Deep Learning-Based Drug–Protein Binding Affinity Prediction Using Sequence Information

Dr.Thilakavathy.P
Department of Computer Science and Engineering
Vels Institute of Science, Technology & Advanced Studies(VISTAS)
Chennai, India
thilakavathy.se@vistas.ac.in

Abina.R
Department of Computer Science and Engineering
Vels Institute of Science, Technology & Advanced Studies(VISTAS)
Chennai, India
abinar81@gmail.com

Ananya.S
Department of Computer Science and Engineering
Vels Institute of Science, Technology & Advanced Studies(VISTAS)
Chennai, India
ananyashri2005@gmail.com

Dr.S. Arun
Department of Computer Science and Engineering
Vels Institute of Science, Technology & Advanced Studies(VISTAS)
Chennai, India
arun.se@vistas.ac.in

***Abstract*— Drug-protein interaction plays a vital role in drug discovery, as the interaction between drug and protein, i.e., binding affinity, plays a key role in drug efficacy and safety. The binding affinity between drug and target protein can be represented as a dissociation constant (Kd). Assessing binding affinity often requires considerable time, money, and other resources, as experimental studies are required to compute binding affinity between drug-target proteins. In that context, this paper discusses a comprehensive approach to utilize a deep drug-target interaction (DTI) approach that can be utilized to predict drug-protein binding affinity, utilizing a deep drug-target interaction approach, considering that drug molecules can be represented as a sequence of SMILES annotation, and similarly, target proteins are represented as a sequence of FASTA annotation. Experiments are conducted utilizing a Davis dataset, considering suitable preprocessing steps, i.e., tokenization, normalization, and division of dataset into train-test sets. The efficacy of the proposed model is assessed in accordance with various regression-based performance metrics such as Mean Square Error (MSE), Mean Absolute Error (MAE), Concordance Index (CI), Pearson correlation coefficient, and R^2 score. The experiments show that the model performs well in generating accurate binding affinity predictions and has substantial applications in computer-aided drug screening.**

Keywords*— *Drug–Protein Interaction, Binding Affinity Prediction, Deep Learning, Machine Learning, Drug Discovery, Davis Dataset.

I. INTRODUCTION

Drug-protein interaction provides the basis upon which biological processes are conducted in the human body. The use of therapeutic drugs in biological systems has a critical dependency upon drug-protein interaction in their development process as well. The efficacy of a drug has primarily been determined by how well it binds to a specific protein targeted in disease pathways. The dissociation constant Kd represents how well this binding has occurred in numbers or factors – less Kd factors define how well a drug binds to a targeted site in relation to its potency.

Normally, binding affinity is measured in drug discovery using experimental approaches in the lab, such as biochemical experiments. While this approach may be the most reliable in providing precise outcomes, it demands precious time, cost, and expertise .Experimental evaluation of all potential drug compounds is not practical in early drug screening with the increasing number of drug compounds in drug discovery.

However, recent developments and achievements in the fields of machine learning have allowed various approaches to use these learning mechanisms to model biochemistry interactions. In this way, machine learning is effective to recognize various complex nonlinear interactions through learning patterns established from usable biochemical dataset sources. In this case, deep learning approaches to Drug-Target Interaction (DTI) models have proven to have remarkable ability to recognize and extract features from raw representations such as SMILE strings and amino acid sequences.

The use of these methods of machine learning for drug–protein interaction analysis open a promising pathway for virtual drug drug development through increasing efficiency and facilitating efficient decisions during early drug development stages.

A. Problem Statement

Accurately predicting how strongly a drug interacts with its target protein is very important in the drug discovery process, as it helps researchers identify drugs that are both effective and safe. Current standard methods for determining affinity involve biochemical assays. The existing computational methods, such as molecule docking and machine learning methods, mostly rely upon hand-crafted information or detailed knowledge of the structure, which might not always be possible to attain or include in a model. In addition to this, it is hard to determine the complex nonlinear relationships within drug molecules and their sequences in proteins.Therefore, there exists the need for an efficient computation model that can predict drug-protein binding affinity accurately using the sequence representation of drugs and protein pairs directly in an efficient manner.

