IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 May 2025, accepted 17 June 2025, date of publication 26 June 2025, date of current version 9 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3583338

RESEARCH ARTICLE

Binding Affinity Prediction for Pancreatic Ductal Adenocarcinoma Using Drug-Target Descriptors and Artificial Intelligence

PRAGYA[1], A. AMALIN PRINCE[2], (Senior Member, IEEE), AND JAC FREDO AGASTINOSE RONICKOM[1]

[1]Computational Neuroscience and Biology Laboratory, School of Biomedical Engineering, Indian Institute of Technology (BHU) Varanasi, Varanasi 221005, India
[2]Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, K K Birla Goa Campus, Sancoale, Goa 403726, India

Corresponding author: A. Amalin Prince (amalinprince@goa.bits-pilani.ac.in)

ABSTRACT Pancreatic ductal adenocarcinoma (PDAC) is the most common and aggressive form of pancreatic cancer, accounting for 90% of all pancreatic malignancies. This study addresses the gap in disease-specific binding affinity prediction by integrating PDAC-derived targets with diverse molecular descriptors and artificial intelligence (AI) models, enabling more accurate therapeutic profiling. Initially, we constructed a drug library using compounds from the DepMap database, targeting proteins such as LIFR, BTG2, EPHX2, and PAK3 identified as differentially expressed genes in a previous PDAC study. We employed descriptors such as Conjoint Triad, amino acid composition (AAC), and Quasi sequence order to represent the targets. Similarly, the drugs were described by Morgan, RDKit, and PubChem descriptors. We used AI algorithms like random forest regressor (RFR), extreme gradient boost regressor (XGBR), and one-dimensional convolutional neural network (1D-CNN) to predict the binding affinity. We also employed two benchmark datasets, DAVIS and BindingDB, to compare our models' performance in binding affinity prediction. We achieved a mean square error (MSE) value of 1.5 using Morgan-RDKit-PubChem-Conjoint descriptors and 1D-CNN on the PDAC dataset. Similarly, 1D-CNN with PubChem-AAC descriptors produced an MSE of 0.27 on the DAVIS dataset. Further, the XGBR model using the PubChem-AAC descriptors produced an MSE of 0.69 on BindingDB. Our study demonstrates the potential of an AI-driven framework as an effective and scalable solution for disease-specific drug-target interaction prediction, with promising implications for drug repurposing in PDAC.

INDEX TERMS Pancreatic ductal adenocarcinoma, drug-target descriptors, drug repurposing, binding affinity, artificial intelligence.

I. INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is one of the most aggressive solid tumors, characterized by a poor prognosis. In 2020, there were an estimated 495,773 new cases and 466,003 deaths worldwide, with a notably high incidence reported in the Western Pacific region, including India. Current treatment strategies for PDAC include surgery, chemotherapy (e.g., gemcitabine), radiation therapy, CAR-T cell therapy, and personalized neoantigen vaccines [1]. Additionally, targeted therapies are being explored for patients with specific genetic mutations [2], [3]. However, these therapies often face challenges such as drug resistance and the limited efficacy of immunotherapy. In such cases, repurposing existing drugs offers a cost-effective alternative to de novo drug development [4], [5]. Drug repurposing involves identifying new therapeutic uses for existing drugs, including those already approved for other diseases. In this study, we aim to repurpose existing drugs for the treatment of PDAC based on identified genetic biomarkers.

Drug repurposing can be achieved through *in vitro*, *in vivo*, and computational (*in silico*) methods. *In vitro* approaches

The associate editor coordinating the review of this manuscript and approving it for publication was Vincenzo Conti.

involve testing existing drugs on cultured cells or organoids to observe their biological effects on disease-related pathways, offering high-throughput and relatively low-cost screening. *In vivo* methods utilize animal models to study the efficacy, toxicity, and pharmacokinetics of repurposed drugs in a whole-organism context, providing more physiologically relevant insights [6]. Meanwhile, computational methods leverage bioinformatics to analyze large-scale datasets such as gene expression profiles, protein interaction networks, and drug-target databases to predict new therapeutic uses for existing compounds. These *in silico*strategies are fast, cost-effective, and hypothesis-generating. In this study, we explored computational methods for drug repurposing in PDAC [7].

Computational drug repurposing involves predicting drug-target interactions (DTIs) for a specific disease condition. A critical step in DTI prediction is the effective representation of drugs and targets in formats suitable for computational processing. The simplified molecular input line entry system (SMILES) is commonly used in computational studies to represent and describe the chemical structure of drugs and the sequences of proteins [8]. Further, SMILES strings are converted into computationally analyzable formats using molecular descriptors, enabling effective drug repurposing for PDAC treatment [9]. Drug molecules can be represented using various 1D descriptors, including Morgan, PubChem, the molecular access system, and RDKit. For target-based descriptors, commonly used representations include Conjoint Triad, amino acid composition (AAC), k-mers, and Quasi-sequence order (Quasi-seq) [10], [11]. Further, they can be represented in 2D derived from molecular topology (connectivity), like extended-connectivity fingerprints, graph-based embeddings, and position-specific scoring matrix. Similarly, 3D descriptors incorporate the spatial geometry of the molecule, like comparative molecular field analysis and a 3D voxel grid of the binding pocket [12]. However, the 2D and 3D descriptors require an increased computational demand, greater complexity, and potential issues with data quality and interpretability compared to the robust and straightforward 1D descriptors. In this study, we utilized Conjoint Triad, AAC, and Quasi-seq for target representation, and Morgan, RDKit, and PubChem fingerprints as drug descriptors [13], [14], [15].

Molecular docking is a computational technique used to predict DTIs. Several toolboxes, such as AutoDock, AutoDock Vina, and GLIDE, are available for molecular docking. However, these tools are often time-consuming and labor-intensive. To overcome these challenges, studies have explored the use of artificial intelligence (AI) algorithms for DTI prediction. Machine learning (ML) has revolutionized molecular docking by enhancing the prediction of protein-ligand binding affinities [16], [17]. Guvenilir et al. used random forest regressor (RFR) and support vector machine (SVM) algorithms for DTI prediction, as they are widely used in benchmark studies [18]. Other studies have employed one-dimensional convolutional neural networks (1D-CNNs) [19] and attention layers with 1D-CNNs to predict DTIs [20]. Additionally, two-dimensional (2D-CNN) [21] and three-dimensional convolutional neural networks (3D-CNNs) have also been used for DTI prediction [22]. However, the use of complex 2D and 3D-CNN architectures on small datasets may introduce additional noise in the supervised DTI prediction stage. This may potentially mask the ability of learned representations to capture ligand interaction-related properties of proteins, hindering a fair evaluation of model performance. In this study, we aim to predict DTIs using RFR, extreme gradient boosting regressor (XGBR), and 1D-CNN models.

The main contributions of this study are as follows:

- Prepared a library of drugs using DepMap for PDAC and compared it with the benchmark datasets.
- Extracted molecular descriptors for targets using Conjoint Triad, AAC, and Quasi-seq, and drugs using Morgan, Rdkit and PubChem.
- Employed the RFR, XGBR and 1D-CNN model for drug target affinity prediction and evaluated its performance using coefficient of determination (R^2), mean square error (MSE), root mean square error (RMSE) and mean absolute error (MAE).

II. METHODOLOGY

Figure 1, shown below, depicts the study process pipeline. The first step is target identification, followed by the preparation of the drug library corresponding to PDAC. Then, we extracted the molecular descriptors of the target and drug of PDAC data, as well as benchmark datasets DAVIS and BindingDB. Further, AI regression models are trained, and the performance is evaluated using different metrics.

A. DATASETS

1) PDAC DATASET

We prepared a library of drugs corresponding to PDAC-identified targets from our previous study, such as LIFR, BTG2, EPHX2, and PAK3. The drugs were selected using the Cancer Dependency Map (DepMap) database, a significant research initiative aimed at identifying essential genes across various cancer cell lines through advanced genomic techniques such as CRISPR and shRNA screens [23]. To identify drugs targeting specific genes, we used the ''Tools'' option in the Data Explorer available through the DepMap portal. We selected the ''Expression Public 23Q4'' dataset and sorted the list of drugs based on available drug sensitivity data. Further, we ranked the drugs based on their correlation values and selected a library of 15 drugs for each target. This led to a drug library of 60 drugs. We performed molecular docking using Autodock with 46 drugs (a few were common uric acid, KU55933, AP-24534, Alisertib, KU-57788, MK-0457, GSK525762A, Decitabine and few drugs such as AY-22989, Bortezomib, Eloxatin, TL-1-85 do not have 3D structure in

