because they prevent the training data from overfitting the model. To better trust the model's anticipated values, we compare them with the actual measurements of affinity acquired from the DAVIS datasets. Figure 6 shows that there is a strong relationship between the projected ESPF-ESPF target affinity score and the measured ESPF-ESPF target affinity score from the Davis dataset. This is why we count on the ideal predictive model to yield estimates (p) that are just as precise as the true values (y). The high density around the $x = y$ line in Fig. 6 indicates that our model performs far better than the current ones.

To circumvent the computational impossibility of searching the entire molecular space (10^{60}) for drug repurposing, we can use DL models to narrow the search space or explore the space. Using AI, ML, and DL, among other computational methods, improved the efficiency and success of drug repositioning. Due to its potential applications in areas such as target prediction and drug repositioning, DL has recently attracted much attention. Undoubtedly, clinical laboratories will save time and money using computational methods, and molecular space can be reduced.

6 Conclusion and Future Scope

This study aimed to provide evidence that DL can accurately predict whether or not a drug will successfully attach to a target protein by examining both sequences. The FNNs are used in affinity prediction tasks based on drug and target protein sequence representations. The Davis kinase drug datasets are utilized for the simulation process. Therefore, the performance of the FNN model on the Davis dataset is significantly improved, and the combination of the dense layer with dropout is used. It was observed that DL techniques produced superior results when compared with the models that were already in use. As the datasets grow, this information's significance will also increase. The CI score went from 0.886 to 0.895, which is a substantial improvement, whereas the MSE went from 0.261 to 0.234, which is a significant decrease. Through the use of DL architecture, it is possible to find hidden or embedded information more efficiently.

The most significant finding of this study is the development of a DL-based model capable of predicting a drug's effectiveness against a given target using just character representations of proteins and compounds. In addition, this paper recommends the best strategy for encoding compounds and proteins above the alternative standard approaches. Many proteins have not been targeted because drug development focuses on a small group of proteins or because they cannot be drugged. Our research is focused on finding the best way to train a deep-learning model to identify potential therapeutic targets. Successful models can be used directly in the evaluation process or alternative hyperparameter combinations can be tried. Predicting drug-target affinities using a trustworthy hybrid model is equally challenging. The optimal values for a model's hyperparameters can be found with the help of a bioinspired algorithm, which can then be used in the

model's development and refinement. Imaging also plays an important role in the research and discovery of drugs by giving data that may be used to locate prospective drug targets or monitor the effects of drugs already in use. The current state-of-the-art in this field is constantly improving and becoming more refined and powerful.

Acknowledgments

The authors declare that there are no conflicts of interest.

References

1. M. A. Yildirim et al., "Drug-target network," *Nat. Biotechnol.* **25**(10), 1119–1126 (2007).
2. S. C. Janga and A. Tzakos, "Structure and organization of drug-target networks: insights from genomic approaches for drug discovery," *Mol. Biosyst.* **5**, 1536–1548 (2009).
3. M. Kuhn et al., "Large-scale prediction of drug-target relationships," *FEBS Lett.* **582**(8), 1283–1290 (2008).
4. Y. C. Wang et al., "Computationally probing drug-protein interactions via support vector machine," *Lett. Drug Des. Discov.* **7**, 370–378 (2010).
5. Z. Xia et al., "Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces," *BMC Syst. Biol.* **4**, S6 (2010).
6. M. R. Dreher et al., "Tumor vascular permeability, accumulation, and penetration of macromolecular drug carriers," *J. Natl. Cancer Inst.* **98**(5), 335–344 (2006).
7. Y. Tabei et al., "Identification of chemogenomic features from drug-target interaction networks using interpretable classifiers," *Bioinformatics* **28**(18), i487–i494 (2012).
8. M. Hao, Y. Wang, and S. H. Bryant, "Improved prediction of drug-target interactions using regularized least squares integrating with kernel fusion technique," *Anal. Chim. Acta* **909**, 41–50 (2016).
9. M. Takarabe et al., "Drug target prediction using adverse event report systems: a pharmacogenomics approach," *Bioinformatics* **28**(18), i611–i618 (2012).
10. J. C. Dearden, "In silico prediction of drug toxicity," *J. Comput. Aided Mol. Des.* **17**(2–4), 119–127 (2003).
11. H. Van de Waterbeemd and E. Gifford, "ADMET in silico modelling: towards prediction paradise?" *Nat. Rev. Drug Discov.* **2**(3), 192–204 (2003).
12. Y. F. Dai and X. M. Zhao, "A survey on the computational approaches to identify drug targets in the postgenomic era," *Biomed. Res. Int.* **2015**, 239654 (2015).
13. M. J. Keiser et al., "Relating protein pharmacology by ligand chemistry," *Nat. Biotechnol.* **25**(2), 197–206 (2007).
14. B. K. Shoichet, "Virtual screening of chemical libraries," *Nature* **432**(7019), 862–865 (2014).
15. G. Klebe, "Virtual ligand screening: strategies, perspectives, and limitations," *Drug Discov. Today* **11**(13–14), 580–594 (2006).
16. R. Z. Cer et al., "IC 50-to-k i: a web-based tool for converting ic 50 to k i values for inhibitors of enzyme activity and ligand binding," *Nucl. Acids Res.* **37**(suppl. 2), W441–W445 (2009).
17. M. I. Davis et al., "Comprehensive analysis of kinase inhibitor selectivity," *Nat. Biotechnol.* **29**(11), 1046–1051 (2011).
18. H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: deep drug-target binding affinity prediction," *Bioinformatics* **34**(17), i821–i829 (2018).
19. J. Shim et al., "Prediction of drug-target binding affinity using similarity-based convolutional neural network," *Sci. Rep.* **11**, 4416 (2021).
20. H. Öztürk, E. Ozkirimli, and A. Özgür, "WideDTA: prediction of drug-target binding affinity," arXiv:1902.04166 (2019).
21. T. He et al., "SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines," *J. Cheminform* **9**(1), 24 (2017).