B. Motivation

The increasing complexity of drug discovery and the rapidly improving availability of biological data create a strong demand for efficient computation functions to analyze drug-protein interactions. Binding affinity, which decides the effectiveness of drugs, cannot be experimentally measured for thousands of drug-protein pairs without much time, cost, and resource wastage. Recent breakthroughs in machine learning, in particular deep learning-based DTI models, hold much promise for directly learning complex patterns from the representations of drug molecules and protein sequences. Empowered by benchmark datasets such as the Davis dataset, these models have enabled the prediction of binding affinity with high accuracy in a scalable way, hence acting as important tools for virtual screening and early-stage drug design.

C. Objective

The objective of this research is to design an efficient model of drug target interactions based on the application of the overall concept of deep learning in machines for the virtual prediction of drugs with a high affinity of binding to substances or proteins by using sequence-based data. This model would be trained by the application of the Davis data set related to the SMILES strings of drugs and the amino acid sequences of target proteins for the efficient recognition of learning characteristics of drug interactions with proteins. This application of efficient nonlinear relationships for the recognition of proteins by drugs would make this model aimed to provide efficient predictions regarding binding with high accuracy without requiring any knowledge of the three-dimensional structures of drugs or proteins. Moreover, there would also be overall evaluation of this model on the basis of various performance measures for effective predictions in relationship with its overall robustness by the application of Mean Squared Error, Mean Absolute Error, Concordance Index, Pearson correlation coefficient, R^2 value, etc.

II. EXISTING SYSTEM

Traditionally, the prediction of drug-protein binding affinity in drug discovery pipelines is done by experimental techniques such as biochemical assays and laboratory testing. These experimental techniques give results that are accurate and reliable but require huge time and cost along with laboratory infrastructure. Owing to these reasons, a large number of drug-protein combinations can hardly be evaluated during preliminary stages of drug screening.Computational methods such as molecular docking and simulation-based techniques supplement the experimental methods to estimate binding interactions. Methods like these reduce experimental effort but often need detailed three-dimensional structural information and involve high computational complexity. Moreover, their performance depends very much on the quality of input structures and parameter settings. Therefore, the present system suffers due to a lack of scalability and efficiency, becoming unrealistic for large-scale and fast drug-protein interaction analysis.

A. Research Paper and Studies

Various studies proposed numerous computational methods for determining drug-protein binding affinity, intending to minimize experimental procedures. Initially, various studies were performed based only on molecular docking and simulation techniques, considering various aspects of determining interaction affinity. Later, along with machine learning, regression-based techniques were also employed for accurate estimation of binding affinity considering various aspects of molecular

behavior, especially based only on handcrafted molecular descriptor-based techniques, which resulted in efficiency in addition to accuracy, as well as the ability of the model in handling complex non-linearity.Recently, various models based on deep learning within the Drug–Target Interaction dataset attracted a lot of attention based on their capacity to learn and extract features from data directly from a dataset in terms of feature presentation in different forms, including SMILES and Amino Acid sequences. The results from models based on a combination of neural networks obtained very high results on existing standard datasets, including the Davis dataset. They are based on a direct approach to bind proteins with a certain bound affinity level. However, there are still plenty of scope and areas to be covered in developing efficient and deep learning models. Therefore, there are plenty of areas to be covered in deep learning models.

III. EXISTING SYSTEM ANALYSIS

The existing system in analyzing the affinity of drug–protein binding utilizes a combination of experimental and computational methodology, usually adopted in any drug discovery study. The experimental study in analyzing and understanding the affinity has relied on biochemical assays; these are usually very accurate and safe in validating any results obtained in assessing and understanding a drug study. Apart from experimental methodology, various models in computer simulation, including docking and machine learning models, are usually applied in analyzing and understanding a systematic study of a drug–protein interaction by predicting various physicochemical properties and past trends in interaction.