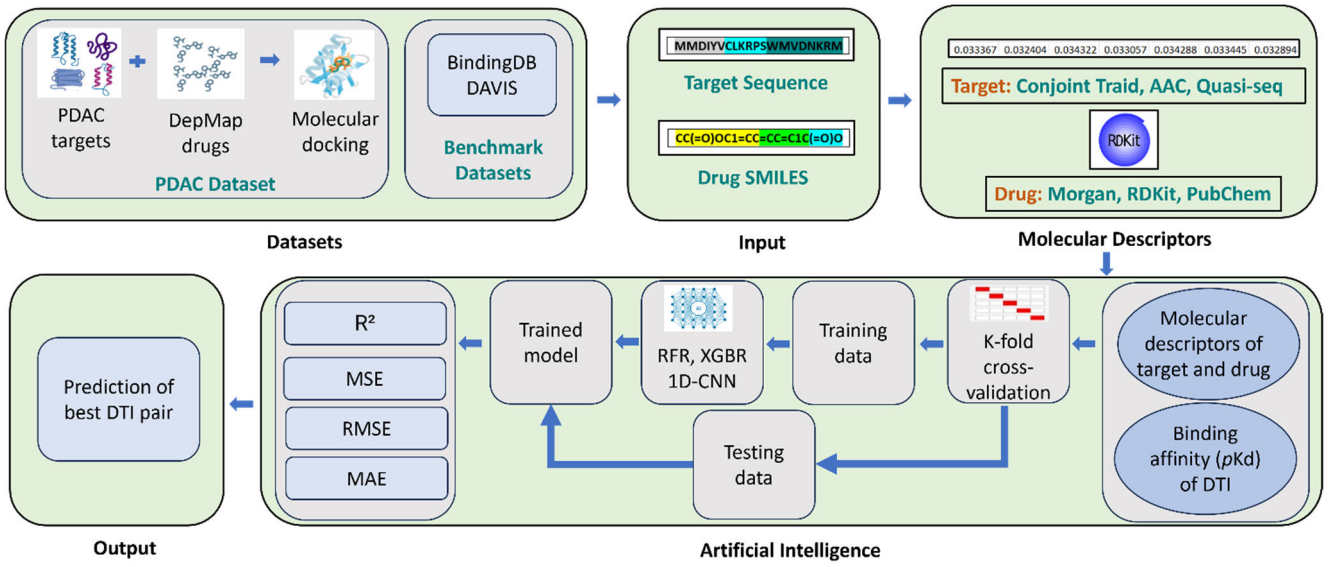

FIGURE 1. Process pipeline of the study.

PubChem), four proteins (LIFR, BTG2, EPHX2, and PAK3), and 184 DTI pairs.

2) BINDINGDB DATASET

It is a publicly accessible database that catalogs experimentally determined binding affinities between small molecules and proteins. This resource facilitates a variety of applications, including medicinal chemistry, the training of AI models, and the development of computational chemistry methods [24], [25]. This dataset consists of 10,665 drugs, 1,413 proteins and 52,284 DTI pairs.

3) DAVIS DATASET

It is a benchmark dataset used for DTI prediction that focuses explicitly on kinase inhibitors and kinases. The dataset includes experimentally measured binding affinities, typically represented as kinase dissociation constants (K_d), which indicate the strength with which a drug binds to a target protein [26]. This dataset consists of 68 drugs, 379 proteins and 25,772 DTI pairs.

B. MOLECULAR DESCRIPTORS

We employed three drug and target descriptors, Morgan, RDKit, PubChem, and Conjoint, AAC, Quasi seq, respectively, for our analysis. Morgan fingerprints encode the structural information of a molecule into a binary vector. This vector is generated on the basis of the arrangement and connectivity of atoms within the molecule [27]. RDKit is an open-source cheminformatics toolkit. It provides advanced functionality for creating molecular fingerprints. PubChem is a binary molecular representation that captures the presence or absence of specific chemical substructures within a molecule. These are categorized by their dimensionality and the specific structural information they represent [28]. The Conjoint Triad descriptor represents protein sequences by grouping amino acids into triads, where the three amino acids in each triad belong to the same class and share functional similarities. It consists of several descriptors that quantify different aspects of the amino acid sequence [29], [30]. AAC serves as a fundamental approach for encoding protein sequences into numerical vector representations. Each component of the resulting vector represents the frequency of a specific amino acid within the sequence. The protein sequences were extracted from FASTA files. Subsequently, the sequences were processed to generate their corresponding numerical representations [31]. Quasi-seq is a combination of sequence composition and correlation of physicochemical properties like hydrophobicity, hydrophilicity, polarity, and side-chain volume [32]. It encodes both the amino acid composition and local sequence-order patterns using physicochemical distance matrices.

C. K_D, K_I AND BINDING ENERGY RELATIONS

Binding energy (ΔG) and K_i values for DTI prediction in PDAC were obtained from molecular docking using AutoDock for the identified targets and the prepared drug library. In the case of the BindingDB and DAVIS datasets, the relation between the target and drug is represented by dissociation constant (K_d). Therefore, we converted K_i values obtained from the molecular docking (AutoDock) into K_d for comparison with both datasets. Initially, the K_i was converted to K_d and then to pK_d to standardize drug-target binding affinity, ensuring consistency and comparability in computational models [33]. The K_d measures ligand-receptor binding affinity, with lower values indicating stronger binding. It is expressed as:

$$\Delta G = -RT \ln(K_d) \quad (1)$$

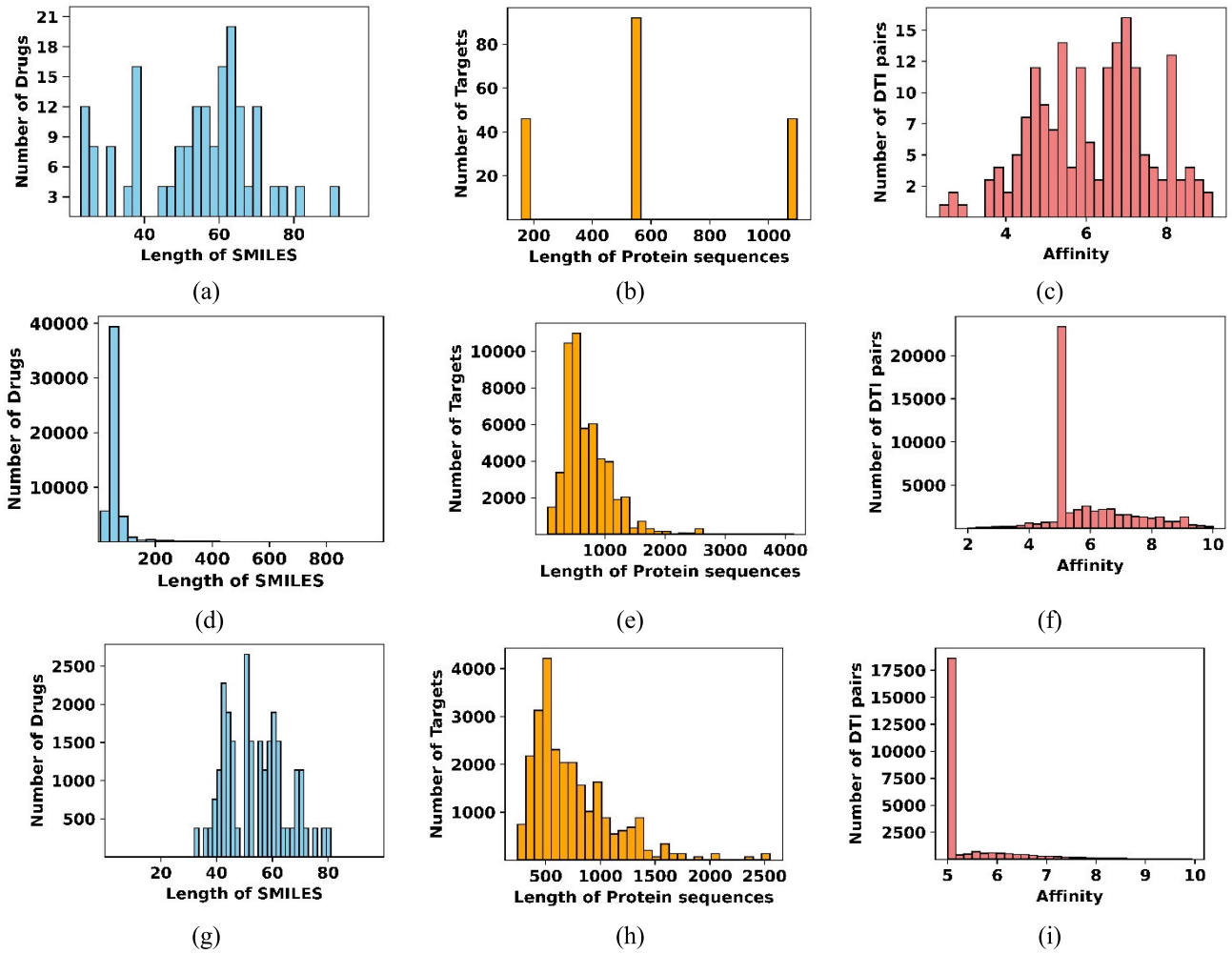

FIGURE 2. Histogram plot of length of SMILES, protein sequence, and affinity in (a-c) PDAC, (d-f) BindingDB, and (c) DAVIS dataset.

where R is the universal gas constant, and T is the temperature (298K). The affinity values were transformed into logarithm space (pK_d) by applying the following equation

$$pK_d = -\log(K_d/1e9) \tag{2}$$

It facilitates a straightforward comparison of binding affinities [34]. Figure 2 (a-c), (d-f), and (g-i) represent the length of SMILES, protein sequences, and the pK_d values distribution for the PDAC, BindingDB and DAVIS datasets, respectively. The length of SMILES of drugs lies between 20 and 90 for the PDAC and DAVIS datasets, whereas for BindingDB it lies from 100 to 900. This shows the high variation between the values in the datasets. The corresponding length of protein sequences for these datasets lies between 150-1200, 250-4500, and 250-2500. The majority of the protein sequences have lengths of 500. It can be noted that the ranges of pk_d are within 2 to 10 in the PDAC and BindingDB datasets. Similarly, for DAVIS, it is 5 to 10, with the majority of the values lying around 5.

