

FusionNet- A Hybrid Deep Learning Approach for Accurate Drug-Target Binding Prediction

Original Scientific Paper

Tintu Vijayan*

Department of CSE, Presidency University,
Bangalore, India
tintuidhaya@gmail.com

Pamela Vinitha Eric

Department of CSE, Presidency University,
Bangalore, India
pamelavinitha.eric@presidencyuniversity.in

*Corresponding author

Abstract – Identifying drug-target binding affinities (DTBA) is crucial in drug discovery, to understand how effectively drugs interact with their targets. However, traditional methods often struggle to accurately capture the complex relationships in biological data, leading to limitations in their predictive power. This paper introduces FusionNet, an advanced deep-learning model designed to improve DTBA prediction. FusionNet combines the strengths of Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and Transformers, to better understand both short-range and long-range interactions in biological sequences and employs the Layer-wise Adaptive Moments (LAMB) optimizer, which ensures the model is more efficient and stable, especially when working with large datasets. FusionNet achieved an MSE of 0.20 and an rm^2 of 0.681 on the Davis dataset and an MSE of 0.18 and an rm^2 of 0.71 on the KIBA dataset, significantly outperforming existing models like SimBoost, GANsDTA, DeepCDA, and DeepDTA, making it a powerful tool for drug discovery and bioinformatics. This work not only enhances the accuracy of DTBA prediction but also sets new performance standards by integrating advanced neural network architectures and optimizing their training process. FusionNet effectively addresses the limitations of previous approaches, offering a more reliable and efficient method for predicting drug-target interactions.

Keywords: Drug Target Binding Affinity (DTBA), Convolutional neural network (CNN), Layer-wise Adaptive Moments (LAMB)

Received: August 22, 2024; Received in revised form: November 14, 2024; Accepted: November 15, 2024

1. INTRODUCTION

Drug discovery and development is a complex, lengthy, and expensive process, often requiring over a decade and billions of dollars to bring a new drug to market. A critical aspect of this process is understanding the interaction between drug molecules and their target proteins, quantified by the drug-target binding affinity (DTBA). Accurate prediction of DTBA is essential for determining the efficacy and safety of potential therapeutics. Traditional experimental methods such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, surface plasmon resonance (SPR), and isothermal titration calorimetry (ITC) are critical for understanding drug-target interactions at the molecular level [1]. However, these methods are labor-intensive, time-consuming, and expensive, often limiting their scalability and practical application in early-stage drug discovery. For example, X-ray crystallography,

while highly precise, can take several months to years for data collection and structural determination, making it impractical for high-throughput screening [2]. Similarly, NMR spectroscopy requires a high concentration of samples and extensive computational resources, limiting its efficiency [3]. Consequently, there is a significant demand for computational methods that can predict DTBA accurately, efficiently, and at scale.

Recent advances in computational power and the availability of large-scale biochemical datasets have catalyzed the development of various computational approaches for DTBA prediction. These methods range from classical machine learning techniques to more sophisticated deep learning models. Traditional machine learning models, such as random forests and support vector machines (SVMs), have been widely used in drug-target binding affinity prediction due to their simplicity and interpretability [4]. However, these models typically rely on manually engineered features,

such as molecular fingerprints and physicochemical descriptors, which fail to capture the complex, non-linear relationships inherent in biological data. This reliance on handcrafted features restricts the models' generalization ability to unseen data, reducing prediction accuracy and reliability [5]. Additionally, SVMs and other classical models are computationally inefficient when dealing with high-dimensional, large-scale datasets, limiting their effectiveness in practical, large-scale drug discovery applications [6]. In contrast, deep learning models can learn feature representations directly from raw data, showing promise in overcoming these limitations. Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) have been particularly effective due to their ability to model spatial and sequential data, respectively. [7]