A. Limitations

Despite the significant contribution these models have made towards drug discovery, there are striking limitations to the majority of the current drug-protein binding affinity prediction systems. While experimental approaches are highly accurate, they are extremely time-consuming, expensive, and require specialized laboratory infrastructure, hence unsuitable for large-scale or high-throughput screening. Computational techniques such as docking and traditional machine learning models decrease the experimental burden. However, most of the current methods require either three-dimensional structural data in great detail or manually engineered features, neither of which is consistently available across all the drug-protein pairs. Moreover, such approaches can also be computationally very costly and may also fail to model the complex nonlinear interactions between drugs and proteins. Because of this, there is a clear need for more efficient, scalable, robust computational models that provide accurate binding affinity predictions directly from sequence-based information.

B. Proposed Improvement

In order to mitigate these challenges faced by existing systems, this proposed model employs a deep model of Drug-Target Interaction that is capable of predicting these interactions. In this model, binding affinity is directly estimated without involving either laboratory experiments or structural information. This is done through various representations such as SMILES and FASTA. Even with these representations, there is no guarantee that accurate predictions will be generated. In order to mitigate this problem, deep learning is used to improve these predictions. In this model, regression metrics such as Mean Squared Errors, Mean Absolute Errors, Concordance Index, Correlation Coefficient, and R^2 Score are used to evaluate model performance. This model is developed to bring efficiency to virtual screening and early-stage drug discovery.

IV. PROPOSED SYSTEM

The proposed system uses a deep Drug–Target Interaction (DTI) framework to estimate the dissociation constant (Kd) and predict drug–protein binding affinity. To discover intricate nonlinear interaction patterns, the model combines biological sequence data from protein targets with chemical structural data from drug molecules. The Davis Dataset, which includes continuous Kd values and drug–target interaction pairs that have been experimentally validated, is used to train and assess the system. In order to support computational drug discovery and minimize experimental screening efforts, the proposed framework aims to accurately predict the binding strength between a drug compound and a disease-related protein.

A. System Architecture

The architecture of the proposed system adheres to a dual-input deep learning framework mean to concurrently process data of drug and protein . The input takes drug molecules in SMILES format and protein sequences in FASTA format. During preprocessing, entries which are invalid or are incomplete are removed, and both SMILES and FASTA sequences are broken down into meaningful units. To make sure that all sequences have dimensions with same number, they are either padded or cut to fixed lengths. Each token is assigned with unique integer index and then passed through embedding layers for creating dense numerical vector representations.

To normalize variance and make regression work better, a logarithmic transformation helps to change Kd values from one order of magnitude to another (pKd = -\log_{10}(Kd)). Deep learning layers process the embedded drug sequences to get structural and chemical features, and to get biologically important sequence patterns and motifs, similar layers process the embedded protein sequences. The drug and protein feature vectors are combined to create a single interaction representation. To model complicated drug-protein interactions, this fused vector goes through various fully connected layers with nonlinear activation functions. Dropout regularization is used to improve generalization and reduce overfitting. Regression output layer is the last layer, uses the Mean Squared Error (MSE) loss function and the Adam optimizer for predicting about the continuous pKd value.

B. Key Features

- End-to-end deep learning framework is used by proposed system, that automatically learns representations from SMILES and FASTA sequences. Manual feature engineering is no needed.
- A dual-branch architecture processes drug and protein inputs separately, helping it easier to obtain chemical and biological features.
- Trainable embedding layers turn separate tokens into dense vector representations, which helps with context.
- Using (pKd = -\log_{10}(Kd)) to change affinity values into logarithmic form makes the numbers more stable and the regression more accurate.
- A feature fusion mechanism merges drug and protein feature vectors to make nonlinear interaction relationships model.
- The framework does continuous regression-based affinity prediction, which gives an accurate value of the binding strength as a numbers.
- The model is made more robust and better at generalizing by dropout regularization.
- The architecture can be changed to work with large-scale drug discovery applications.

