RESEARCH PAPER

Artificial Intelligence Framework for Drug-Target Affinity Prediction Using Deep Neural Networks and Integrated Molecular and Protein Feature Engineering

Hency Juliet A[1], R. Padmajavalli[2], A. Jerline Amutha[3], R. Christy Puspaleela[4], B.Shathya [5], N. Legapriyadharshini[6]

[1] Professor, Department of Computer Applications, Saveetha College of Liberal Arts and Sciences, SIMATS, Chennai, Tamil Nadu India, hencyjulieta.sclas@saveetha.com

[2] Professor & Head, Department of Computer Science, Bhaktavatsalam Memorial College for Women, Chennai, padmahari2002@yahoo.com

[3] Associate Professor, Department of Data Science, Women's Christian College, Chennai, jerlineamutha@wcc.edu.in

[4] Associate Professor, P.G Department of Computer Science and Technology, Women's Christian College, Chennai, christypushpaleela@wcc.edu.in

[5]Asst. Professor, Department of Computer Applications, Ethiraj College for Women, Chennai, Tamil Nadu, India. shathya04@gmail.com

[6] Associate Professor, Department of Computer Applications, Saveetha College of Liberal Arts, SIMATS, Chennai, India. legapriyadharshinin.sclas@saveetha.com

Abstract:

Objective: To develop an artificial intelligence framework that integrates deep neural networks with molecular and protein feature engineering to accurately predict drug-target affinity, thereby supporting efficient computational drug discovery and therapeutic candidate identification. The proposed study employed a publicly available drug-target affinity dataset consisting of molecular SMILES representations, protein amino acid sequences, and experimentally determined affinity values, partitioned into training, validation, and testing subsets for model development. **Material and Methods**: A publicly available drug-target affinity dataset containing training, validation, and testing subsets was employed, in which molecular SMILES representations and protein sequences were converted into integrated numerical descriptors through feature engineering and standardized before model training. Subsequently, a Deep Neural Network (DNN) was developed to capture the complex relationships between the engineered molecular and protein features for accurate drug-target affinity prediction. Extracting representative features from molecular SMILES representations and protein sequences remains a major challenge in accurately predicting drug–target affinity. Capturing the complex nonlinear relationships between drug molecules and target proteins is further complicated by the high dimensionality and intricate nature of biological data. **Results**: To mitigate these challenges, a Deep Neural Network integrated with molecular and protein feature engineering was employed to effectively model complex drug–target interactions and enhance the accuracy of drug–target affinity prediction. The proposed Deep Neural Network efficiently learns complex nonlinear patterns from integrated molecular and protein features, thereby enhancing the accuracy of drug–target affinity prediction. The proposed Deep Neural Network achieved consistent performance across the training, validation, and testing datasets, with Pearson correlation coefficients of 0.6586, 0.6465, and 0.6343, respectively. The obtained R^2 scores of 0.2585, 0.2163, and 0.1599 demonstrate the model's capability to effectively predict drug–target affinity using integrated molecular and protein features. **Conclusion**: The proposed Artificial Intelligence framework effectively combined Deep Neural Networks with integrated molecular and protein feature engineering for drug–target affinity prediction. The findings indicate that the framework provides a reliable and efficient approach for modelling drug–target interactions, thereby supporting computational drug discovery.

Keywords: Artificial Intelligence, Drug–Target Affinity Prediction, Deep Neural Networks, Molecular Feature Engineering, Protein Sequence Analysis, Computational Drug Discovery, Healthy lives, Well beings, Medical.

How to cite this article: Hency Juliet A, Padmajavalli R, Jerline Amutha A, Christy Puspaleela R, Shathya B, Legapriyadharshini N. Artificial intelligence framework for drug-target affinity prediction using deep neural networks and integrated molecular and protein feature engineering. Int J Drug Deliv Technol. 2026;16(75s): 203-217. DOI: 10.25258/ijddt.16.75s.24

1. INTRODUCTION

Drug–target affinity prediction plays a crucial role in computational drug discovery by facilitating the identification of potential interactions between drug compounds and target proteins while minimizing the time, cost, and resources required for conventional laboratory-based screening [1]. The increasing availability of biomedical data, including molecular SMILES representations and protein amino acid sequences, has accelerated the adoption of artificial intelligence techniques for extracting biologically relevant information [2]. Nevertheless, the heterogeneous characteristics of molecular and protein data, along with the complex nonlinear relationships governing drug–target interactions, continue to pose significant challenges for accurate affinity prediction [3].

Artificial intelligence, particularly deep learning, has gained considerable attention for its ability to effectively analyze complex biological datasets and uncover hidden interaction patterns. Deep Neural Networks (DNNs) have demonstrated superior capability in learning nonlinear relationships from high-dimensional data, making them well suited for drug–target affinity prediction tasks [4]. Moreover, the integration of engineered molecular features derived from SMILES representations with protein sequence-based descriptors provides a richer representation of drug–target pairs, thereby enhancing the model's ability to characterize their biological interactions [5].

To address these challenges, this study presents an Artificial Intelligence Framework for Drug–Target Affinity Prediction using Deep Neural Networks and Integrated Molecular and Protein Feature Engineering [6]. The proposed framework converts molecular SMILES representations and protein sequences into standardized numerical descriptors through feature engineering, followed by Deep Neural Network-based learning to model the underlying drug–target relationships [7]. By combining integrated feature representations with deep learning, the proposed approach aims to improve the accuracy and robustness of drug–target affinity prediction, thereby contributing to efficient computational drug discovery and the identification of potential therapeutic candidates [8].

2.1 RELATED WORK

Drug–target affinity prediction has become an important area of research because of its potential to accelerate computational drug discovery and facilitate the identification of promising therapeutic compounds. A variety of artificial intelligence and deep learning techniques have been developed to enhance prediction performance by effectively modelling complex interactions between molecular and protein features. [9] Limbu and Dakshanamurthy introduced a hybrid deep learning model for predicting protein–ligand binding affinity and facilitating de novo drug design. The proposed framework employed a hybrid neural network to capture intricate protein–ligand relationships. The results demonstrated enhanced prediction accuracy and supported efficient drug discovery. [10] Ma et al. proposed a drug–target affinity prediction framework by integrating protein knowledge derived from biological networks. The study utilized deep learning with network-based protein representations to strengthen feature extraction. The proposed method improved affinity prediction by effectively incorporating biological network information. [11] Kumar and Singh developed an optimized deep learning model for predicting drug–target interactions and binding affinity. The framework adopted an optimized neural network architecture to model complex biological interactions. The findings showed improved predictive accuracy and robust affinity estimation.

[12] Çevik et al. developed a sequence-based framework for drug–target binding prediction without requiring three-dimensional structural information. The study evaluated machine learning, deep learning, and ensemble learning models using sequence-derived features. The results revealed that deep learning and ensemble techniques produced superior predictive performance. [13] Zhang et al. presented a deep learning approach that incorporated protein secondary structure information for drug–target binding affinity prediction. The proposed model integrated sequence and structural descriptors to generate enriched feature representations. The experimental findings confirmed improved prediction accuracy through secondary structure integration. [14] Wang et al. reviewed the application of artificial intelligence techniques in drug development with emphasis on drug–target interactions. The study examined various machine learning and deep learning methodologies employed in computational drug discovery. The review highlighted the significant contribution of artificial intelligence to improving drug discovery efficiency.

[15] Zhang et al. conducted a comprehensive survey of graph neural network-based approaches for drug–target interaction and affinity prediction. The review analysed different graph neural network architectures and their biomedical applications. The study emphasized the effectiveness of graph-based models in enhancing prediction performance. [16] Hudson explored the application of machine learning-based data integration techniques in drug discovery and molecular biology. The study focused on integrating heterogeneous biological datasets to improve analytical performance. The findings demonstrated that data integration enhances predictive modelling and biological knowledge extraction. [17] Li et al. proposed a neural network-based multi-feature fusion framework for drug–target prediction. The approach integrated multiple biological feature representations to strengthen the learning process. The proposed model achieved improved prediction performance through effective feature fusion.

[18] Huang et al. investigated the integration of artificial intelligence and deep learning techniques for drug discovery and target identification. The study reviewed recent AI-driven methodologies applied in pharmaceutical research. The findings indicated that deep learning substantially improves the effectiveness of computational drug discovery. [19] Xu et al. proposed a graph unseen node embedding learning network for drug–target affinity prediction. The framework utilized graph embedding techniques to learn meaningful representations of previously unseen nodes. The experimental results demonstrated enhanced prediction accuracy and better model generalization. [20] Liu introduced a drug–target affinity prediction method based on the consistent expression of heterogeneous biological data. The proposed framework integrated diverse biological information to improve feature representation. The results showed improved affinity prediction by effectively leveraging heterogeneous datasets.