D. DTI PREDICTION USING AI ALGORITHMS

In this study, we employed three AI algorithms, RFR, XGBR and 1D-CNN to predict the DTIs. We built the models on different combinations of inputs as follows: Morgan+ Conjoint Traid, RDKit+Conjoint Traid, PubChem+Conjoint Traid, Morgan+AAC, RDKit+AAC, PubChem+AAC, Morgan+Quasi-seq, RDKit+Quasi-seq, PubChem+Quasi-seq, Morgan+ RDKit+ PubChem+Conjoint Traid, Morgan+ RDKit+PubChem+AAC, Morgan+RDKit+ PubChem+ Quasi-seq, Morgan+Conjoint Traid+AAC+ Quasi-seq, RDKit+ Conjoint Traid+ AAC+Quasi-seq and PubChem+Conjoint Traid+ AAC+Quasi-seq. We chose these combinations, ensuring one target and one drug descriptor in addition to one target with three drug descriptors and vice versa, to know the role of molecular descriptors in DTI prediction. RFR can determine the most influential molecular descriptors, play a crucial role in predicting outcomes, and offer valuable guidance for experimental validation of critical interactions [35]. It is an ensemble method that builds multiple unpruned regression trees using bootstrap

samples of the training data along with random feature selection during tree construction. Predictions are generated by combining the outputs of all trees, through averaging (for regression). Additionally, it offers three valuable features: built-in performance evaluation, the ability to assess the relative importance of descriptors, and a compound similarity measure that incorporates descriptor importance weighting [36], [37]. DTI prediction using XGBR has emerged as a gradient boosting algorithm based on decision trees. This model addresses the complex nonlinear relationships between input features and interaction outcomes in computational drug discovery, combining feature engineering efficiency with robust predictive performance [38]. It is an approach to predict DTIs based on structured numerical representations of both drug molecules and target proteins. Drug features were derived from molecular fingerprints or physicochemical descriptors, while target proteins were encoded using sequence-based or structural features. These representations were concatenated into a single feature vector for each drug-target pair, capturing the combined chemical and biological information necessary for interaction prediction. Due to its ability to handle feature interactions and prevent overfitting through built-in regularization, it served as an effective and interpretable method for DTI prediction. 1D-CNN is a deep learning model that effectively captures local patterns in sequential data, such as SMILES strings or protein sequences. In DTI prediction, 1D-CNNs are used to extract meaningful features from molecular and protein sequences, enabling accurate interaction modeling [39] from vectorized representations of drug molecules and target proteins. Separate CNN branches were constructed for drugs and targets, each composed of stacked 1D convolutional layers with increasing filter sizes to capture local and global features. Global Max Pooling was applied to retain the most informative features from each branch. The extracted features from both branches were then concatenated and passed through fully connected layers to model the interaction between drugs and targets. Dropout layers were used to mitigate overfitting, and the model was trained using the mean squared error loss function with the Adam optimizer. This deep learning approach allows the model to automatically learn relevant features for DTI prediction without the need for manual feature engineering.

E. CROSS-VALIDATION AND PERFORMANCE METRICS CALCULATION

We performed 5-fold cross-validation on PDAC, as the dataset has 184 DTI pairs. The BindingDB and DAVIS datasets have more data, so 10-fold cross-validation was implemented. This helps to assess the performance of models on different data segments. The data were split into five or ten folds, and in each iteration, the model is trained on 4 or 9 folds of the data and tested on the remaining fold, allowing every data point to be used for both training and validation at least once. This process provides a more reliable estimate of model performance compared to a simple train-test split.

TABLE 1. Drugs corresponding to the identified targets for PDAC.

Targets	Drugs
LIFR	AP-24534*, P 22077, PILOCARPINE, ARAVA, URIC-ACID*, ALISERTIB*, AZ 960, DECITABINE*, BIIB-057, SELUMETINIB:DECITABINE (4:1 MOL/MOL), TL-1-85#, GSK525762A, EX-8678, WEHI-539, CYTARABINE HYDROCHLORIDE
BTG2	4-IODO-6-PHENYLPYRIMIDINE, KU-55933*, AZ-628, KU-57788*, CS-110266, AB1010, URIC-ACID*, LY2606368, VE 821, ARRY-886, 944328-88-5, 284461-73-0, RO-3306, FTI-277, OSI930
EPHX2	ELOXATIN#, KU-55933*, NIRAPARIB, BI-2536, 7-NITROINDAZOLE, ENOXIMONE, GSK525762A, ALISERTIB*, TROLOX, FLUVASTATIN, MK 0457, VX-680, REPAGLINIDE, KU-57788*, AK-55409
PAK3	NILUTAMIDE, URIC-ACID*, ESTRAMUSTINE-PHOSPHATE, GAMMA-SECRETASE INHIBITOR 1, GSK-2141795, PLX 4720, MK 0457, BORTEZOMIB#, ACETYLCYSTEINE, MEPIVACAINE, GSK-J4, CGP-082996, AP-24534*, AY-22989#, CHLORINDANOL

* Depicts the drugs that were common between the targets, # Depicts the drugs having no 3D structure on PubChem

We evaluated the performance of the models using R^2, MSE, RMSE and MAE. The average of the 5 or 10 folds was reported as the performance of the model. R^2 quantifies the extent to which a regression model explains the variance in the observed data. It ranges from 0 to 1, where 1 indicates a perfect fit, and 0 means the model explains no variance in the data.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\Sigma (y_i - \bar{y})^2} \tag{3}$$

where y_i is the actual/observed values, $\hat{y}_i$ is the predicted values, $\bar{y}_i$ is the mean of actual values, and N is the number of observations. MSE evaluates the quality of an estimator or predictor by quantifying the average squared difference between the estimated values and the actual values.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{n} (yi - \hat{y}i)^2 \tag{4}$$

RMSE quantifies the difference between predicted values and observed values, providing insight into the model's performance.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (yi - \hat{y}i)^2} \tag{5}$$

MAE quantifies the average absolute differences between predicted values and actual/observed values. It is beneficial because it treats all errors equally, regardless of their direction (positive or negative), and is given by