Although CNNs and LSTMs have achieved great success in DIBA prediction, there are still some significant challenges. CNNs excel at capturing local spatial patterns but struggle to model long-distance dependencies, which are critical for understanding interactions between distant residues or a drug and its binding site on a protein. LSTMs, targeted at sequential data processing, can partially capture sequential dependencies. However, they may run into difficulties such as vanishing gradients and computational inefficiency when dealing with lengthy sequences, resulting in their model for the full complexity of drug-target interactions. [8]

Moreover, both CNNs and LSTMs fall short of effectively leveraging contextual information, which is essential for accurately modeling biological sequences. For instance, the interaction between a drug and a protein is influenced by the broader context of the protein's structure and the physicochemical properties of the drug, which these models struggle to capture comprehensively. This limitation hinders their ability to provide accurate and reliable DTBA predictions. Another limitation lies in the optimization of deep learning models. Standard optimizers, such as stochastic gradient descent (SGD) and Adam, may not be efficient enough to train large models on complex datasets, leading to suboptimal performance and longer training times, particularly when processing large bioinformatics datasets [9].

To address these challenges, we propose a hybrid deep learning model that integrates CNNs, LSTMs, and Transformers to improve DTBA prediction. The model leverages CNNs for local feature extraction, LSTMs for capturing sequential dependencies, and Transformers for modeling long-range dependencies and contextual information via self-attention [10]. This combination provides a comprehensive representation of sequences, enhancing accuracy and reliability. CNNs extract key patterns, LSTMs maintain temporal order, and Transformers capture broader interactions between drug and protein sequences.

To further enhance the training efficiency and performance of the hybrid model, the Layer-wise Adap-

tive Moments (LAMB) optimizer is employed. LAMB is specifically designed for large-batch training, making it well-suited for deep learning models on large datasets. It adapts the learning rate for each layer individually, ensuring stable and efficient convergence, and addressing the shortcomings of traditional optimizers in training large models efficiently, which is particularly beneficial in the context of our complex hybrid model.

The contributions are:

- A novel hybrid model combining CNNs, LSTMs, and Transformers to improve DTBA prediction by addressing key limitations of existing methods
- Implementation of the LAMB optimizer to enhance training efficiency and performance.
- Extensive experiments demonstrating the superiority of the model over traditional approaches.

2. RELATED WORKS

Deep learning has become a powerful tool for predicting drug-target binding affinity (DTBA). Öztürk et al. introduced DeepDTA, a model that uses CNNs to analyze sequence information from both targets and drugs, achieving high accuracy in predicting binding affinities [11]. Feng et al. developed PADME, which consistently outperformed baseline methods across multiple datasets using a deep learning-based framework [12]. Furthering this, Öztürk et al. introduced WideDTA, combining chemical and biological textual sequence information to enhance binding affinity predictions [13]. Zeng et al. improved upon these approaches by integrating multiple attention blocks, effectively encoding correlations between atoms and modeling drug-target interactions [14]. In other applications, Elansary et al. developed a bat-inspired optimizer using RNNs for predicting anti-viral cure drugs, highlighting the versatility of deep learning in drug discovery [15]. Makowski et al. explored machine learning models for co-optimizing therapeutic antibody affinity and specificity, emphasizing their importance in therapeutic development [16].

Hybrid models that integrate CNNs and LSTMs have shown significant improvements in drug discovery and drug repurposing. These models leverage CNNs' ability to capture local features and LSTMs' strengths in modeling sequential dependencies. For example, Yoon et al. outperformed traditional models in predicting DNA-protein binding sites using this hybrid approach [17]. In DTBA, hybrid CNN-LSTM models have demonstrated remarkable success by extracting key features from SMILES representations and protein sequences before processing them with LSTMs for accurate affinity predictions. DeepBind, which also integrates CNNs and LSTMs, has been used to predict protein-DNA binding affinity with high accuracy.