[21] Zhou et al. developed the XG-DTA framework for drug–target affinity prediction by integrating drug molecular graphs with protein sequences. The proposed method combined graph neural networks and the XLNet architecture for comprehensive feature learning. The framework achieved improved prediction accuracy through effective representation of drugs and proteins. [22] Yang et al. proposed GraphCL-DTA, a graph contrastive learning framework for drug–target binding affinity prediction. The model employed graph contrastive learning together with molecular semantic information to learn discriminative feature representations. The experimental evaluation demonstrated improved prediction accuracy and robust molecular feature learning. Despite the considerable progress achieved by existing artificial intelligence and deep learning approaches, effectively capturing the intricate nonlinear interactions between molecular and protein features while ensuring computational efficiency remains a persistent challenge in drug–target affinity prediction [23]. To address this limitation, the present study introduces an Artificial Intelligence Framework that combines Deep Neural Networks with integrated molecular and protein feature engineering to learn informative feature representations and enhance the accuracy and reliability of drug–target affinity prediction [24].

2.2 Existing System Overview

Conventional drug–target affinity prediction methods predominantly employ machine learning and deep learning algorithms to estimate the interaction strength between drug compounds and target proteins using molecular and biological characteristics [25]. While these approaches have improved the efficiency of computational drug discovery through the utilization of molecular SMILES representations, protein sequences, and feature extraction methods, they often struggle to accurately model the intricate nonlinear relationships present in heterogeneous biological data [26]. In addition, inadequate integration of molecular and protein features and the challenges associated with learning high-dimensional representations can limit predictive performance and model generalizability, thereby emphasizing the need for more effective artificial intelligence-driven frameworks.

2.2.1 *Difficulty in Capturing Complex Biological Relationships*

Existing drug–target affinity prediction approaches often struggle to effectively capture the intricate nonlinear interactions between molecular SMILES representations and protein sequence information. The complex nature of biological interactions and the high dimensionality of the input data can limit the capability of conventional machine learning and deep learning models, resulting in suboptimal prediction accuracy and reduced robustness.

2.2.2 *Inadequate Integration of Heterogeneous Features*

Many existing methods exhibit limitations in integrating molecular and protein features into a unified and informative representation for affinity prediction. Consequently, the models may fail to fully exploit the complementary information present in heterogeneous biological data, thereby affecting feature learning, computational efficiency, and the overall predictive performance of drug–target affinity prediction systems.

2.3 Proposed System

The proposed framework introduces an Artificial Intelligence-based approach for drug–target affinity prediction by integrating Deep Neural Networks (DNNs) with engineered molecular and protein features to effectively characterize drug–protein interactions. In the proposed methodology, molecular SMILES representations and protein amino acid sequences are converted into standardized numerical descriptors through feature engineering, followed by feature integration and normalization to construct comprehensive representations of drug–target pairs [27]. The resulting feature vectors are then utilized to train a Deep Neural Network capable of capturing complex nonlinear relationships within high-dimensional biological data, thereby enabling accurate affinity prediction [28]. By leveraging integrated feature engineering and deep learning, the proposed framework improves prediction accuracy, model robustness, and computational efficiency, making it a reliable solution for computational drug discovery and therapeutic candidate identification.

2.3.1 Effective Modelling of Complex Drug–Target Interactions

The proposed Artificial Intelligence framework employs Deep Neural Networks together with engineered molecular and protein features to effectively model the complex nonlinear interactions between drug molecules and target proteins. By learning rich and discriminative representations from high-dimensional biological data, the framework enhances prediction accuracy, improves

model robustness, and provides more reliable drug–target affinity estimation [29].

2.3.2 Unified Representation of Molecular and Protein Features

The proposed framework integrates molecular descriptors extracted from SMILES representations with protein sequence-derived features to construct a comprehensive and unified feature representation for each drug–target pair. This integrated feature engineering strategy effectively captures complementary chemical and biological information, leading to improved feature representation, increased computational efficiency, and enhanced overall performance in drug–target affinity prediction [30].

3. MATERIAL AND METHODS

3.1 Research Area

Drug–target affinity prediction has become a significant research domain in computational drug discovery, aiming to accurately estimate the interaction strength between drug compounds and target proteins through computational approaches. The increasing availability of molecular SMILES representations and protein sequence data has encouraged the widespread adoption of artificial intelligence techniques for analysing complex biological information, enabling faster and more cost-effective identification of promising therapeutic candidates compared with conventional experimental methods.

Advances in artificial intelligence, particularly Deep Neural Networks (DNNs), have provided effective solutions for modelling the intricate nonlinear relationships inherent in high-dimensional molecular and protein data. Recent studies have increasingly focused on integrating engineered molecular descriptors with protein sequence-derived features to construct informative feature representations that improve affinity prediction. Building upon these developments, the present research introduces an Artificial Intelligence Framework that integrates molecular and protein feature engineering with Deep Neural Networks to enhance the accuracy, robustness, and computational efficiency of drug–target affinity prediction.

3.2 Dataset Description

The proposed framework employs a publicly available drug–target affinity dataset containing 52,274 drug–target interaction instances, which are divided into 36,592 training samples, 5,227 validation samples, and 10,455 testing samples for model development and evaluation [31]. Each instance includes a unique drug identifier, the molecular SMILES representation of the drug, a target protein identifier, the corresponding protein amino acid sequence, and an experimentally measured binding affinity value (Y). The molecular SMILES strings and protein sequences are utilized as the primary input features, whereas the binding affinity value serves as the target variable for supervised learning. By incorporating diverse molecular and protein characteristics, the dataset enables the proposed Deep Neural Network framework to effectively learn complex drug–target interactions through integrated molecular and protein feature engineering, thereby facilitating accurate affinity prediction**.** Table 1 presents the overview of the Drug–Target Affinity dataset, including the dataset size, molecular and protein data representations, target output variable, and the allocation of samples for training, validation, and testing used in the proposed framework. Fig. 3.1 presents a representative sample of the training dataset, illustrating drug–target pairs along with their corresponding molecular SMILES, protein amino acid sequences, and experimentally determined binding affinity values used to train the proposed Deep Neural Network framework. Fig. 3.2 depicts a representative subset of the testing dataset, showcasing drug–target pairs along with their associated molecular SMILES, protein amino acid sequences, and experimentally measured binding affinity values utilized to assess the predictive capability and generalization performance of the proposed Deep Neural Network framework. Fig. 3.3 presents a representative subset of the validation dataset, illustrating drug–target pairs together with their corresponding molecular SMILES, protein amino acid sequences, and experimentally measured binding affinity values used to validate the predictive performance and generalization capability of the proposed Deep Neural Network framework.

Table 1 Drug–Target Affinity dataset

Parameter	**Value**
Dataset	Drug–Target Affinity Dataset
Total Samples	52,274
Drug Representation	Molecular SMILES
Protein Representation	Protein Amino Acid Sequence
Output	Drug–Target Binding Affinity (Y)
Training Samples	36,592
Validation Samples	5,227
Testing Samples	10,455

	A	B	C	D	E
1	Drug_ID	Drug	Target ID	Target	Y
2	444607	Cc1ccc(CNS(=O)(=O)c2ccc(S(N)(=O)=O)s2)cc1	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAK	0.46
3	4316	COc1ccc(CNS(=O)(=O)c2ccc(S(N)(=O)=O)s2)cc1	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAK	0.49
4	4293	NS(=O)(=O)c1ccc(S(=O)(=O)NCc2ccco2)s1	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAK	0.83
5	4369102	CCN[C@H]1CN(CCOC)S(=O)(=O)c2sc(S(N)(=O)=O)cc21	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAK	0.32
6	3013848	CN[C@H]1CN(c2cccc(OC)c2)S(=O)(=O)c2sc(S(N)(=O)=O)cc21	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAK	0.1
7	4369101	CN[C@@H]1CN(c2cccc(OC)c2)S(=O)(=O)c2sc(S(N)(=O)=O)cc21	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAK	1.7
8	1514	COc1cccc(N2C=Cc3cc(S(N)(=O)=O)sc3S2(=O)=O)c1	P00918	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAK	0.13
9	4347	NS(=O)(=O)c1ccc(C(=O)NCc2ccccc2)cc1		SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKY	5.6
10	4290	NS(=O)(=O)c1ccc(C(=O)NCc2ccccc2F)cc1		SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKY	2.3
11	1700	NS(=O)(=O)c1ccc(C(=O)NCc2ccccc(F)c2F)cc1		SHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKY	1.6