$$MAE = \frac{1}{N} \sum_{i=1}^{n} |yi - \hat{y}i| \tag{6}$$

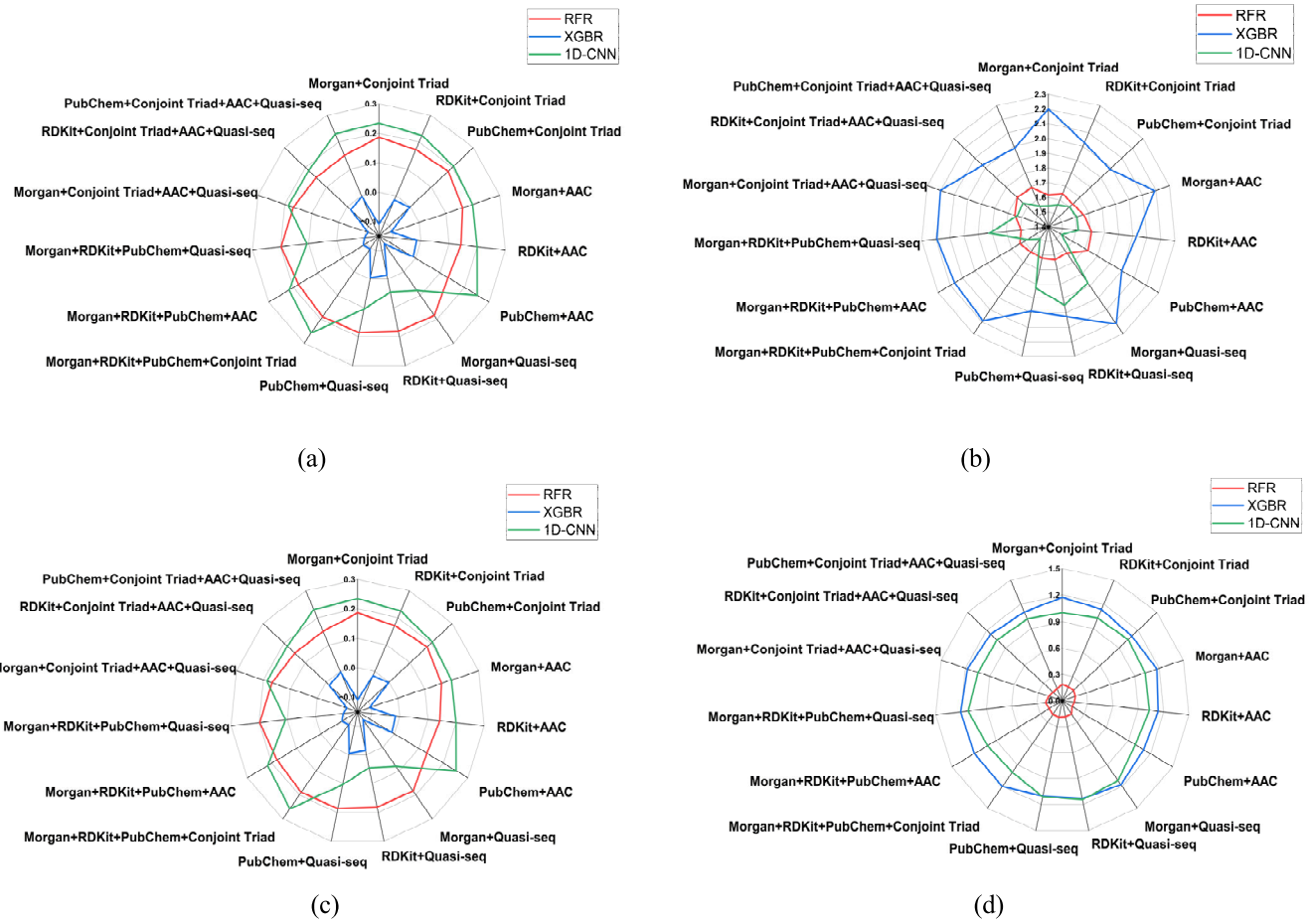

FIGURE 3. Performance evaluation of AI models on PDAC datasets based on (a) R^2, (b) MSE, (c) RMSE, and (d) MAE.

III. RESULTS

Table 1 shows the library of drugs selected from the DepMap database for the PDAC targets (LIFR, BTG2, EPHX2, and PAK3). To ensure the relevance of the selected compounds, we considered the top 15 drugs for each, which exhibit a significant relationship with specific gene targets. Thereby prioritizing those with stronger associations to the drug sensitivity profiles. This rigorous selection process resulted in a final library comprising 60 compounds representing diverse mechanisms of action. Finally, 46 drugs (a few were common uric acid, KU55933, AP-24534, Alisertib, KU-57788, MK-0457, GSK525762A, Decitabine and few drugs such as AY-22989, Bortezomib, Eloxatin, TL-1-85 do not have 3D structure in PubChem).

Figure 3 (a-d) represents the performance of AI models on the PDAC dataset using the different combinations of descriptors. We received the highest performance on RFR using the Morgan+RDKit+PubChem+Quasi-seq combination with R^2, MSE, RMSE, and MAE of 0.19, 1.5, 1.2, and 1, respectively. In the case of XGBR, the PubChem+Conjoint Traid+AAC+Quasi-seq combination produced the performance with R^2 of −0.003, MSE of 1.9, RMSE of 1.4, and MAE of 1.1. We achieved the highest prediction performance with the 1D-CNN using Morgan+RDKit+PubChem+Conjoint Traid combination with R^2 of 0.25, MSE of 1.5, RMSE of 1.21, and MAE of 1. We can see that 1D-CNN seems to outperform RFR and XGBR on our PDAC dataset. RFR and 1D-CNN perform well with three drug descriptors and one target descriptor, whereas in XGBR, one drug and three target descriptors show better outcomes. It can be noted that the multiple drug and target descriptors perform well compared to individual target and drug descriptors. This emphasizes the importance of using various descriptors to fit the model. However, we were not able to find a pattern between the models.

Figure 4 (a-d) represents the performance of the AI models on the BindingDB dataset using different combinations of descriptors. RFR model produced the highest performance using Morgan+RDKit+PubChem+AAC combination, with $R^2$of 0.6, MSE of 0.74, RMSE of 0.86, and MAE of 0.53. We achieved the highest accuracy with XGBR using PubChem+AAC combination with the R^2, MSE, RMSE, and MAE of 0.63, 0.7, 0.83, and 0.54, respectively. Similarly, the 1D-CNN using RDKit+AAC combination showed the

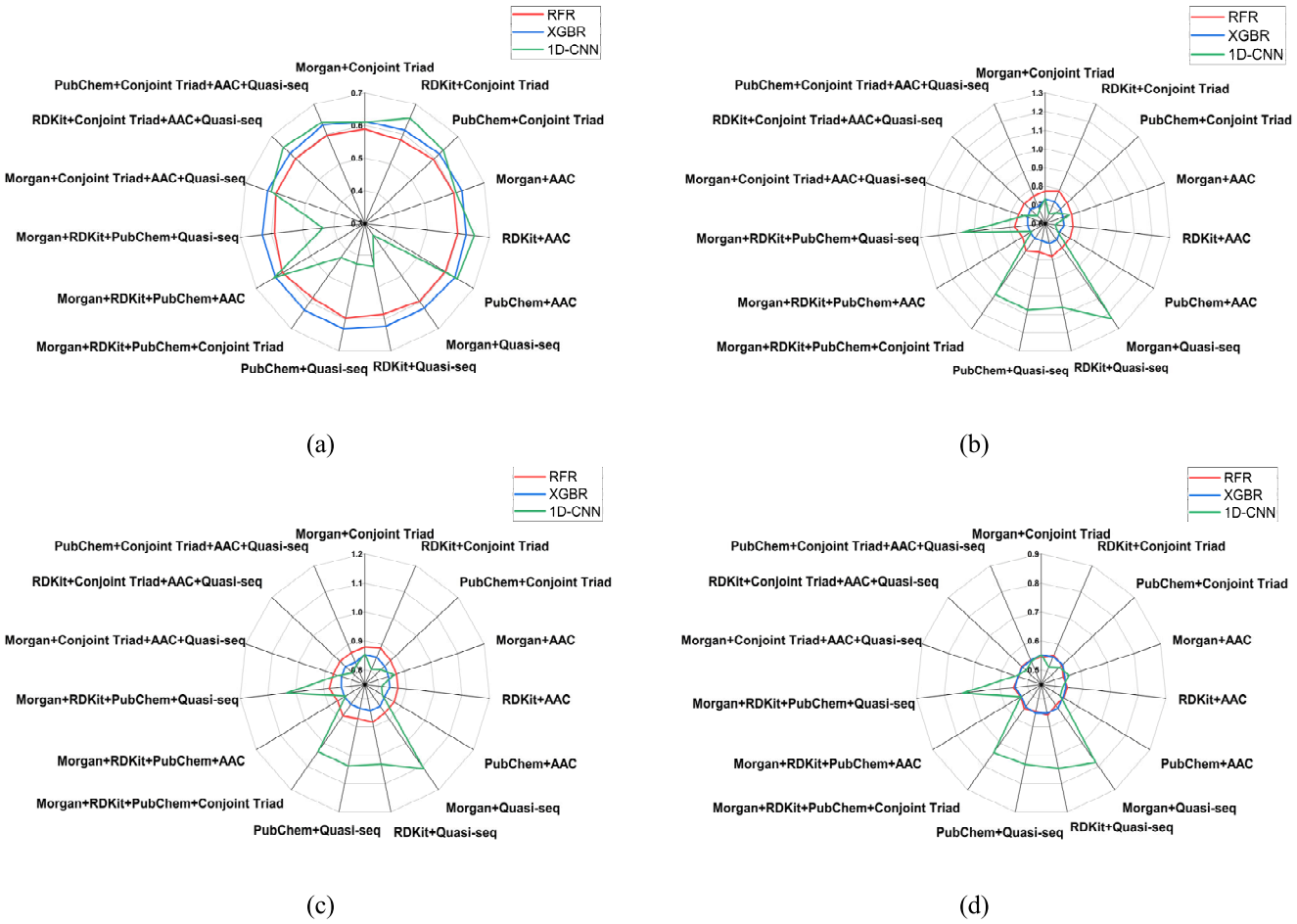

FIGURE 4. Performance evaluation of AI models on BindingDB datasets based on (a) R^2, (b) MSE, (c) RMSE, and (d) MAE.

best performance, with R^2 of 0.65, MSE of 0.65, RMSE of 0.81, and MAE of 0.52. Notably, the features computed using AAC (target descriptor) are common in all top-performing models, indicating their significance in binding affinity prediction. Furthermore, while ensemble learning models like XGBR and RFR exhibited robust performance with multiple descriptor sets, the 1D-CNN model consistently demonstrated better results, particularly when paired with sequence-based descriptors. This suggests that deep learning architectures are better at capturing intricate patterns in multimodal descriptor data.