Transformers have revolutionized sequence modeling by effectively capturing long-range dependencies

with self-attention mechanisms. In bioinformatics, transformers have been applied to tasks like protein structure prediction, achieving state-of-the-art results. The LAMB optimizer, developed by You et al., enhances training efficiency and performance in large-batch scenarios by adapting learning rates layer-wise, which is especially beneficial for models with extensive parameter spaces [18]. LAMB builds on Adam [19], offering better generalization and faster convergence. The proposed hybrid model, combining CNNs, LSTMs, and Transformers with the LAMB optimizer, enhances DTBA prediction by leveraging their strengths for improved accuracy and generalization.

3. METHODOLOGY

3.1. DATASETS

The study utilizes two well-established datasets for predicting drug-target binding affinities: the KIBA dataset and the Davis dataset [20]. These datasets provide

comprehensive and experimentally validated information on drug-target interactions, making them suitable benchmarks for evaluating the performance of the proposed hybrid model. These datasets provide comprehensive and experimentally validated information on drug-target interactions, making them suitable benchmarks for evaluating the performance of the proposed hybrid model.

Table 1. The statistics of datasets

Dataset	No of drugs	No of proteins	Known DTI
Davis	68	442	30,056
Kiba	2,116	229	118,254

The KIBA dataset integrates information from multiple sources, including Kinase Inhibitor Bioactivity data, to provide a unified measure of drug-target binding affinities. It combines data from Ki, Kd, and IC50 measurements, offering a robust and comprehensive resource for kinase inhibitor bioactivity.

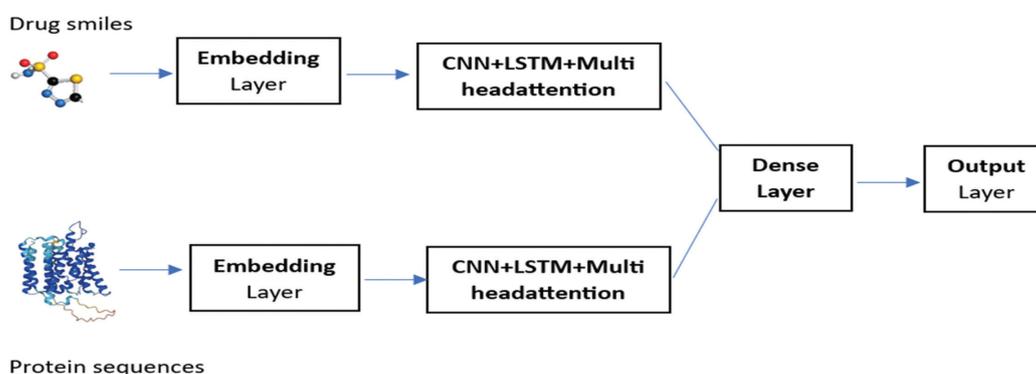


Fig.1. Proposed Model Structure

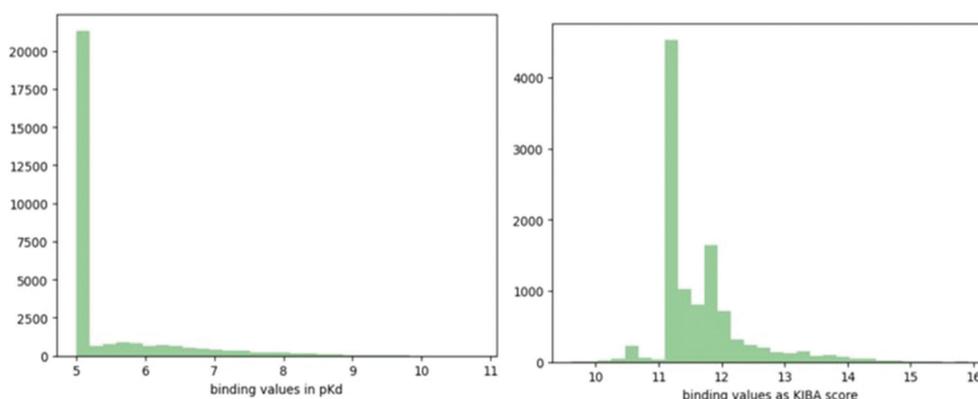


Fig. 2. Distribution of binding affinity values in Davis and Kiba dataset

3.2. PROPOSED METHODOLOGY

Predicting drug-target binding affinity (DTBA) is crucial in drug discovery, where the goal is to quantify the interaction strength between a drug molecule and a target protein. The problem can be mathematically stated as, given a drug represented by its SMILES (Simplified Molecular Input Line Entry System) string, S , and a tar-

get protein represented by its amino acid sequence, P , the task is to predict the binding affinity y , which can be mathematically represented as

$$y = f(S, P) \quad (1)$$

where f is the predictive model.