C. Workflow

The data obtaining from the dataset, including SMILES strings, FASTA sequences, and corresponding Kd values, is the first step in the workflow of the proposed system. Preprocessing of the input data includes logarithmic transformation of affinity values, integer encoding, tokenization, and padding or truncation. By passing the processed sequences through embedding layers dense vector representations are obtained. Deep learning layers extract high-level structural characteristics from drug sequences and biologically significant patterns from protein sequences. For regression-based pKd value prediction, the extracted feature vectors are combined to create a unified interaction representation, which is subsequently given into fully connected layers. Binding strength is calculated by interpreting the predicted pKd values; higher Kd denotes weak interaction and lower Kd (higher pKd) indicates strong binding affinity. Potential therapeutic drug-target interactions are identified and ranked through the final output.

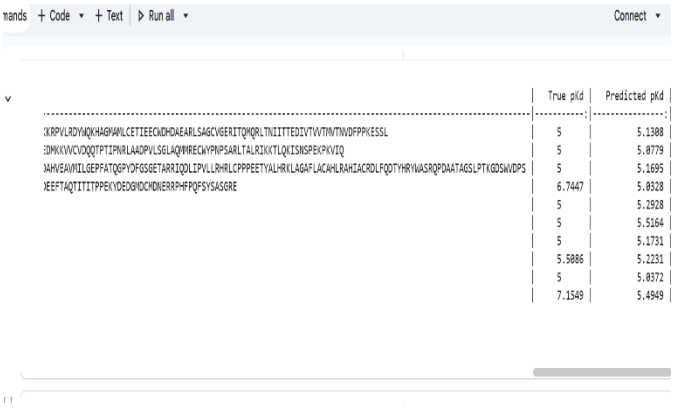
V.IMPLEMENTATION DETAILS

The proposed system is implemented using Python with deep learning libraries like TensorFlow and PyTorch. Davis dataset is used as input, including SMILES as string of drug, FASTA as sequence of protein, and experimentally measured binding affinity value Kd. Data preprocessing is used for tokenization of sequences, normalization of binding affinity values, and padding is used to ensure that all the inputs to the model have similar lengths.

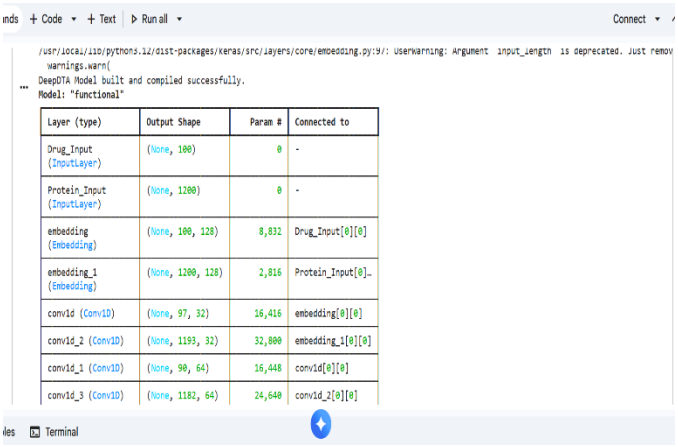
The core model is a deep DTI architecture which uses separate encoders for drugs and proteins. The representations are fused, modeling interactions, and theregression layer predicts continuous binding affinity values. The datasetis split into training, validation, and test sets; the model is then trained using gradient-descent-based optimization with Mean Squared Error loss. Some of the regularization techniques includes withdrawal and early stopping to prevent overfitting. The model will then be assessed with MSE, MAE, Concordance Index, Pearson correlation, and R^2 score which provides an end-to-end pipeline for the efficient and scalable prediction of binding affinity.