Figure 5 (a-d) represents the performance of the models on the DAVIS dataset using different combinations of descriptors considered in the study. In RFR model, Morgan+Conjoint Traid+AAC+Quasi-seq combination shows the best performance with R^2 of 0.5, MSE of 0.32, RMSE of 0.57, and MAE of 0.34. In the case of XGBR, Morgan+RDKit+PubChem+AAC combination shows the best performance with R^2 of 0.59, MSE of 0.27, RMSE of 0.52, and MAE of 0.33. Similar performance was produced by 1D-CNN on PubChem+AAC combination with R^2 of 0.59, MSE of 0.27, RMSE of 0.52, and MAE of 0.33. On this dataset, the 1D-CNN and XGBR models outperform the RFR model. In terms of descriptors in 1D-CNN, single drug and target (PubChem+AAC) combination performed well, but in the case of RFR and XGBR, multiple combinations of descriptors showed better performance. This highlights that deep learning models may generalize from fewer but informative features, whereas ensemble models benefit from diverse input representations.

Figure 6 (a) shows the predicted values of the RFR model using the Morgan+RDKit+PubChem+Conjoint Traid combination on the PDAC dataset. The figure helps visually compare model effectiveness in terms of closely predictions align with actual values. We can observe a spread around the regression line, which indicates moderate prediction accuracy. Figure 6 (b) shows the predicted values XGBR model using the PubChem+Conjoint Triad+AAC+Quasi-seq combination on the PDAC dataset. In this model, the data points show a lesser alignment with the ideal fit compared to others; variance remains, which suggests prediction limitations. Figure 6 (c) shows 1D-CNN using Morgan+RDKit+PubChem+Conjoint Traid combination. In this regression, the data points are more tightly

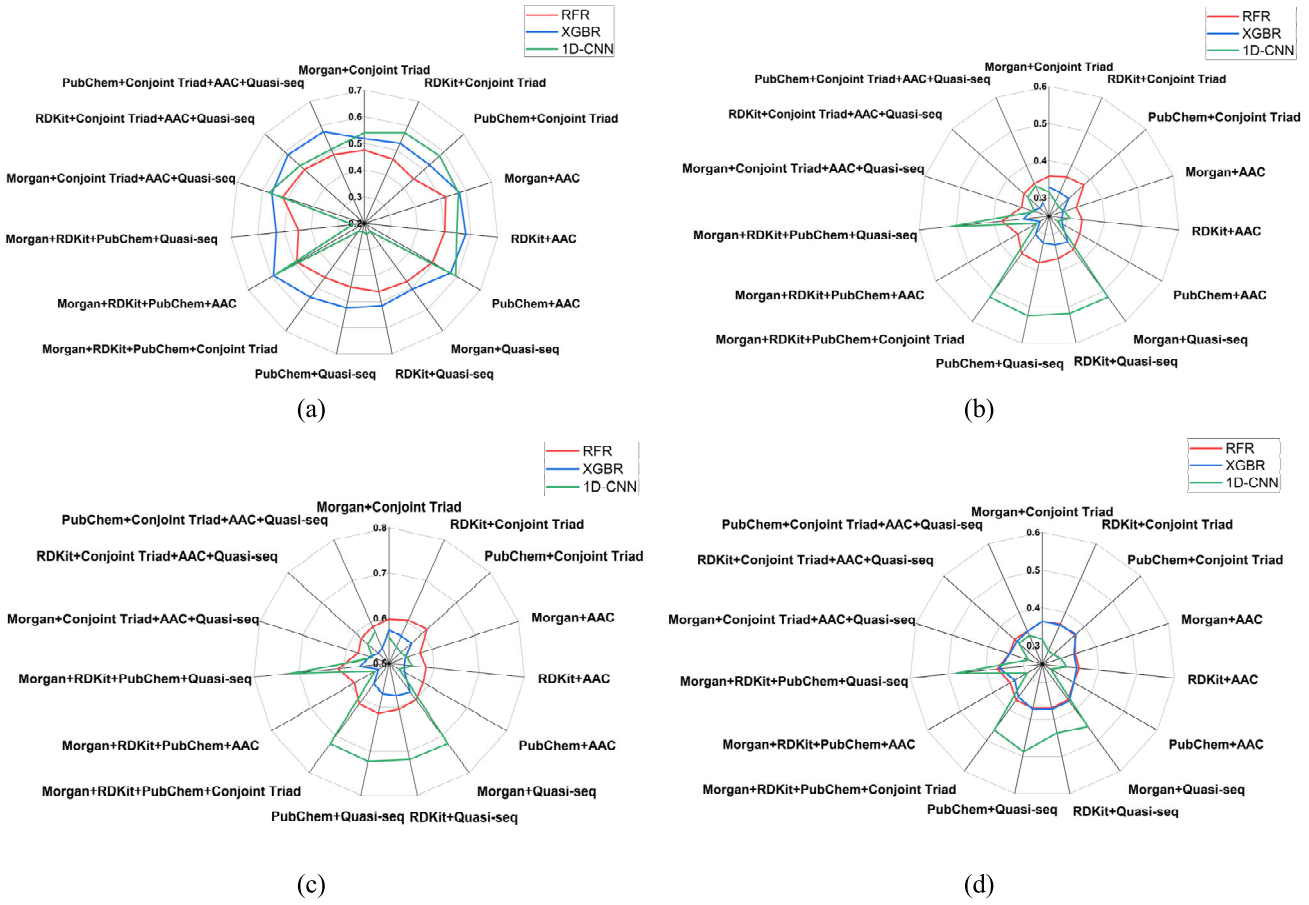

FIGURE 5. Performance evaluation of AI models on DAVIS datasets based on (a) R^2, (b) MSE, (c) RMSE, and (d) MAE.

clustered around the line, indicating improved predictive performance and reduced error. This suggests that the 1D-CNN model effectively captures complex, nonlinear relationships between molecular and sequence-based features, outperforming traditional ML models (RFR and XGBR). This also indicates the generalization capacity of 1D-CNN in predicting drug-target binding affinity on the PDAC dataset.

IV. DISCUSSIONS

In our study, we demonstrated that several critical factors can significantly influence the performance and interpretation of DTI prediction models. Specifically, we highlighted the impact of (i) Datasets (PDAC, BindingDB, and DAVIS), (ii) Drug and target descriptors (RDKit, PubChem, Morgan, AAC, Quasi-Seq, and Conjoint Triad descriptors), and (iii) AI algorithms (RFR, XGBR, and 1D-CNN). We explored 15 different combinations of these drugs and target features to evaluate model performance comprehensively. Our results underscore the necessity of carefully considering these factors when developing DTI models. Table 2 summarizes recent studies on drug-target binding affinity prediction using AI models, highlighting the datasets, molecular descriptors, model architectures, and their respective performance metrics.

A. EFFECT OF DATA CHOSEN FOR THE ANALYSIS

Our models performed well on DAVIS followed by BindingDB and PDAC dataset. The difference in MSE between DAVIS and BindingDB may be partly attributed to the variation in their affinity value ranges. DAVIS has a narrower affinity range (5 to 10), while BindingDB spans a broader range (2 to 10). Since MSE is sensitive to the scale and variability of the target variable, the broader range in BindingDB increases the complexity of the prediction task. This increased variability in BindingDB likely results in higher MSE compared to the more constrained and homogeneous affinity values in the DAVIS dataset. In addition to the wider affinity value range (2 to 10), the higher MSE observed in the PDAC dataset may also be influenced by its small sample size, containing only 184 binding affinity values. This limited the model's ability to learn complex patterns and generalize effectively. Consequently, the combination of a wide target range and a limited number of samples likely contributes to the elevated MSE in PDAC compared to larger datasets like DAVIS and BindingDB. In a study by Abbasi et al., the proposed model achieved best performance on the KIBA with the lowest MSE. BindingDB exhibited intermediate performance, while DAVIS had the highest MSE. These differences highlight dataset characteristics as well as the

interaction with the AI model to influence prediction accuracy [5]. Kalemeti et al. evaluated their proposed model on the different benchmark datasets such as KIBA, BindingDB, DAVIS and PDBbind. It was reported that the PDBbind and DAVIS performed exceptionally well, demonstrating superior prediction accuracy and robustness in drug-target binding affinity [48]. This matches our results, where the models perform well in the DAVIS dataset compared to BindingDB.

B. EFFECT OF DRUG-TARGET DESCRIPTORS

The study indicates that the AAC descriptor consistently contributes to high model performance across all datasets and algorithms, highlighting its significance in target (protein) representation for binding affinity prediction. This may be due to the fact that AAC provides a fixed-length numerical representation of proteins. It captures the overall frequency of each amino acid in a protein sequence. It highlights global patterns in protein composition that are often enough to distinguish functionally different proteins. Additionally, drug descriptors such as PubChem and Morgan fingerprints frequently appear in top-performing combinations, underscoring the value of integrating multiple molecular feature types. While no single descriptor combination consistently outperforms across all datasets, models tend to achieve better predictive accuracy when both drug and target descriptors are used together. Ji et al. compared the models performance using Morgan and behavioral descriptors. The study found that the combination of Morgan and behavior performs well compared to the individual descriptors [49]. Alkhadrawi et al. employed multiple individual combinations of drug and target descriptors to identify potential membrane transporters for glycyrrhetinic acid and reported that the Morgan-Conjoint Traid combination showed the highest accuracy. AAC placed sixth and occurred frequently 6 times out of the 12 combinations [29]. This matches our results showing the influence of AAC on the higher performance of the AI models.