Fig 3.1 Sample Training data - drug–target pairs [31]

	A	B	C	D	E
1	Drug_ID	Drug	Target_ID	Target	Y
2	9951964	O=C(c1ccc(F)cc1)C1CCN(CCC2Cc3ccc(F)ccc3C2=O)CC1	P14842	MEILCEDNISLSSIPNSLMQLG	7.59
3	24889392	CC(C)(C)c1cc(NC(=O)Nc2ccc(-c3cn4c(n3)sc3ccc(OCCN5CCOCC5)ccc34)cc2)n	O14733	MAASSLEQKLSRLEAKLKQE	10000
4	153999	CN(C)C[C@@H]1CCn2cc(c3ccccc32)C2=C(C(=O)NC2=O)c2cn(c3ccccc23)CC	P33981	MESEDLSGRELTIDSIMNKVR	10000
5	54756927	O=C(CCCCCNC(=O)COc1ccc(/C=C/c2ccc3n[B-](F)(F)[N+]2=C(c4ccs4)C=C	P07550	MGQPGNGSAFLLAPNGSHAP	0.7413
6	44259	CN[C@@H]1C[C@H]2O[C@@](C)([C@@H]1OC)n1c3ccccc3c3c4c(c5c6ccccc	Q9P289	MAHSPVAVQVPGMQNNIAD	140
7	15578577	COC(=O)CCNCC(O)COc1ccc2ccccc12		MGAGVLALGASEPCNLSSTA	79
8	11676971	CN(C)CC(=O)N1CCC(c2ccc(NC(=O)c3ncc(C#N)[nH]3)c(C3=CCCCC3)c2)CC1	Q02156	MVVFNGLLKIKICEAVSLKPT	10000
9	96869421	NCCN1C(=O)/C(=C/c2ccc3c(c2)OCO3)SC1=S		MAAAAAAAGPEMVRGQVFDV	8100
10	91898531	CC(C)=CCC/C(C)=C/C[C@@]12C[C@H](CC=C(C)C)C(C)(C)[C@@](CC=C(C	P42224	MSQWYELQQLDSKFLEQVH	4670
11	11712649	O=C(O)c1ccc(Nc2ncc3c(n2)-c2ccc(Cl)cc2C(c2ccc(F)cccc2F)=NC3)cc1	Q9BRS2	MDYRRLLMSRVVPGQFDDA	10000

Fig 3.2 Sample Test data - drug–target pairs with their associated molecular SMILES [31]

	A	B	C	D	E
1	Drug_ID	Drug	Target_ID	Target	Y
2	5287969	CN1CC[C@H](c2c(O)cc(O)c3c(=O)cc(-c4ccccc4Cl)oc23)[C@H](C	Q15303	MKPATGLWVWVSLLVAAGTVQPSDSQSVCAGT	5000
3	11314340	Cc1n[nH]c2ccc(-c3ncnc(OC[C@@H]3N)c4ccccc43)cc12	Q04759	MSPFLRIGLSNFDCGSCQSCQGEAVNPYCAVLVI	12
4	15462117	CN(C)C[C@@H](c1ccc2ccccc2c1)[C@@H](O)C(C)(C)C	P23975	MLLARMNPQVQPENNGADTGPEQPLRARKTAE	30
5	3038525	O=c1ncn2nc(Sc3ccc(F)cc3F)ccc2c1-c1c(Cl)cccc1Cl	P15056	MAALSGGGGGAEPGQALFNGDMEPEAGAGAG	1
6	153999	CN(C)C[C@@H]1CCn2cc(c3ccccc32)C2=C(C(=O)NC2=O)c2cn(c3	P42680	MNFNTLLEEILIKRSQQKKTSPLNYKERLFVLTI	10000
7	73776640	CC(C)(C)Nc1c(F)c(S(N)(=O)=O)c(F)c1S(=O)(=O)CCc1ccccc1	Q16790	MAPLCPSPWLPLLIPAPAPGLTVQLLLSLLLLVP	6.4
8	10138259	Cc1[nH]c(/C=C2/C(=O)Nc3ccc(F)cc32)c(C)c1C(=O)NC[C@H](O)C	P11802	MATSRYEPVAEIGVGAYGTVYKARDPHSGHFV	7100
9	11664012	Cc1ccnc(NS(=O)(=O)c2ccc(NC(=O)c3ccc4nc(-c5cccs5)c(-c5cccs5	P62937	MVNPTVFFDIAVDGEPLGRVSFELFADKVPKTA	41000
10	6445562	CCOc1cc2ncc(C#N)c(Nc3ccc(F)c(Cl)c3)c2cc1NC(=O)/C=C/CN(C)	Q99683	MSTEADEGITFSVPPFAPSGFCTIPEGGICRRGGA	10000
11	5705529	O=C(O[C@@H]1[C@@H]2O[C@]2(CO)[C@H]2[C@H](O[C@@	P07900	MPEETQTQDQPMEEEEVETFAFQAEIAQLMSLII	16

Fig 3.3 Sample Validation data - drug–target pairs [31]

3.3 Proposed Architecture

The proposed architecture presents an Artificial Intelligence-based framework for drug–target affinity prediction by integrating Deep Neural Networks (DNNs) with molecular and protein feature engineering to effectively characterize drug–protein interactions. Initially, a publicly available drug–target affinity dataset containing molecular SMILES representations, protein amino acid sequences, and experimentally measured binding affinity values is utilized for model development. The SMILES strings and protein sequences are independently transformed into informative numerical representations through feature engineering, after which the extracted molecular and protein features are integrated to construct a unified feature vector. The integrated features are subsequently normalized to ensure consistent feature scaling and to enhance the efficiency and stability of Deep Neural Network training. The normalized feature vectors are then provided as input to the DNN, which learns the complex nonlinear relationships between drug compounds and target proteins to predict binding affinity accurately. Finally, the effectiveness of the proposed framework is assessed using standard regression performance metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Pearson Correlation Coefficient, and the coefficient of determination (R^2), demonstrating its capability to support reliable computational drug discovery and therapeutic candidate identification. Fig. 3.4 presents the proposed Artificial Intelligence framework for drug–target affinity prediction, illustrating the complete processing pipeline from dataset preparation through feature engineering, feature integration, normalization, Deep Neural Network-based affinity prediction, and final performance evaluation.

3.4 Molecular Feature Engineering

The molecular feature engineering process transforms the molecular SMILES (Simplified Molecular Input Line Entry System) representations of drug compounds into informative numerical descriptors suitable for machine learning. Initially, each SMILES string is processed to extract relevant molecular characteristics, including physicochemical and structural properties that describe the chemical composition and topology of the drug molecule. In addition, molecular fingerprints are generated to encode the presence or absence of specific substructures and functional groups as fixed-length binary or numerical vectors, thereby preserving essential structural information. The extracted descriptors and molecular fingerprints are subsequently combined and numerically encoded to construct a comprehensive molecular feature vector, which serves as the drug representation for Deep Neural Network training. Mathematically, the molecular feature vector can be represented as

M= {m1, m2, m3, …..m ----------(1)

where M denotes the molecular feature vector and m represents the i^{th} molecular descriptor or fingerprint feature.

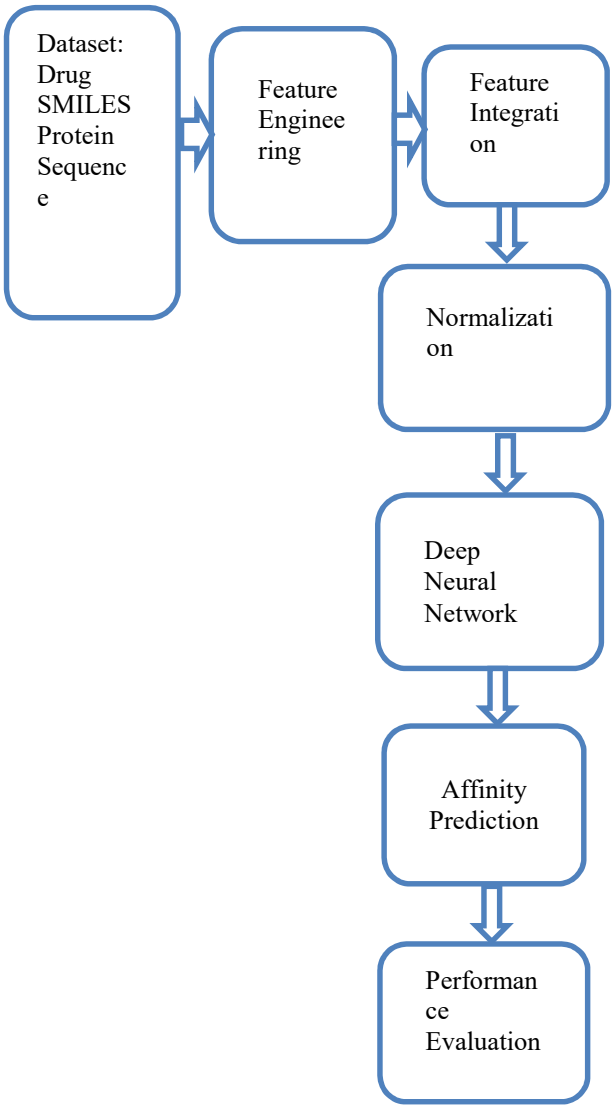

Fig. 3.4 proposed Artificial Intelligence framework for drug–target affinity prediction