22. T. Pahikkala et al., "Toward more realistic drug-target interaction predictions," *Brief Bioinf.* **16**, 328–337 (2014).
23. H. Chen et al., "The rise of deep learning in drug discovery," *Drug Discov. Today* **23**, 1241–1250 (2018).
24. G. Papadatos et al., "Activity, assay, and target data curation and quality in the ChEMBL database," *J. Comput. Aided Mol. Des.* **29**, 885–896 (2015).
25. S. Kim et al., "PubChem substance and compound databases," *Nucl. Acids Res* **44**(D1), D1202–D1213 (2016).
26. M. K. Gilson et al., "BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology," *Nucl. Acids Res* **44**(D1), D1045–D1053 (2016).
27. M. Ammad-ud-Din et al., "Drug response prediction by inferring pathway-response associations with kernelized Bayesian matrix factorization," *Bioinformatics* **32**(17), i455–i463 (2016).
28. G. E. Dahl et al., "Multi-task neural networks for QSAR predictions," arXiv:1406.1231 (2014).
29. S. Perumal et al., "ANN-based novel approach to detect node failure in wireless sensor network," *Comput. Mater. Contin.* **69**(2), 1447–1462 (2021).
30. M. Tabassum et al., "Enhance data availability and network consistency using artificial neural network for IoT," *Multimedia Tools Appl.* **81**, 1–21 (2022).
31. A. Mayr et al., "DeepTox: toxicity prediction using deep learning," *Front. Environ. Sci.* **3**(PP), 80 (2016).
32. J. Ma et al., "Deep neural nets as a method for quantitative structure-activity relationships," *J. Chem. Inf. Model.* **55**, 263–274 (2015).
33. B. Ramsundar et al., "Is multitask deep learning practical for pharma?" *J. Chem. Inf. Model.* **57**, 2068–2076 (2017).
34. A. Koutsoukas et al., "Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data," *J. Cheminf.* **9**, 42 (2017).
35. K. Abbasi et al., "DeepCDA: deep cross-domain compound-protein affinity prediction through LSTM and convolutional neural networks," *Bioinformatics* **36**, 4633–4642 (2020).
36. http://staff.cs.utu.fi/~aatapa/data/DrugTarget/.
37. R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," in *ACL 2015* (2015).
38. E. E. Bolton et al., "Pubchem: integrated platform of small molecules and biological activities," *Annu. Rep. Comput. Chem.* **4**, 217–241 (2008).

Moolchand Sharma is an assistant professor in the Department of Computer Science and Engineering at Maharaja Agrasen Institute of Technology, GGSIPU, Delhi, India. Currently, he is a doctoral researcher at DCR University of Science and Technology, Haryana. He has published scientific research studies in international journals and conferences, including SCI-indexed and scopus indexed journals, such as *Cognitive Systems Research* (Elsevier), *Physical Communication* (Elsevier), *Intelligent Decision Technologies: An International Journal*, and many more.

Suman Deswal received her BTech degree in computer science and engineering from CR State College of Engineering, Murthal, Haryana, India, and her MTech (CSE) degree from Kurukshetra University, Kurukshetra, Haryana, India, in 1998 and 2009, respectively. She received her PhD from DCR University of Science and Technology, Murthal, Haryana, India. She possesses 18 years of teaching experience and is currently a professor in the Department of Computer Science and Engineering at DCRUST, Murthal, Haryana, India