C. AI REGRESSOR PERFORMANCE VARIABILITY

In this study, 1D-CNN performed well consistently across all the datasets compared to RFR and XGBR. This may be due to the ability of deep learning architectures to use convolutional filters to capture local and global patterns of multimodal descriptor data. In addition, regularization techniques such as dropout help in performing well. For instance, convolutional architectures like 1D-CNN inherently optimize feature extraction from sequential or grid-like data, which explains their robustness with sequence-based descriptors like AAC. Meanwhile, ensemble methods like RFR and XGBR rely on feature engineering. A study by Ismail et al. compared ML and deep learning and reported that deep learning models (especially SEQ) consistently outperformed ML models in classifying whether DTIs are active or inactive [50]. Ji et al. compared different ML classifier models such as logistic regression, k-nearest neighbor, Naïve Bayes, RF and decision tree. They reported that RF performed well in comparison to

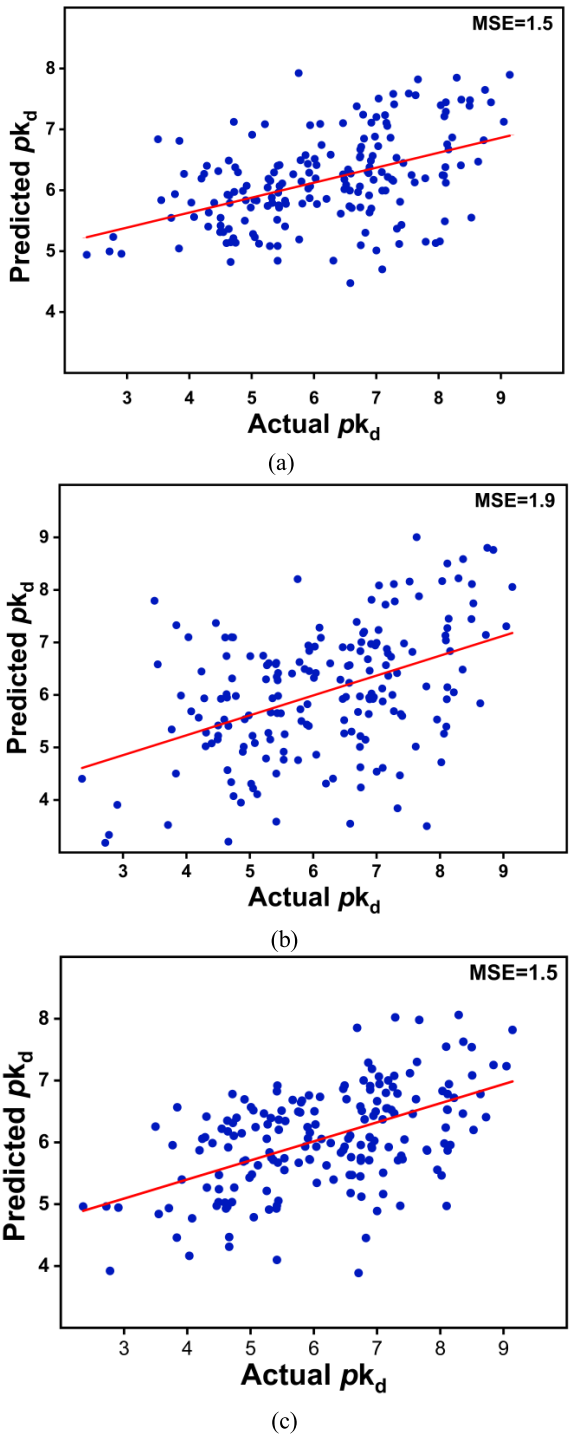

FIGURE 6. Actual vs Predicted pKd of DTI on PDAC dataset (a) RFR, (b) XGBR, and (c) 1D-CNN models.

the other classifiers considered in the study [49]. Chu et al. emphasize that neural networks excel in fusing heterogeneous features through hierarchical representations [51].

D. LIMITATIONS AND FUTURE SCOPE

This study highlights the potential of AI in predicting drug-target binding affinities for PDAC. While this study

TABLE 2. Comparison of studies that used the AI algorithms for DTI prediction.

Author	Dataset	Type of Dataset	Descriptors	AI Algorithms	Performance metrics
Our study	PDAC, BindingDB and DAVIS	Target specific dataset (PDAC) and benchmark dataset	Conjoint triad, AAC, and Quasi-seq (targets) Morgan, RDKit, and PubChem (drugs)	RFR, XGBR, 1D-CNN	MSE PDAC-1.5 DAVIS-0.27 BindingDB-0.65
Ranjan et al. (2024) [40]	DAVIS and KIBA	Benchmark Dataset	1D, 2D, and 3D (sequence, graph, and structure features)	MDF-DTA	MSE KIBA-0.14 DAVIS-0.17
Liyaqat et al. (2023) [41]	DAVIS and KIBA	Benchmark dataset	Mol2Vec, substructure fingerprints, and graph fingerprints (drugs) ProtBERT-BFD from ProtTrans and PSSM (targets)	TeM-DTBA	MSE KIBA-0.18 DAVIS-0.23
Voitsitskyi. et al. (2023) [42]	DAVIS and KIBA	Benchmark dataset	One-hot encoding	GNN	MSE DAVIS-0.28 KIBA-0.23
Mengmeng Gao et al. (2023) [43]	DrugBank, DAVIS, and KIBA	Target-specific and benchmark dataset	Molecular graph (drug) and 1D-CNN- embedding layer (target)	GraphormerDTI	AUC DrugBank-0.86 DAVIS-0.89 KIBA-0.92
Zhijian et al. (2022) [21]	DAVIS and KIBA	Benchmark dataset	2D molecule graphs (drugs) 128-dimensional feature (targets)	GDGRU-DTA	MSE KIBA-0.13 DAVIS-0.20
Souza et al. (2022) [19]	DAVIS and KIBA	Benchmark dataset	Encoder-decoder	CNN	MSE DAVIS-0.28 KIBA-0.21
Hua et al. (2022) [16]	DAVIS, KIBA and sc-PDB	Benchmark and target specific dataset	FCFPs and GNN-derived features (drug), AAE, and WE (target)	MFR-DTA	MSE DAVIS-0.22 KIBA-0.13
Alkhadrawi, et. al (2022) [29]	DAVIS, Glycyrrhetinic acid	Benchmark and, Target specific Dataset	Morgan, Pubchem, Daylight, RDkit, ESPF, and ErG (drug) Conjoint Triad, and AAC (target)	CNN, SVM	Morgan-Conjoint Triad AUROC-0.95
Shim et al. (2021) [44]	DAVIS and KIBA	Benchmark dataset	Tanimoto similarity matrix and Smith (drugs) Waterman similarity (targets)	SimCNN-DTA	MSE KIBA-0.27 DAVIS-0.31
Aleb et al. (2021) [45]	DAVIS and KIBA	Benchmark dataset	One-hot encoding	Multilevel attention model	MSE KIBA-0.13 DAVIS- 0.18
Abbasi et al. (2020) [5]	KIBA, DAVIS and BindingDB	Benchmark dataset	Embedding layer technique	DeepCDA	MSE KIBA-0.17 DAVIS-0.73 BindingDB-0.45
Rifaioglu, et al 2020 [46]	ChEMBL	Target-specific dataset	2D molecular images	DEEPScreen, CNN	Accuracy- 0.87

demonstrates the effectiveness of combining multiple molecular descriptors with ML and deep learning models for DTIs prediction, several limitations must be acknowledged. The analysis was confined to three datasets (PDAC, BindingDB,

TABLE 2. *(Continued.)* **Comparison of studies that used the AI algorithms for DTI prediction.**

You et al. (2019) [47]	Drugbank	Breast cancer (Target-specific dataset)	Chemical, topological, and geometrical properties (drugs), AAC, dipeptide composition, and tripeptide composition (targets)	LASSO-DNN, SVM, SLG regression	Accuracy -0.81 AUC- 0.89
He et al. (2017) [17]	DAVIS, KIBA, and Metz dataset	Benchmark dataset	Drug similarity Target similarity	SimBoost (Gradient boosting)	RMSE DAVIS- 0.24 Metz-0.16 KIBA- 0.204