3.5 Protein Feature Engineering

Protein feature engineering aims to convert protein amino

acid sequences into informative numerical representations that capture their biological and structural characteristics. Each protein sequence is initially encoded using sequence-based descriptors derived from the amino acid composition and physicochemical properties of individual residues. Feature extraction techniques are then employed to generate discriminative protein descriptors that preserve meaningful sequence information while reducing redundancy. These encoded protein features provide a compact representation of the target protein and facilitate effective learning of drug–protein interactions. The resulting protein feature vector is mathematically expressed as

P= {p1, p2, p3, ….p} ----------(2)

where P represents the protein feature vector and p_j denotes the j^{th} protein descriptor extracted from the amino acid sequence. The generated molecular and protein feature vectors are subsequently integrated to form the input representation for the proposed Deep Neural Network-based drug–target affinity prediction framework.

3.6 Integrated Feature Engineering

The integrated feature engineering stage is designed to generate a unified representation by combining the extracted molecular and protein features for each drug–target pair. The molecular feature vector obtained from the SMILES representation and the protein feature vector derived from the amino acid sequence are concatenated to capture complementary chemical and biological information within a single feature space. This integrated representation enables the proposed Deep Neural Network to effectively learn the complex nonlinear relationships between drugs and target proteins, thereby improving the accuracy and robustness of drug–target affinity prediction. The integrated feature vector is represented as

$$F = [D \;||\; P], \qquad \text{----------(3)}$$

where Ddenotes the molecular feature vector, Prepresents the protein feature vector, and Fcorresponds to the concatenated feature vector used as the input to the Deep Neural Network.

3.7 Data Preprocessing

Data preprocessing is performed to prepare the integrated molecular and protein features for efficient Deep Neural Network training. Initially, the dataset is inspected to detect and address missing or incomplete values, ensuring the consistency and integrity of the input data. Subsequently, feature scaling is applied to minimize variations in feature magnitude, and the Standard Scaler technique is employed to standardize each feature by transforming it to a common scale with zero mean and unit variance. Finally, the standardized dataset is divided into training, validation, and testing subsets to facilitate model training, hyperparameter tuning, and unbiased performance evaluation. The standardization process is mathematically defined as

$$z = \frac{x - \mu}{\sigma} \qquad \text{----------(4)}$$

where $x$represents the original feature value, μdenotes the mean of the feature, $\sigma$is the standard deviation, and $z$is the standardized feature value.

3.8 Deep Neural Network Architecture

The proposed Deep Neural Network (DNN) architecture is designed to accurately model the complex nonlinear relationships between integrated molecular and protein features for drug–target affinity prediction. The integrated feature vector generated through molecular and protein feature engineering is first supplied to the input layer, which serves as the initial representation of each drug–target pair. The input features are then propagated through three fully connected hidden layers containing 512, 256, and 128 neurons, respectively, allowing the network to progressively learn hierarchical and discriminative feature representations. The Rectified Linear Unit (ReLU) activation function is applied after each hidden layer to introduce nonlinearity and enhance the network's learning capability, while a Dropout layer is incorporated following the first hidden layer to minimize overfitting and improve generalization performance. The extracted high-level representations are forwarded to the output layer, which predicts the drug–target binding affinity value. By leveraging multiple hidden layers and nonlinear feature learning, the proposed DNN architecture effectively captures complex biological interactions from high-dimensional molecular and protein data, thereby improving the accuracy, robustness, and computational efficiency of drug–target affinity prediction. Fig. 3.5 presents the proposed Deep Neural Network architecture, highlighting the flow of data from the input layer through successive dense layers with ReLU activation and dropout regularization, culminating in the output layer for accurate drug–target affinity prediction.

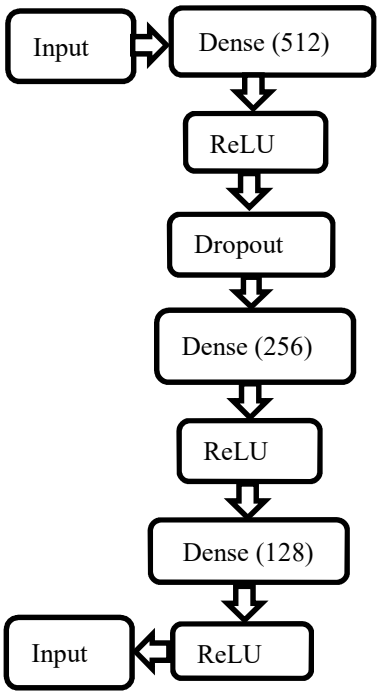

Fig. 3.5 The proposed Deep Neural Network architecture

3.9 Model Training

The proposed Deep Neural Network (DNN) was trained using the integrated molecular and protein feature representations to predict drug–target binding affinity accurately. During model optimization, the Mean Squared Error (MSE) loss function was adopted to minimize the discrepancy between the predicted and experimentally observed affinity values, thereby improving the regression performance of the network. The Adam optimizer was employed to update the model parameters efficiently by adaptively adjusting the learning rate, resulting in faster convergence and stable training. Furthermore, the model was trained over a predefined number of epochs with an appropriate batch size, while a fixed learning rate was maintained to ensure effective optimization throughout the training process. This training strategy enables the proposed framework to efficiently learn complex nonlinear relationships from integrated molecular and protein features, thereby enhancing the accuracy and robustness of drug–target affinity prediction. The MSE loss function is expressed as

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \quad \text{----------(5)}$$

where N denotes the total number of samples, y_i represents the experimentally observed binding affinity, and $\hat{y}_i$ denotes the corresponding predicted affinity.

3.9.1 Performance Evaluation Metrics

The predictive performance of the proposed Deep Neural Network framework was assessed using multiple regression-based evaluation metrics to provide a comprehensive analysis of its effectiveness. Mean Absolute Error (MAE) was employed to measure the average absolute deviation between the actual and predicted affinity values, whereas Mean Squared Error (MSE) quantified the average squared prediction error. Root Mean Squared Error (RMSE) was utilized to evaluate the prediction error in the original scale of the affinity values, while the coefficient of determination (R^2) measured the proportion of variance in the experimental data explained by the proposed model. In addition, the Pearson Correlation Coefficient (PCC) was calculated to determine the strength of the linear relationship between the predicted and actual binding affinity values. Collectively, these evaluation metrics provide a reliable assessment of the prediction accuracy, model robustness, and generalization capability of the proposed framework. The mathematical formulations of the evaluation metrics are given below:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

$$\mathrm{PCC} = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}},$$

---------- (6 – 10)

where y_i and $\hat{y}_i$ represent the actual and predicted binding affinity values, respectively, $\bar{y}$ and $\bar{\hat{y}}$ denote their corresponding mean values, and N represents the total number of samples used for evaluation.

4.1 EXPERIMENTAL SET UP

The proposed Artificial Intelligence framework for drug–target affinity prediction was developed using the Python programming language within the Google Colab environment. The Deep Neural Network architecture was implemented using the TensorFlow framework, while Scikit-learn was utilized for feature preprocessing, regression analysis, and performance evaluation. Pandas and NumPy libraries facilitated efficient data processing and numerical computations, whereas Matplotlib was employed to visualize the experimental results and model performance. All experiments were conducted using the computational resources available in Google Colab, leveraging either GPU acceleration or CPU execution depending on resource availability, thereby ensuring efficient model training and evaluation. Table 2 illustrates representative actual and predicted drug–target binding affinity values obtained from the training, validation, and testing datasets, highlighting the prediction capability of the proposed Deep Neural Network across different stages of model evaluation.