VI. OUTPUT SCREENSHOTS

This proposed system identifies continuous binding affinity values Kd, for each and every drug protein pair in the dataset. The output gives a numerical value of the strength of interaction. Lower values of Kd represent stronger binding. Predicted Kd values are created for all test samples. These are then compared with experimentally measured values to obtain accuracy. Scatter plots of predicted against true values ensures evaluation of correlation and error patterns. Performances include the Mean Squared Error (MSE), Mean Absolute Error (MAE), Concordance Index (CI), Pearson correlation coefficient, and R^2 score that altogether provide an insight into the effectiveness of the model. The outputs can enable prioritization of promising drug candidates for further experimental validation or virtual screening at the initial stages of drug discovery.



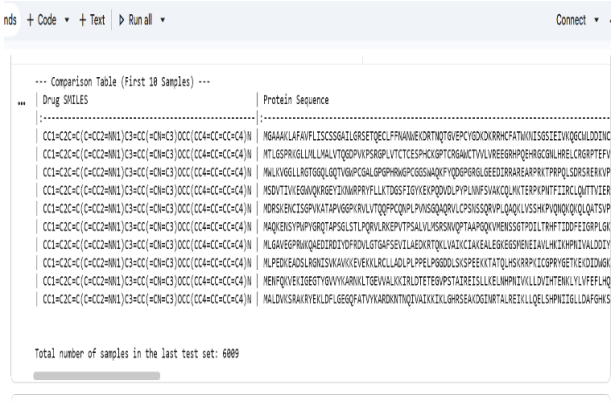
	True pKd	Predicted pKd
CKRPVLRDYNQKHAGMAMLCETIEECWDHDAEARLSAGCVGERITQMQRLTNIITTEDIVTVVTMVTNVDFPPKESSL	5	5.1308
EDMKKVVCVDQQTPTIPNRLAADPVLSGLAQWNRECWYPNPSARLTALRIKKTLQKISNSPEKPKVIQ	5	5.0779
DAHVEAVMILGEPFATQGPYDFGSGETARRIQQLIPVLLRHRLCPPPEETYALHRKLAGAFLACAHLRAHIACRDLFQDTYHRYWASRQPDAATAGSLPTKGDSWDPS	5	5.1695
KEEFTAQTITITPPEKVDEDGHDCMDNERRPHFPQFSYSASGRE	6.7447	5.0328
	5	5.2928
	5	5.5164
	5	5.1731
	5.5086	5.2231
	5	5.0372
	7.1549	5.4949



/usr/local/lib/python3.12/dist-packages/keras/src/layers/core/embedding.py:97: UserWarning: Argument `input_length` is deprecated. Just remov
warnings.warn(
DeepDTA Model built and compiled successfully.
Model: "functional"

Layer (type)	Output Shape	Param #	Connected to
Drug_Input (InputLayer)	(None, 100)	0	-
Protein_Input (InputLayer)	(None, 1200)	0	-
embedding (Embedding)	(None, 100, 128)	8,832	Drug_Input[0][0]
embedding_1 (Embedding)	(None, 1200, 128)	2,816	Protein_Input[0]…
conv1d (Conv1D)	(None, 97, 32)	16,416	embedding[0][0]
conv1d_2 (Conv1D)	(None, 1193, 32)	32,800	embedding_1[0][0]
conv1d_1 (Conv1D)	(None, 98, 64)	16,448	conv1d[0][0]
conv1d_3 (Conv1D)	(None, 1182, 64)	24,640	conv1d_2[0][0]