MDF-DTA-Multi-dimensional fusion for drug target affinity prediction, TeM-DTBA-Time-efficient drug target binding affinity, ProtBERT-BFD- Protein Bidirectional Encoder Representation from Transformers-Big Fantastic Database, PSSM-Position-Specific Scoring Matrix, GNN-Graph neural networks, GraphormerDTI model-Graph transformer neural network, GDGRU-DTA-Graph based Gate Recurrent Unit drug and target affinity, FCFPs-circular fingerprints, AAE-Amino acid embeddings, WE-Word embeddings, MFR-DTA-Multi-functional and robust drug-target binding affinity prediction, ESPF-Explainable substructure partition fingerprint, ErG-2D -Pharmacophore descriptions for scaffold hopping, SimCNN-DTA-Sim convolutional neural network drug target affinity , DeepCDA-Deep combined drug affinity, LASSO-DNN-Least absolute shrinkage and selection operator-Deep neural network, SLG-standard logistic

and DAVIS), potentially limiting the generalizability of the findings to broader chemical and biological spaces. Using a small PDAC dataset with docking-based scores may limit model generalizability and risk overfitting. Additionally, the reliance on predefined descriptors may have restricted the ability to fully capture complex DTIs, as structural, dynamic, and contextual biological information (such as 3D conformations or pathway data) was not included. Moreover, the deep learning models, particularly the 1D-CNN, were used in this study. In future research, we will focus on integrating graph neural networks and transformer-based models to enable richer, data-driven representations of drugs and proteins. Incorporating 2D/3D descriptors, structural biology data, multi-omics profiles, and real-world clinical or pharmacogenomic information could significantly enhance prediction accuracy and translational relevance. We will incorporate the semi-supervised approaches to increase the sample count in future studies. Furthermore, adopting explainable AI techniques (*SHAP* or *LIME*) would make these models more transparent and informative, facilitating their application in drug discovery, repurposing, and precision medicine.

V. CONCLUSION

In conclusion, the study indicates that the 1D-CNN model can predict binding affinity for potential PDAC therapeutics, though with varying degrees of accuracy, depending on the descriptor combinations. Across all three datasets (PDAC, BindingDB, and DAVIS), the 1D-CNN model consistently outperformed both traditional ML models (RFR and XGBR) across all evaluation metrics (R^2, MSE, RMSE, and MAE). While XGBR showed comparable performance, especially in the BindingDB dataset, it generally lagged behind 1D-CNN. RFR exhibited the lowest predictive accuracy, particularly when using individual descriptor sets. In addition, combining multiple molecular descriptor types, such as Morgan, PubChem, AAC, Conjoint Triad, and Quasi-seq features, significantly improved model performance across datasets, highlighting the critical role of rich feature representation for drug-target affinity prediction. These findings suggest that descriptor and AI selection are crucial for accurate binding affinity prediction, and further research is needed to improve the models' predictive power for PDAC targets.

ACKNOWLEDGMENT

The authors would like to thank the support of the PARAM Shivay supercomputer facility of Indian Institute of Technology (BHU), Varanasi, India, used for this research.

REFERENCES

[1] S. Gupta and S. Shukla, "Limitations of immunotherapy in cancer," *Cureus*, vol. 14, Oct. 2022, Art. no. e30856.

[2] G. Lippi and C. Mattiuzzi, "The global burden of pancreatic cancer," *Arch. Med. Sci.*, vol. 16, no. 4, pp. 820–824, 2020.

[3] R. H. Gaidhani and G. Balasubramaniam, "An epidemiological review of pancreatic cancer with special reference to India," *Indian J. Med. Sci.*, vol. 73, pp. 99–109, Oct. 2020.

[4] C. Li and W.-Q. He, "Global prediction of primary liver cancer incidences and mortality in 2040," *J. Hepatol.*, vol. 78, no. 4, pp. e144–e146, Apr. 2023.

[5] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J. B. Ghasemi, and A. Masoudi-Nejad, "DeepCDA: Deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks," *Bioinformatics*, vol. 36, no. 17, pp. 4633–4642, May 2020.

[6] D. M. Gaber, S. S. Ibrahim, A. K. Awaad, Y. M. Shahine, S. Elmallah, H. S. Barakat, and N. I. Khamis, "A drug repurposing approach of atorvastatin calcium for its antiproliferative activity for effective treatment of breast cancer: In vitro and in vivo assessment," *Int. J. Pharmaceutics, X*, vol. 7, Apr. 2024, Art. no. 100249.

[7] T. U. Singh, S. Parida, M. C. Lingaraju, M. Kesavan, D. Kumar, and R. K. Singh, "Drug repurposing approach to fight COVID-19," *Pharmacolog. Rep.*, vol. 72, no. 6, pp. 1479–1508, Dec. 2020.

[8] M. E. Mswahili and Y.-S. Jeong, "Transformer-based models for chemical SMILES representation: A comprehensive literature review," *Heliyon*, vol. 10, no. 20, Oct. 2024, Art. no. e39038.

[9] J. Dong, D.-S. Cao, H.-Y. Miao, S. Liu, B.-C. Deng, Y.-H. Yun, N.-N. Wang, A.-P. Lu, W.-B. Zeng, and A. F. Chen, "ChemDes: An integrated web-based platform for molecular descriptor and fingerprint computation," *J. Cheminformatics*, vol. 7, no. 1, pp. 1–10, Dec. 2015.

[10] P. Carracedo-Reboredo, J. Liñares-Blanco, N. Rodríguez-Fernández, F. Cedrón, F. J. Novoa, A. Carballal, V. Maojo, A. Pazos, and C. Fernández-Lozano, "A review on machine learning approaches and trends in drug discovery," *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 4538–4558, Jan. 2021.

[11] A. Schulman, J. Rousu, T. Aittokallio, and Z. Tanoli, "Attention-based approach to predict drug–target interactions across seven target superfamilies," *Bioinformatics*, vol. 40, no. 8, Aug. 2024, Art. no. btae496.

[12] M. S. Bahia, O. Kaspi, M. Touitou, I. Binayev, S. Dhail, J. Spiegel, N. Khazanov, A. Yosipof, and H. Senderowitz, "A comparison between 2D and 3D descriptors in QSAR modeling based on bio-active conformations," *Mol. Informat.*, vol. 42, no. 4, Apr. 2023, Art. no. 2200186.

[13] Á. Orosz, K. Héberger, and A. Rácz, "Comparison of descriptor- and fingerprint sets in machine learning models for ADME-tox targets," *Frontiers Chem.*, vol. 10, Jun. 2022, Art. no. 852893.

[14] J. Sieg, C. W. Feldmann, J. Hemmerich, C. Stork, F. Sandfort, P. Eiden, and M. Mathea, "MolPipeline: A Python package for processing molecules with RDKit in scikit-learn," *J. Chem. Inf. Model.*, vol. 64, no. 24, pp. 9027–9033, Dec. 2024.

[15] M. Xiao, Q. Zheng, P. Popa, X. Mi, J. Hu, F. Zou, and B. Zou, "Drug molecular representations for drug response predictions: A comprehensive investigation via machine learning methods," *Sci. Rep.*, vol. 15, no. 1, p. 20, Jan. 2025.

[16] Y. Hua, X. Song, Z. Feng, and X. Wu, "MFR-DTA: A multi-functional and robust model for predicting drug–target binding affinity and region," *Bioinformatics*, vol. 39, no. 2, Feb. 2023, Art. no. btad056.

[17] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, and M. Ester, "SimBoost: A read-across approach for predicting drug–target binding affinities using gradient boosting machines," *J. Cheminformatics*, vol. 9, no. 1, pp. 1–14, Dec. 2017.

[18] H. Atas Guvenilir and T. Doğan, "How to approach machine learning-based prediction of drug/compound–target interactions," *J. Cheminformatics*, vol. 15, no. 1, p. 16, Feb. 2023.

[19] S. D'Souza, K. V. Prema, S. Balaji, and R. Shah, "Deep learning-based modeling of drug–target interaction prediction incorporating binding site information of proteins," *Interdiscipl. Sci., Comput. Life Sci.*, vol. 15, no. 2, pp. 306–315, Jun. 2023.

[20] Q. Zhao, G. Duan, M. Yang, Z. Cheng, Y. Li, and J. Wang, "AttentionDTA: Drug–target binding affinity prediction by sequence-based deep learning with attention mechanism," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 2, pp. 852–863, Mar. 2023.

[21] L. Zhijian, J. Shaohua, L. Yigao, and G. Min, "GDGRU-DTA: Predicting drug-target binding affinity based on GNN and double GRU," 2022, *arXiv:2204.11857*.

[22] W. Torng and R. B. Altman, "Graph convolutional neural networks for predicting drug-target interactions," *J. Chem. Inf. Model.*, vol. 59, no. 10, pp. 4131–4149, Oct. 2019.

[23] K. Shimada, J. A. Bachman, J. L. Muhlich, and T. J. Mitchison, "ShinyDepMap, a tool to identify targetable cancer genes and their functional connections from cancer dependency map data," *eLife*, vol. 10, Feb. 2021, Art. no. e57116.