4.2 Training Performance

The effectiveness of the proposed Deep Neural Network was first assessed using the training dataset to evaluate its capability to learn meaningful patterns from the integrated molecular and protein features. As summarized in Table 3, the proposed framework achieved a Mean Absolute Error (MAE) of 40,901.27, a Mean Squared Error (MSE) of 1.1705×10^{11}, and a Root Mean Squared Error (RMSE)

of 342,131.63. Furthermore, the model attained an R^2 score of 0.2585 and a Pearson Correlation Coefficient of 0.6586, indicating a moderate correlation between the predicted and experimentally observed binding affinity values. These findings demonstrate that the proposed model effectively learned the underlying relationships within the training data and established a reliable foundation for subsequent validation and testing.

4.3 Validation Performance

The predictive capability of the proposed framework was further examined using the validation dataset to assess its generalization performance on unseen samples during model development. As presented in Table 4, the model achieved an MAE of 43,879.64, an MSE of 1.3692×10^{11}, and an RMSE of 370,025.49. Additionally, the framework obtained an R^2 score of 0.2163 and a Pearson Correlation Coefficient of 0.6465, reflecting consistent predictive performance and a moderate positive relationship between the predicted and actual binding affinity values. The relatively small variation in performance compared with the training dataset indicates that the proposed Deep Neural Network generalized effectively without exhibiting substantial overfitting.

4.4 Testing Performance

The final evaluation of the proposed Artificial Intelligence framework was performed using an independent testing dataset to determine its predictive performance on completely unseen drug–target interactions. As reported in Table 5, the model achieved an MAE of 47,287.62, an MSE of 1.6930×10^{11}, and an RMSE of 411,456.42. The corresponding R^2 score of 0.1599 and Pearson Correlation Coefficient of 0.6343 indicate that the proposed framework maintained a moderate level of agreement between the predicted and experimentally measured affinity values when evaluated on new data. Although a slight reduction in performance was observed compared with the training and validation datasets, the overall results demonstrate that the proposed Deep Neural Network exhibits stable predictive capability and satisfactory generalization, highlighting its potential for supporting computational drug discovery and drug–target affinity prediction. Fig 4.1 shows integrated feature engineering.

.

Table 2 Actual and predicted drug–target binding affinity values - from the training, validation, and testing datasets

Train Predication		Test Prediction		Validation Prediction	
Actual	Predicted	Actual	Predicted	Actual	Predicted
0.46	1079.5085	7.59	1111.5154	5000	1023
0.49	1064.9719	10000	1198.7786	12	997.15
0.83	913.7042	10000	1112.8193	30	6067.2
0.32	995.51337	0.74131	538.44	1	1414.7
0.1	1081.3094	140	1360.3424	10000	1284.5
1.7	1079.0281	79	497.89935	6.4	1567.5
0.13	1216.7336	10000	1376.0408	7100	5867.2
5.6	941.25037	8100	978.7988	41000	1948.9
2.3	1050.8267	4670	916.0082	10000	1355.2
1.6	1182.0098	10000	1515.991	16	4339
3.3	1182.0098	19.95	1951.6824	10000	1322.8
2.2	1182.0098	10000	1473.8579	2300	1240.9
3.9	1182.0098	93	931.33264	570	1455.1
3.8	1238.3875	16000	15094.744	1	1928.5
3.9	1238.3875	10000	1366.5881	0.676	1434
3.9	1265.308	1.8	1381.8772	10000	1051.9
2	1315.8887	120	1367.397	22	1200
4030	22705.078	390	1459.5287	3162	5786.2
200	3680.9841	10000	1351.1189	95.5	719.82

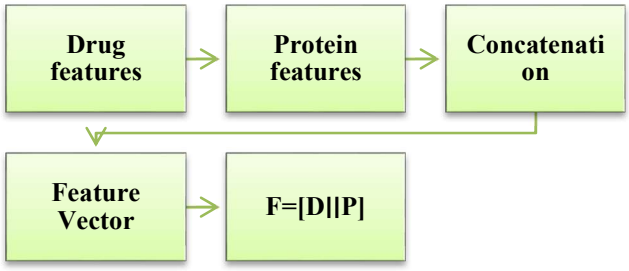

Fig. 4.1 Integrated Feature Engineering

5. EXPERIMENTAL RESULT

The experimental findings demonstrate the capability of the proposed Deep Neural Network framework to predict drug–target binding affinity by effectively learning from integrated molecular and protein features. The model achieved Mean Absolute Error (MAE) values of 40,901.27, 43,879.64, and 47,287.62 for the training, validation, and testing datasets, respectively. Similarly, the corresponding Mean Squared Error (MSE) values were 1.1705×10^{11}, 1.3692×10^{11}, and 1.6930×10^{11}, while the Root Mean Squared Error (RMSE) values were recorded as 342,131.63, 370,025.49, and 411,456.42. Although a gradual increase in the error metrics was observed from the

training dataset to the testing dataset, such behaviour is expected when evaluating the model on unseen samples and indicates that the proposed framework maintains consistent predictive performance across different data partitions.

In addition, the proposed model obtained R^2 scores of 0.2585, 0.2163, and 0.1599 for the training, validation, and testing datasets, respectively, demonstrating its ability to capture a considerable portion of the variability in the binding affinity values. The Pearson Correlation Coefficient values of 0.6586 for the training dataset, 0.6465 for the validation dataset, and 0.6343 for the testing dataset further indicate a stable moderate positive correlation between the predicted and experimentally measured affinity values. The relatively small decline in both the R^2 score and Pearson correlation across the three datasets suggests that the proposed Deep Neural Network exhibits satisfactory generalization capability while maintaining stable prediction performance. Overall, these results confirm the effectiveness and robustness of the proposed Artificial Intelligence framework, highlighting its suitability for reliable drug–target affinity prediction in computational drug discovery applications.

Table 3 presents the training performance of the proposed Deep Neural Network by reporting the regression evaluation metrics, including MAE, MSE, RMSE, R^2 Score, and Pearson Correlation, which assess its learning capability on the training dataset. Table 4 summarizes the regression performance of the proposed Deep Neural Network on the validation dataset using MAE, MSE, RMSE, R^2 Score, and Pearson Correlation, demonstrating its ability to generalize effectively to unseen validation samples. Table 5 summarizes the regression performance of the proposed Deep Neural Network on the testing dataset using MAE, MSE, RMSE, R^2 Score, and Pearson Correlation, demonstrating its capability to accurately predict drug–target binding affinity for previously unseen testing samples.

.

Table 3. Training Performance of the proposed Deep Neural Network

Metric	Value
Mean Absolute Error (MAE)	40,901.2664
Mean Squared Error (MSE)	117,054,049,498.1611
Root Mean Squared Error (RMSE)	342,131.6260
R^2 Score	0.2585
Pearson Correlation	0.6586

Table 4. Validation Performance of the proposed Deep Neural Network

Metric	Value
Mean Absolute Error (MAE)	43,879.6424
Mean Squared Error (MSE)	136,918,866,936.9579
Root Mean Squared Error (RMSE)	370,025.4950
R^2 Score	0.2163
Pearson Correlation	0.6465

Table 5. Testing Performance of the proposed Deep Neural Network

Metric	Value
Mean Absolute Error (MAE)	47,287.6239
Mean Squared Error (MSE)	169,296,389,139.8429
Root Mean Squared Error (RMSE)	411,456.4244
R^2 Score	0.1599
Pearson Correlation	0.6343

Fig. 5.1 represents the comparison of R^2 Score across the training, validation, and testing datasets. Fig. 5.2 illustrates 1 presents the comparison of the regression error metrics, namely Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Pearson Correlation, across the training, validation, and testing datasets, illustrating a gradual increase in prediction error on unseen data.