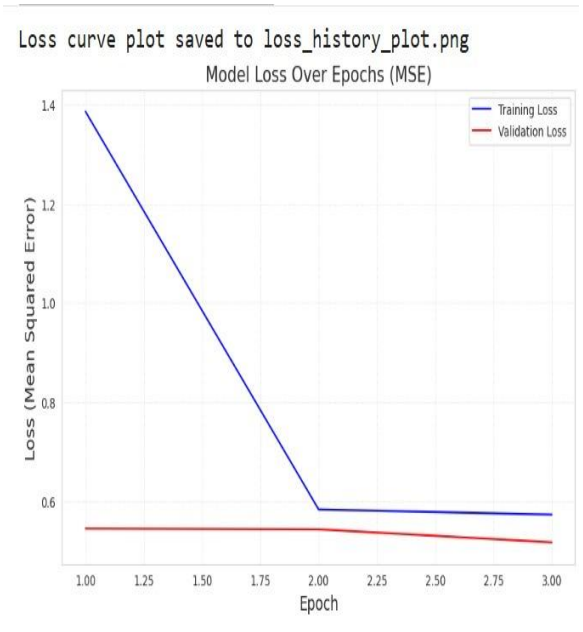
Terminal

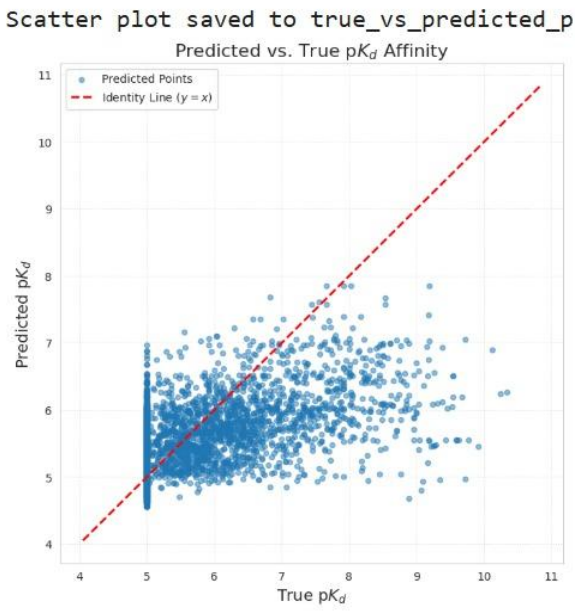


--- Comparison Table (First 10 Samples) ---

Drug SMILES	Protein Sequence
CC1=C2C=C(C=CC2=NN1)C3=CC(=CN=C3)OCC(CC4=CC=CC=C4)N	MGAAAKLAFAVFLISCSSGAILGRSETQECLFPNAMWEKQRTNQTGVEPCYGDKDKRRHCFATWKNISGSIEIVKQGCWLDDINC
CC1=C2C=C(C=CC2=NN1)C3=CC(=CN=C3)OCC(CC4=CC=CC=C4)N	MTLGSPRKGLLMLMALVTQGDPVKPSRGPLVTCTCESPHCKGPTCRGAWCTVVLVREEGRHPQEHRGCGHLHRELCRGRPTEFV
CC1=C2C=C(C=CC2=NN1)C3=CC(=CN=C3)OCC(CC4=CC=CC=C4)N	MWLKVGGLLRGTGGQLGQTVGWPCGALGPGPHRWGPCGGSWAQKFYQDGPGRGLGEEDTRRAREAPRKTPRPQLSDRSRERKVP
CC1=C2C=C(C=CC2=NN1)C3=CC(=CN=C3)OCC(CC4=CC=CC=C4)N	MSDVTIVVEGWQKRGEYIKMWRPRYFLLKTDGSFIGYKEKPQDVDLPYPLNNFSVAKCQLMKTERPKPNTFIIRCLQWTTVIER
CC1=C2C=C(C=CC2=NN1)C3=CC(=CN=C3)OCC(CC4=CC=CC=C4)N	MDRSKENCISGPVKATAPVGGPKRVLVTQQFPCQNPLPVNSGQAQRVLCPSNSGQRVPLQAQKLVSSHMPVQNQKQKQLQATSVP
CC1=C2C=C(C=CC2=NN1)C3=CC(=CN=C3)OCC(CC4=CC=CC=C4)N	MAQKENSYPWPYGRQTAPSGSLSTLPQRVLRKEPVTPSALVLHSRSNVQPTAAPGQKVMENSSGTPDILTRHFTIDDFEIGRPLGK
CC1=C2C=C(C=CC2=NN1)C3=CC(=CN=C3)OCC(CC4=CC=CC=C4)N	MLGAVEGPRWKQAEDIRDIYDFRDVLGTGAFSEVILAEDKRTQKLVAIKCIAKEALEGKEGSHENEIAVLHKIKHPNIVALDDIY
CC1=C2C=C(C=CC2=NN1)C3=CC(=CN=C3)OCC(CC4=CC=CC=C4)N	MLPEDKEADSLRGNISVKAVKKEVEKKLRCLLADLPLPPELPGGDDLSKVSPEEKTATQLHSVKRPKICGPRYGETKEKDIDWGK
CC1=C2C=C(C=CC2=NN1)C3=CC(=CN=C3)OCC(CC4=CC=CC=C4)N	MENFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIRLDTETEGVPSTAIREISLLKELNHPNIVKLLDVIHTENKLYLVFEFLHQ
CC1=C2C=C(C=CC2=NN1)C3=CC(=CN=C3)OCC(CC4=CC=CC=C4)N	MALDVKSRAKRYEKLDFLGEGQFATVYKARDKNTNQIVAIKKIKLGHRSEAKDGINRTALREIKLLQELSHPNIIGLLDAFGHKS