[24] T. Liu, L. Hwang, S. K. Burley, C. I. Nitsche, C. Southan, W. P. Walters, and M. K. Gilson, "BindingDB in 2024: A FAIR knowledgebase of protein-small molecule binding data," *Nucleic Acids Res.*, vol. 53, no. D1, pp. D1633–D1644, Jan. 2025.

[25] G. Nicola, M. R. Berthold, M. P. Hedrick, and M. K. Gilson, "Connecting proteins with drug-like compounds: Open source drug discovery workflows with BindingDB and KNIME," *Database*, vol. 2015, Jan. 2015, Art. no. bav087.

[26] M. Abdel-Basset, H. Hawash, M. Elhoseny, R. K. Chakrabortty, and M. Ryan, "DeepH-DTA: Deep learning for predicting drug-target interactions: A case study of COVID-19 drug repurposing," *IEEE Access*, vol. 8, pp. 170433–170451, 2020.

[27] S. Zhong and X. Guan, "Count-based Morgan fingerprint: A more efficient and interpretable molecular representation in developing machine learning-based predictive regression models for water contaminants' activities and properties," *Environ. Sci. Technol.*, vol. 57, no. 46, pp. 18193–18202, Jul. 2023.

[28] E. Fernández-De Gortari, C. R. García-Jacas, K. Martinez-Mayorga, and J. L. Medina-Franco, "Database fingerprint (DFP): An approach to represent molecular databases," *J. Cheminformatics*, vol. 9, no. 1, pp. 1–9, Dec. 2017.

[29] A. M. Alkhadrawi, Y. Wang, and C. Li, "In-silico screening of potential target transporters for glycyrrhetinic acid (GA) via deep learning prediction of drug-target interactions," *Biochem. Eng. J.*, vol. 181, Apr. 2022, Art. no. 108375.

[30] J. Shen, J. Zhang, X. Luo, W. Zhu, K. Yu, K. Chen, Y. Li, and H. Jiang, "Predicting protein–protein interactions based only on sequences information," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 11, pp. 4337–4341, Mar. 2007.

[31] M. Taheri, M. R. Keyvanpour, and M. S. Mousavi, "Improving drug-target interaction prediction using enhanced feature selection," in *Proc. 15th Int. Conf. Inf. Knowl. Technol. (IKT)*, Dec. 2024, pp. 157–161.

[32] S. A. Ong, H. H. Lin, Y. Z. Chen, Z. R. Li, and Z. Cao, "Efficacy of different protein descriptors in predicting protein functional families," *BMC Bioinf.*, vol. 8, no. 1, pp. 1–14, Dec. 2007.

[33] X. Lin, "DeepGS: Deep representation learning of graphs and sequences for drug-target binding affinity prediction," in *Proc. ECAI*, Jan. 2020, pp. 1301–1308.

[34] S. A. Mir, R. K. Meher, I. Baitharu, and B. Nayak, "Molecular dynamic simulation, free binding energy calculation of Thiazolo-[2,3-b]quinazolinone derivatives against EGFR-TKD and their anticancer activity," *Results Chem.*, vol. 4, Jan. 2022, Art. no. 100418.

[35] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: A classification and regression tool for compound classification and QSAR modeling," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 1947–1958, Nov. 2003.

[36] C. Zhuo, J. Gao, A. Li, X. Liu, and Y. Zhao, "A machine learning method for RNA–small molecule binding preference prediction," *J. Chem. Inf. Model.*, pp. 7386–7397, Sep. 2024.

[37] M. F. Sanner, L. Dieguez, S. Forli, and E. Lis, "Improving docking power for short peptides using random forest," *J. Chem. Inf. Model.*, vol. 61, no. 6, pp. 3074–3090, Jun. 2021.

[38] C. Chen, H. Shi, Z. Jiang, A. Salhi, R. Chen, X. Cui, and B. Yu, "DNN-DTIs: Improved drug-target interactions prediction using XGBoost feature selection and deep neural network," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104676.

[39] V. P. Mishra, Y. N. Singh, F. Khan, and M. K. Dutta, "SeqDPI: A 1D-CNN approach for predicting binding affinity of kinase inhibitors," *J. Comput. Chem.*, vol. 46, no. 1, pp. e2751–8, Jan. 2025.

[40] A. Ranjan, A. Bess, C. Alvin, and S. Mukhopadhyay, "MDF-DTA: A multi-dimensional fusion approach for drug-target binding affinity prediction," *J. Chem. Inf. Model.*, vol. 64, no. 13, pp. 4980–4990, Jul. 2024.

[41] T. Liyaqat, T. Ahmad, and C. Saxena, "TeM-DTBA: Time-efficient drug target binding affinity prediction using multiple modalities with lasso feature selection," *J. Comput.-Aided Mol. Des.*, vol. 37, no. 12, pp. 573–584, Dec. 2023.

[42] T. Voitsitskyi, R. Stratiichuk, I. Koleiev, L. Popryho, Z. Ostrovsky, P. Henitsoi, I. Khropachov, V. Vozniak, R. Zhytar, D. Nechepurenko, S. Yesylevskyy, A. Nafiiev, and S. Starosyla, "3DProtDTA: A deep learning model for drug-target affinity prediction based on residue-level protein graphs," *RSC Adv.*, vol. 13, no. 15, pp. 10261–10272, 2023.

[43] M. Gao, D. Zhang, Y. Chen, Y. Zhang, Z. Wang, X. Wang, S. Li, Y. Guo, G. I. Webb, A. T. N. Nguyen, L. May, and J. Song, "GraphormerDTI: A graph transformer-based approach for drug-target interaction prediction," *Comput. Biol. Med.*, vol. 173, May 2024, Art. no. 108339.

[44] J. Shim, Z.-Y. Hong, I. Sohn, and C. Hwang, "Prediction of drug–target binding affinity using similarity-based convolutional neural network," *Sci. Rep.*, vol. 11, no. 1, p. 4416, Feb. 2021.

[45] N. Aleb, "Multilevel attention models for drug target binding affinity prediction," *Neural Process. Lett.*, vol. 53, no. 6, pp. 4659–4676, Dec. 2021.

[46] A. S. Rifaioglu, E. Nalbat, V. Atalay, M. J. Martin, R. Cetin-Atalay, and T. Doğan, "DEEPScreen: High performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations," *Chem. Sci.*, vol. 11, no. 9, pp. 2531–2557, 2020.

[47] J. You, R. D. McLeod, and P. Hu, "Predicting drug-target interaction network using deep learning model," *Comput. Biol. Chem.*, vol. 80, pp. 90–101, Jun. 2019.

[48] M. Kalemati, M. Zamani Emani, and S. Koohi, "BiComp-DTA: Drug-target binding affinity prediction through complementary biological-related and compression-based featurization approach," *PLOS Comput. Biol.*, vol. 19, no. 3, Mar. 2023, Art. no. e1011036.

[49] B.-Y. Ji, Z.-H. You, H.-J. Jiang, Z.-H. Guo, and K. Zheng, "Prediction of drug-target interactions from multi-molecular network based on LINE network representation method," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–11, Dec. 2020.

[50] H. Ismail, N. H. A. H. Malim, S. Z. M. Zobir, and H. A. Wahab, "Comparative studies on drug-target interaction prediction using machine learning and deep learning methods with different molecular descriptors," *Artif. Intell., Drug Discovery Clin. Pharmacol.*, vol. 12, pp. 1–6, Mar. 2021.

[51] Z. Chu, F. Huang, H. Fu, Y. Quan, X. Zhou, S. Liu, and W. Zhang, "Hierarchical graph representation learning for the prediction of drug-target binding affinity," *Inf. Sci.*, vol. 613, pp. 507–523, Oct. 2022.

PRAGYA received the B.Tech. degree in biotechnology from the College of Basic Science and Humanities, DRPCAU, Pusa, in 2022. She is currently pursuing the Ph.D. degree with the School of Biomedical Engineering, Indian Institute of Technology (BHU), Varanasi, India. Her research interests include genomics, transcriptomics, machine learning, deep learning, and drug repurposing.

A. AMALIN PRINCE (Senior Member, IEEE) received the Ph.D. degree from BITS Pilani, Pilani, India, in 2011. He is currently a Professor and the Head of the Department of Electrical and Electronics Engineering, BITS Pilani, Goa, India. His research interests include FPGA-based system design, hardware accelerated data processing, affective computing, and applications of artificial intelligence.

JAC FREDO AGASTINOSE RONICKOM received the degree in electrical and electronics engineering and the Ph.D. degree in neuroinformatics from Anna University, Chennai, in 2015. He continued his postdoctoral research at the Brain Development Imaging Laboratory, San Diego State University, USA, in 2016. He joined as a Research Fellow at Nanyang Technological University, Singapore, in 2018, and got Advanced Research Opportunities (AROP) Scholarship at RWTH Aachen, Germany, in 2020. He is currently an Associate Professor with the School of Biomedical Engineering, Indian Institute of Technology (BHU), Varanasi, India. His research interests include bio-signal and image processing, biomedical instrumentation, computational neuroscience, and neuroinformatics.