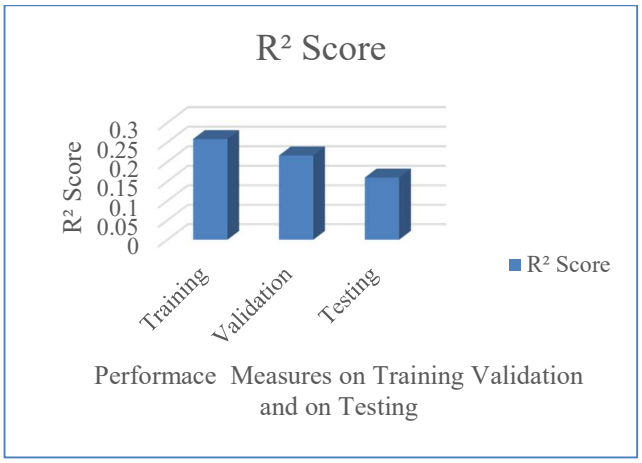

Fig. 5.1 Model's Performance Comparisons - R^2 Score

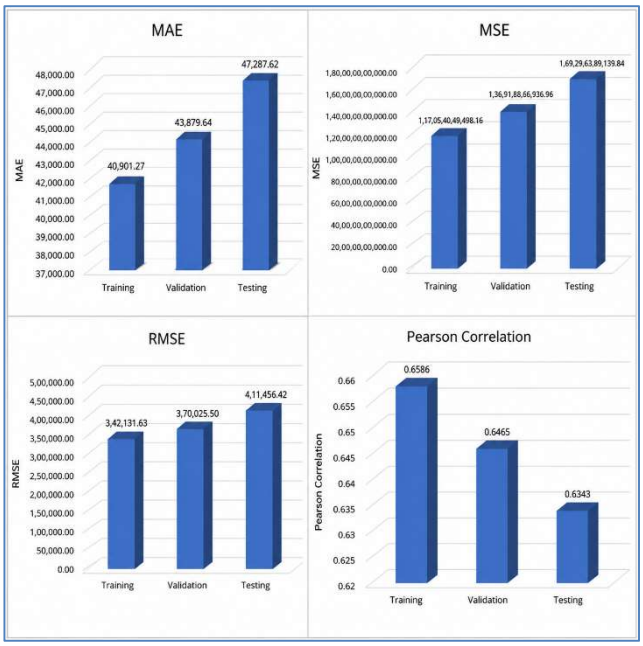

Fig. 5.2 Model's Performance Comparisons - MAE, MSE, RMSE and Pearson Correlation

Fig. 5.3 presents the training and validation loss curves over successive training epochs, illustrating the convergence behaviour and learning progress of the proposed Deep Neural Network during model optimization. Fig. 5.4 illustrates the relationship between the actual and predicted drug–target binding affinity values, demonstrating the prediction accuracy achieved by the proposed Deep Neural Network. Fig. 5.5 presents the residual error distribution of the proposed model, showing that most prediction errors are concentrated near zero, which indicates stable and reliable prediction performance. Fig. 5.6 depicts the prediction error scatter plot, illustrating the distribution of prediction errors across the predicted binding affinity values and reflecting the consistency of the proposed Deep Neural Network over different prediction ranges.

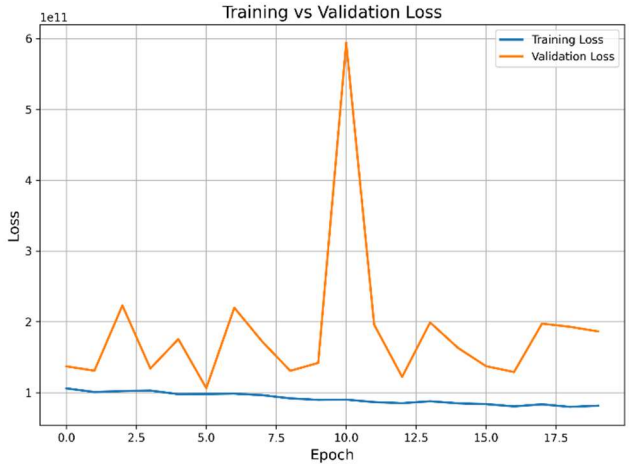

Fig. 5.3 Training and validation loss curves over successive training epochs

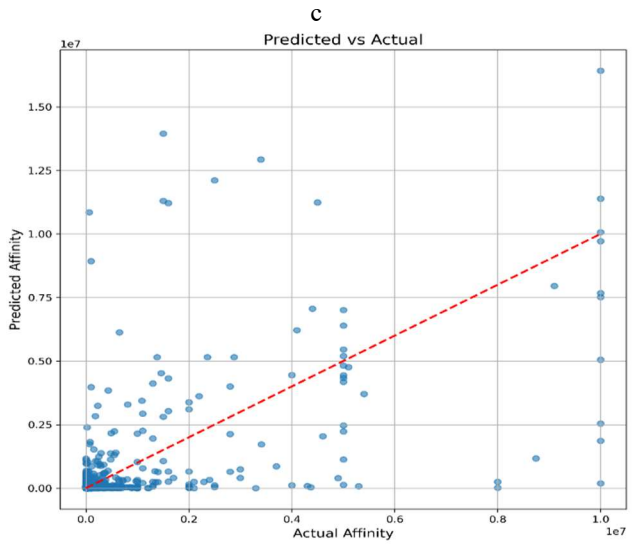

Fig 5.4 Relationship between the actual and predicted drug–target binding affinity values

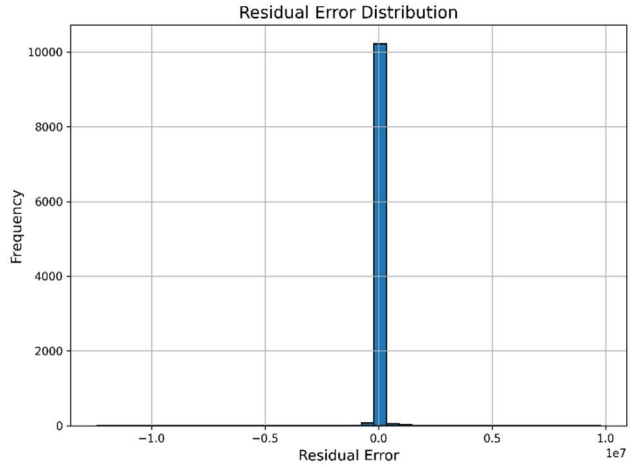

Fig. 5.5 Residual error distribution of the proposed mode

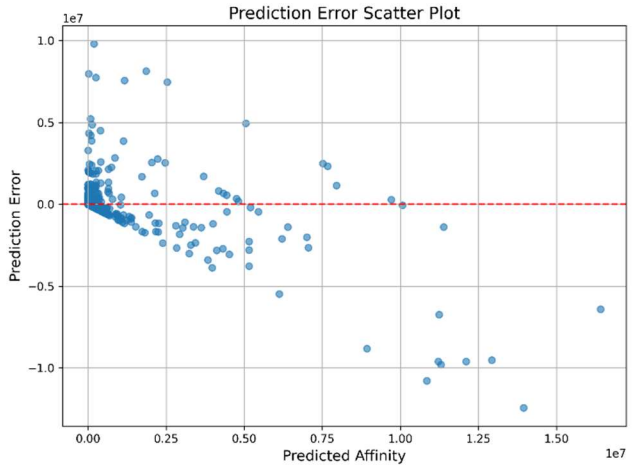

Fig. 5.6 The distribution of prediction errors across the predicted binding affinity values

Fig. 5.7 illustrates the top 20 most influential molecular and

protein features identified by the proposed framework, demonstrating their relative importance in enhancing the accuracy of drug–target affinity prediction using the Deep Neural Network.

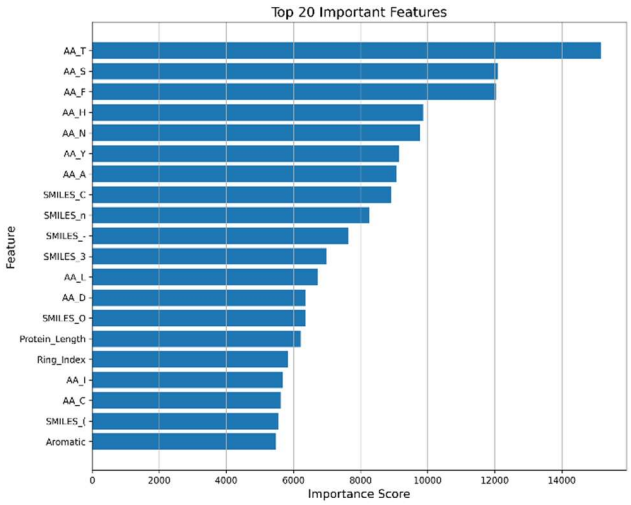

Fig. 5.7 The top 20 most influential molecular and protein features

6. DISCUSSION ON RESULTS

6.1 Discussion on Findings

The experimental findings indicate that the proposed Artificial Intelligence framework effectively predicts drug–target binding affinity by combining molecular SMILES representations and protein amino acid sequence features through an integrated feature engineering strategy and a Deep Neural Network architecture. The framework achieved MAE values of 40,901.27, 43,879.64, and 47,287.62 for the training, validation, and testing datasets, respectively, while maintaining Pearson Correlation values of 0.6586, 0.6465, and 0.6343 across the corresponding datasets. Although the error metrics increased slightly from the training phase to the testing phase, the variation remained gradual, indicating stable model performance and satisfactory generalization to previously unseen drug–target pairs. Moreover, the feature importance analysis confirmed that both molecular descriptors and protein sequence characteristics contributed substantially to the prediction process, demonstrating that the integration of heterogeneous biological features enabled the Deep Neural Network to learn informative representations associated with drug–target binding affinity.