Total number of samples in the last test set: 6009





A. Performance Metric

The performance of the proposed model will be verified through various traditional metrics of regression. This is to ensure accuracy, reliability, and robustness of the proposed technique. Regression metrics will be useful in validating the accuracy of a proposed method over various parameters like, accuracy, reliability, and robustness. It measures the boosted values of binding affinity between a drug and a skin protein, thereby measuring the required accuracy through various parameters like accuracy, reliability, and robustness. It also measures the precise nature of a drug-binding simulation using a precise technique, thereby evaluating the necessary accuracy through various parameters, ensuring accuracy, reliability, and robustness over a desired level.

B. Challenges

One of the major challenges in predicting binding affinity of a drug molecule and a protein is handling effectively with the complex interactions of nonlinear processes that happen in biological systems, which cannot easily be accounted for using a simple sequential model. An appropriate model selection method is required for developing a model that effectively learns hidden nuances of this system without structural details. However, another challenge is related with preprocessing and representation of the given data because, in terms of representation, there is a variation between the length of string representations of drugs and proteins. In this case, there is a chance of losing or adding information that is required to make predictions. The process of training a model and its overall generalization also contains some challenges, especially in preventing model over fitting. This is a common issue due to a lack of sufficient labeled training instances in available benchmarks, such as in the Davis dataset. The selection of apropriate hyper parameters, as well as model regularization, is an important key in achieving a more stable model convergence.

C. Discussion

It is noted from the experimental results that the proposed deep Drug–Target Interaction model is efficient in effectively predicting the binding affinity of drug–protein through sequence-based representations, as shown in the experimental outcome.The predicted affinity values show strong correlation in accuracy with experimental Kd value results, as demonstrated through low error metric values and high correlation scores. The employment of various metrics, including Mean Squared Error, Mean Absolute Error, Concordance Index, Pearson correlation, and R^2 metric scores, provides a comprehensive understanding of model efficiency in terms of accuracy, rankings, and overall generalization.

The experimental outcomes demonstrate that complex nonlinear associations in Drug-Protein interaction, based on various drug molecules, are well captured through the proposed deep DTI model, thereby showcasing its capabilities in efficient applications of virtual screening in the field of drug discovery.

It is, therefore, inferred that a deep DTI model, based on Machine Learning, is a reliable method of computing Drug-Protein interaction.