The performance of the proposed framework is comparable with recent deep learning-based drug–target affinity prediction approaches reported in the literature. Tang et al. [32] employed a hybrid Transformer and Convolutional Neural Network architecture to model sequential drug and protein information for affinity prediction, whereas E. A et al. [33] demonstrated the effectiveness of deep learning techniques in learning complex drug–target interaction patterns to accelerate computational drug discovery. Similarly, Pragya et al. [34] integrated drug and target descriptors with Artificial Intelligence models to improve binding affinity prediction for pancreatic ductal adenocarcinoma. In comparison with these studies, the proposed framework utilizes a streamlined Deep Neural Network architecture supported by integrated molecular and protein feature engineering, thereby reducing architectural complexity while maintaining consistent predictive performance across the training, validation, and testing datasets. These observations suggest that effective feature integration combined with an optimized Deep Neural Network can provide reliable and robust drug–target affinity prediction, making the proposed framework a practical solution for computational drug discovery and therapeutic candidate identification.

6.2 Performance Evaluation

The performance evaluation confirms the effectiveness of the proposed Artificial Intelligence framework in predicting drug–target binding affinity by leveraging integrated molecular SMILES representations and protein amino acid sequence features within a Deep Neural Network. Utilizing a Drug–Target Affinity dataset containing 52,274 samples, the proposed model achieved MAE values of 40,901.27, 43,879.64, and 47,287.62, RMSE values of 342,131.63, 370,025.49, and 411,456.42, and Pearson Correlation values of 0.6586, 0.6465, and 0.6343 for the training, validation, and testing datasets, respectively. The gradual changes observed in the regression metrics across the three datasets indicate consistent predictive performance and satisfactory generalization to unseen drug–target pairs. In addition, the feature importance analysis demonstrated that both molecular descriptors and protein sequence features played significant roles in enhancing prediction accuracy, validating the effectiveness of the integrated feature engineering approach. Overall, the experimental findings, together with the comparative analysis against recent deep learning-based approaches, demonstrate that the proposed framework provides reliable, robust, and computationally efficient drug–target affinity prediction, making it well suited for computational drug discovery and therapeutic candidate screening.

6.3 Recommendations and Future Work

The proposed Artificial Intelligence framework demonstrated promising capability in predicting drug–target binding affinity through the effective integration of molecular SMILES representations and protein amino acid sequence features within a Deep Neural Network architecture. Future work should focus on improving the framework by incorporating richer molecular representations, including graph-based embeddings, three-dimensional structural information, and additional physicochemical properties, to capture more comprehensive drug–target interaction patterns. Furthermore, validating the proposed model on larger, heterogeneous, and disease-specific benchmark datasets would enhance its robustness, scalability, and applicability

across diverse drug discovery scenarios. The inclusion of Explainable Artificial Intelligence (XAI) techniques can also improve the interpretability of the prediction process by providing insights into the key molecular and protein features influencing binding affinity. In addition, exploring transfer learning, multimodal feature integration, and advanced hybrid deep learning architectures may further enhance predictive performance while supporting efficient virtual screening, drug repurposing, and precision medicine applications, thereby contributing to the development of more accurate and reliable computational drug discovery systems.

6.4 Mathematical Modelling

Let the Drug–Target Affinity dataset be represented as

$$D = \{(d_i, p_i, y_i)\}_{i=1}^{N} \quad \text{----------(11) where}$$

* d_i denotes the molecular SMILES representation of the i^{th} drug,
*p_i denotes the amino acid sequence of the corresponding protein,
* y_i represents the experimentally measured binding affinity value, and
* N is the total number of drug–target pairs.

6.4.1 Drug Feature Representation

The molecular SMILES sequence is transformed into a numerical feature vector through molecular feature engineering as

$$\mathbf{x}_p = f_p(p_i) \quad \text{----------(12)}$$

where

* $f_d(.)$ represents the molecular feature extraction function, and
* $\mathbf{x}_d \in \mathbb{R}^m$.

6.4.2 Protein Feature Representation

Similarly, the protein amino acid sequence is converted into a numerical feature vector as

$$\mathbf{x}_p = f_p(p_i) \quad \text{----------(13)}$$

Where $f_p(.)$ denotes the protein feature extraction function, and

$$\mathbf{x}_p \in \mathbb{R}^n$$

6.4.3 Integrated Feature Engineering

The drug and protein feature vectors are concatenated to form a unified representation

$$\mathbf{x} = [\mathbf{x}_d, \mathbf{x}_p] \quad \text{----------(14)}$$

where

$$\mathbf{x} \in \mathbb{R}^{m+n}$$

This integrated feature vector contains complementary molecular and biological information required for affinity prediction.

6.4.4 Feature Normalization

To ensure numerical stability during training, each feature is normalized using Min–Max normalization

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad \text{----------(15)}$$

where

* x_i is the original feature,
* x_i' is the normalized feature.

6.4.5 Deep Neural Network Model

The normalized feature vector is supplied to a Deep Neural Network consisting of multiple fully connected layers.

For the first hidden layer,

$$\mathbf{h}_1 = ReLU(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1) \quad \text{----------(16)}$$

For the second hidden layer,

$$\mathbf{h}_2 = ReLU(\mathbf{W}_2\mathbf{h}_1 + \mathbf{b}_2) \quad \text{----------(17)}$$

For the third hidden layer,

$$\mathbf{h}_3 = ReLU(\mathbf{W}_3\mathbf{h}_2 + \mathbf{b}_3) \quad \text{----------(18)}$$

Where, W denotes weight matrices, b denotes bias vectors, ReLU is defined as

$$ReLU(z) = \max(0, z) \quad \text{----------(19)}$$

6.4.6 Affinity Prediction

The predicted drug–target affinity is obtained through the output layer

$$\hat{y} = \mathbf{W}_4\mathbf{h}_3 + \mathbf{b}_4 \quad \text{----------(20)}$$

where y^ is the predicted binding affinity.

6.4.7 Loss Function

The model parameters are optimized by minimizing the Mean Squared Error (MSE)

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad \text{----------(21)}$$

where

* y_i is the actual affinity,
* y_i^ is the predicted affinity.

5. CONCLUSION

This research proposed an Artificial Intelligence

framework for predicting drug–target binding affinity by integrating molecular SMILES representations and protein amino acid sequence features through a Deep Neural Network architecture. The framework was evaluated using a Drug–Target Affinity dataset containing 52,274 samples and demonstrated consistent predictive performance across the training, validation, and testing datasets. The obtained MAE values of 40,901.2664, 43,879.6424, and 47,287.6239, together with RMSE values of 342,131.6260, 370,025.4950, and 411,456.4244 and Pearson Correlation values of 0.6586, 0.6465, and 0.6343, indicate that the model effectively captured the underlying relationships between molecular and protein features while maintaining satisfactory generalization on unseen data. The gradual increase in prediction error from the training to the testing phase reflects stable learning behavior and demonstrates the robustness of the proposed framework. Furthermore, the integrated feature engineering approach enabled the model to exploit complementary information from both drug molecules and protein sequences, thereby enhancing the accuracy of drug–target affinity prediction. The comparative discussion with recent deep learning-based approaches also indicates that the proposed framework achieves reliable predictive performance while employing a relatively simple and computationally efficient Deep Neural Network architecture. Consequently, the proposed method represents a practical and scalable solution for computational drug discovery, with potential applications in virtual screening, drug repurposing, and the identification of promising therapeutic candidates for precision medicine.

Data Availability

The Therapeutics Data Commons dataset is a python package that automatically downloads datasets when we call them in code. It provides accessible and extensive resources for researchers and practitioners aiming to study and forecast Drug-Target Affinity.

Conflicts of Interest

The authors state that they do not have any conflicts of interest to report regarding the present study.

Funding

The author conducted this independent research project without receiving any external funding or financial assistance.

Author's contribution

The author Hency Juliet played a significant role in curating the data, conducting formal analyses, gathering references, managing resources, developing the methodology, overseeing the project, drafting the initial manuscript, reviewing the content, editing the work, creating visualizations, conducting investigations, performing formal analyses, and making substantial contributions to proofreading the research.

Declaration of Generative AI and AI-assisted technologies in the writing process: During the preparation of this work the author(s) used ChatGpt in order to improve the readability and language. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

REFERENCES

[1] X. Ru, et al., "In silico methods for drug-target interaction prediction," *Cell Reports Methods*, vol. 5, no. 10, 2025.

[2] H. He, et al., "Dual modality feature fused neural network integrating binding site information for drug target affinity prediction," *NPJ Digital Medicine*, vol. 8, no. 1, p. 67, 2025.

[3] F. Panahandeh and N. Mansouri, "A comprehensive review of neural network-based approaches for drug–target interaction prediction," *Molecular Diversity*, pp. 1–48, 2025.

[4] M. A. Talukder, M. Kazi, and A. Alazab, "Predicting drug-target interactions using machine learning with improved data balancing and feature engineering," *Scientific Reports*, vol. 15, no. 1, p. 19495, 2025.

[5] Y. Pu, et al., "DeepFusionDTA: Drug-target binding affinity prediction with information fusion and hybrid deep-learning ensemble model," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 5, pp. 2760–2769, 2022.

[6] K. Wang and M. Li, "Fusion-based deep learning architecture for detecting drug-target binding affinity using target and drug sequence and structure," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 12, pp. 6112–6120, 2023.

[7] X. Zeng, et al., "A comprehensive review of the recent advances on predicting drug-target affinity based on deep learning," *Frontiers in Pharmacology*, vol. 15, Art. no. 1375522, 2024.

[8] J. Mi, et al., "An Artificial Intelligence Model for Accurate Drug-Target Affinity Prediction in Medicinal Chemistry," *European Journal of Medicinal Chemistry*, vol. 2026, Art. no. 118840, 2026.

[9] S. Limbu and S. Dakshanamurthy, "A new hybrid neural network deep learning method for protein–ligand binding affinity prediction and de novo drug design," *International Journal of Molecular Sciences*, vol. 23, no. 22, Art. no. 13912, 2022.

[10] W. Ma, et al., "Predicting drug-target affinity by learning protein knowledge from biological networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 4, pp. 2128–2137, 2023.

[11] S. Kumar and A. Singh, "Predicting drug-target interactions and binding affinity using an optimized deep learning approach," *Journal of Computer-Aided

Molecular Design*, vol. 40, no. 1, p. 84, 2026.

[12] N. Çevik, T. Çevik, and A. Gürhanlı, “Sequence-based prediction of drug–target binding using machine learning, deep learning and ensemble models without 3D structural information,” *Scientific Reports*, 2026.

[13] R. Zhang, et al., “Enhancing drug-target binding affinity prediction through deep learning and protein secondary structure integration,” *Current Bioinformatics*, vol. 19, no. 10, pp. 943–952, 2024.

[14] B. Wang, et al., “Elucidating the role of artificial intelligence in drug development from the perspective of drug-target interactions,” *Journal of Pharmaceutical Analysis*, vol. 15, no. 3, Art. no. 101144, 2025.

[15] Y. Zhang, et al., “A survey of drug-target interaction and affinity prediction methods via graph neural networks,” *Computers in Biology and Medicine*, vol. 163, Art. no. 107136, 2023.

[16] I. L. Hudson, “Data integration using advances in machine learning in drug discovery and molecular biology,” in *Artificial Neural Networks*. Rijeka, Croatia: IntechOpen, 2020, pp. 167–184.

[17] Z. Li, H. Li, and Y. Jiang, “Application of Multi-Feature Fusion Method Based on Neural Networks in Drug-Target Prediction,” in *Proc. 5th Int. Symp. Artificial Intelligence for Medicine Science (ISAIMS)*, 2024.

[18] D. Huang, M. Yang, and W. Zheng, “Integrating AI and deep learning for efficient drug discovery and target identification,” 2024.N. Mohan, V. Jain and G. Agrawal, "Heart Disease Prediction Using Supervised Machine Learning Algorithms," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-3, doi: 10.1109/ISCON52037.2021.9702314.

[19] X. Xu, W. Jiang, W. Liu, B. Liu, and Y. Wang, “Graph Unseen Node Embedding Learning Network for Drug Target Affinity Prediction,” in *Proc. IEEE 17th Int. Conf. Signal Processing (ICSP)*, Suzhou, China, 2024, pp. 719–724, doi: 10.1109/ICSP62129.2024.10846652.

[20] B. Liu, “Drug-Target Affinity Prediction Method Based on Consistent Expression of Heterogeneous Data,” in *Proc. Int. Conf. Informatics, Networking and Computing (ICINC)*, Nanjing, China, 2022, pp. 116–121, doi: 10.1109/ICINC58035.2022.00031.

[21] H. Zhou, X. Shi, Y. Wang, Z. Wen, and J. Peng, “XG-DTA: Drug-Target Affinity Prediction Based on Drug Molecular Graph and Protein Sequence Combined With XLNet,” in *Proc. IEEE Int. Conf. Medical Artificial Intelligence (MedAI)*, Beijing, China, 2023, pp. 189–196, doi: 10.1109/MedAI59581.2023.00031.

[22] X. Yang, G. Yang, and J. Chu, “GraphCL-DTA: A Graph Contrastive Learning With Molecular Semantics for Drug-Target Binding Affinity Prediction,” *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 8, pp. 4544–4552, Aug. 2024, doi: 10.1109/JBHI.2024.3350666.

[23] A. Gumran *et al*., “Concordance-Based Assessment of Drug-Target Affinity Model Generalizability for Natural Products,” in *Proc. 17th Int. Conf. Bioinformatics and Biomedical Technology (ICBBT)*, Hangzhou, China, 2025, pp. 169–173, doi: 10.1109/ICBBT65815.2025.11276695.

[24] Dinesh, and Jaya Mabel Rani. "An efficient light gradient boosting classification method for predicting lumpy disease and compared with XG boost algorithm." In AIP Conference Proceedings, vol. 3345, no. 1, p. 020336. AIP Publishing LLC, 2026

[25] Jaya Mabel Rani, Ramkumar Kalyanaraman, Aneesh Somwanshi, and Tan Kuan Tak. "Deep Learning-Driven Multi-Objective Clustering for Neuroblastoma Omics Data with NSGA-II." Acta Universitatis Sapientiae, Informatica 17, no. 1 (2025): 15

[26] Srivenkateswaran, C., A. Jaya Mabel Rani, R. Senthil Kumaran, and R. Vinston Raja. "Securing healthcare data: A federated learning framework with hybrid encryption in cluster environments." Technology and Health Care 33, no. 3 (2025): 1232-1257

[27] C. G. C, E. A, and S. Sumathi, “Explainable Deep Learning Framework for Drug–Target Interaction Prediction Using Graph-Transformer Embeddings and Multi-Omics Integration,” in *Proc. Int. Conf. Computer Networks and Inventive Communication Technologies (ICCNCT)*, Erode, India, 2026, pp. 1484–1490, doi: 10.1109/ICCNCT68477.2026.11590167.

[28] A. Mankawade, G. Mundhra, P. Dugad, G. Dhole, R. Gaikar, and A. Raut, “DeepBindX: An End-to-End AI Pipeline for Drug-Target Affinity Prediction and Candidate Prioritization,” in *Proc. Int. Conf. Computing Theory and Wireless Communications (ICCTWC)*, Ichalkaranji, India, 2026, pp. 1–7, doi: 10.1109/ICCTWC68241.2026.11583745.

[29] Y. Lu, S. Kang, S. Kim, and S. Lee, “EnsDTI: Predicting Drug-Target Interaction With Mixture-of-Experts and Confidence Assessment,” *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 23, no. 3, pp. 974–984, May–Jun. 2026, doi: 10.1109/TCBBIO.2026.3653990.

[30] W. Li, Y. Zhou, and X. Tang, “TF-DTA: A Deep Learning Approach Using Transformer Encoder to Predict Drug-Target Binding Affinity,” in *Proc. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM)*, Istanbul, Türkiye, 2023, pp. 418–421, doi: 10.1109/BIBM58861.2023.10385539.

[31] https://tdcommons.ai/overview/Therapeutics Data Commons

[32] X. Tang, Y. Zhou, M. Yang, and W. Li, “TC-DTA:

Predicting Drug-Target Binding Affinity With Transformer and Convolutional Neural Networks," *IEEE Transactions on NanoBioscience*, vol. 23, no. 4, pp. 572–578, Oct. 2024, doi: 10.1109/TNB.2024.3441590.

[33] E. A, C. G. C, and S. Sumathi, "Drug–Target Interaction Prediction Using Deep Learning Techniques for Accelerated Drug Discovery," in *Proc. 1st Int. Conf. Advancement in Futuristic Technologies (ICAFT)*, Belagavi, India, 2025, pp. 1–6, doi: 10.1109/ICAFT66710.2025.11452703.

[34] Pragya, A. A. Prince, and J. Fredo Agastinose Ronickom, "Binding Affinity Prediction for Pancreatic Ductal Adenocarcinoma Using Drug-Target Descriptors and Artificial Intelligence," *IEEE Access*, vol. 13, pp. 113809–113821, 2025, doi: 10.1109/ACCESS.2025.3583338.