VII. CONCLUSION AND FUTUREWORK

A. Conclusion

This project proposes a method based on deep learning models and their ability in predicting drug–protein binding affinity by using sequence-level data from a dataset named Davis. Through its application of a deep Drug–Target Interaction model with its

ability to learn complicated patterns from its dual input encoders, it has resulted in a more precise and accurate approach in predicting binding affinities. Various results are presented to prove its efficiency and accuracy in its predictions, focusing on multiple regression and various ranking values including MSE, MAE, Conc. Index, Pearson Correlation, and R^2 score. The study signifies how DTI models based on machine learning would greatly help in saving time and money during experimental processes by providing a more efficient method of virtual screening in early-stage drug discovery.

B. Future Work

The proposed system may also be improved further using features of drug and protein structures and physicochemical properties besides sequence-based features to obtain a better representation of molecular interaction. Sophisticated techniques of deep learning methods like attention models, graph CNNs, and transformer-based approaches may also be pursued to better handle long-range relationships and complex interaction schemes. The training of the model with additional large and different datasets as well as several standards of binding affinity will make it more capable of handling different biological entities.

Future extensions may also consider model understandability, allowing us to pinpoint the important substructures in drugs and proteins that are mostly mean to affect the accuracy of binding affinity predictions.

VIII. REFERENCES

[1] A deep learning-based model for drug-target affinity prediction with application to drug repurposing and virtual drug screening - Discover Pharmaceutical Sciences, vol. 1, no. 16, Sept. 2025.

[2] D. Ma, S. Li, and Z. Chen, Drug-target binding affinity prediction method based on a deep graph neural network, - Mathematical Biosciences and Engineering, vol. 20, no. 1, pp. 269–282, 2023.

[3] Prediction of drug-target binding affinity based on deep learning models, - Elsevier, 2024.

[4] Structure-based, deep-learning models for protein-ligand binding affinity prediction - Journal of Cheminformatics, vol. 16, Art. no. 2, Jan. 2024.

[5] Drug–target binding affinity prediction with extended graph learning-convolutional networks - BMC Bioinformatics, 2024.

[6] GraphDTA: Predicting drug-target binding affinity with graph neural networks - Bioinformatics, 2020.

[7] Ö. Öztürk, A. Özgür, and E. Özgür, DeepDTA: Deep drug-target binding affinity prediction,- Bioinformatics, vol. 34, no. 17, pp. i821–i829, 2018.

[8] GramSeq-DTA: A grammar-based drug–target affinity prediction approach fusing gene expression information,- Biomolecules, MDPI.

[9] Y. Kalakoti, S. Yadav, and D. Sundar, TransDTI: Transformer-Based Language Models for Estimating Drug-Target Interactions, - ACS Omega, vol. 7, no. 3, pp. 2706–2717, 2022.

[10] Prediction of drug-target binding affinity based on deep learning models - Elsevier, 2024.

[11] J. Yang, M. Li, S. Zhang, S. Wang, X. Wang, Q. Yuan, and Z. Wei, Drug–target affinity prediction using graph neural network and contact maps - RSC Advances, vol. 10, pp. 20701–20712, 2020.

[12] J. Lee, B. Park, and S. Kim, CSatDTA: Prediction of Drug–Target Binding Affinity Using Convolution Model with Self-Attention,- International Journal of Molecular Sciences, vol. 23, no. 15, p. 8453, 2022.

[13] S. Wang, X. Yang, and G. Yang, DeepMHADTA: Prediction of Drug-Target Binding Affinity Using Multi-Head Self-Attention and Convolutional Neural Network - Molecules, vol. 44, no. 5, p. 155, 2024.

[14] X. Yang, G. Yang, and J. Chu, GraphCL-DTA: Graph contrastive learning with molecular semantics for drug-target binding affinity prediction - arXiv, Jul. 2023.

[15] C. Zhuang, L. Shi, Y. Chen et al., Drug-target binding affinity prediction using message passing neural network and self-supervised learning, - BMC Genomics, vol. 24, p. 557, Sept. 2023.