The ultimate objective is to optimize the function, $f(S, P)$ such that the predicted binding affinity y' close-

ly approximates the true binding affinity y minimizing the mean squared error (MSE) across the dataset.

In the proposed model as in Fig.1, Convolutional Neural Networks (CNNs) to extract local features from SMILES strings and protein sequences, utilizing multiple 1D convolutional layers with filter sizes of 4, 6, and 8, chosen based on empirical testing to capture various sequence patterns. Bidirectional Long Short-Term Memory (LSTM) networks handle sequential dependencies, with 64 units selected to balance complexity and efficiency. To model long-range dependencies, Transformer layers with multi-head self-attention (4 heads and key dimension of 64) are incorporated, chosen through preliminary experiments to optimize performance while managing computational demands. Hyperparameters were tuned using cross-validation and grid search, including filter sizes, LSTM dimensions, and dropout rates (set at 0.5 to mitigate overfitting). The LAMB optimizer was selected for its ability to adjust learning rates for each layer adaptively, improving training efficiency, particularly with large batch sizes (128) and a learning rate chosen to ensure stable and rapid convergence. The model was trained for 100 epochs, with Mean Squared Error (MSE) as the loss function and evaluation metric to measure the alignment between predicted and actual binding affinities.

Mathematically, the convolution operation for each filter size i can be expressed as

$$\text{conv}(x) = \text{ReLU}(\text{Conv1D}(x, \text{filters}_i, \text{Kernel_size}_i)) \quad (2)$$

where x represents the input sequence, and filters_i , Kernel_size_i correspond to the filter and kernel size for the i -th convolutional layer. Following convolution, max-pooling layers are applied to reduce the dimensionality and retain the most significant features, expressed as

$$\text{MaxPool}_i = \text{Maxpooling1D}(\text{Conv}_i(x)) \quad (3)$$

The pooled features are then concatenated to form a comprehensive feature representation. Subsequently, Long Short-Term Memory (LSTM) networks are employed to capture sequential dependencies in the data. The bidirectional LSTM processes the CNN-extracted features f by considering both forward and backward contexts, which can be mathematically represented as

$$\text{LSTMbi}(f) = \text{Concat}(\text{LSTMforward}(f), \text{LSTMbackward}(f)) \quad (4)$$

Where $\text{LSTMforward}(f)$, $\text{LSTMbackward}(f)$ are the LSTM operations in the forward and backward directions, respectively. Finally, Transformer layers are incorporated to model long-range dependencies more effectively. The Transformer utilizes a multi-head self-attention mechanism, which allows the model to assign different weights to various parts of the sequences during the interaction between SMILES and protein features.

The multi-head self-attention mechanism applies multiple attention layers in parallel (in our model, 4 heads), each learning distinct patterns from the input

sequences. The attention weights are calculated by taking the dot product of queries (Q), keys (K), and values (V), where Q , K , and V are derived from the input. This can be expressed as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

where d_k is the dimension of the keys (set to 64 in our model). This mechanism allows the model to focus on different parts of the sequences, making it well-suited for tasks like DTBA, where interactions between distant parts of the sequences are important.

Following the attention mechanism, the outputs are passed through a feed-forward neural network, and layer normalization is applied.

The multi-head self-attention applied to the LSTM-processed features H can be expressed as

$$H_{\text{attn}} = \text{Multi_Head_Self_Attention}(H) \quad (6)$$

This is followed by a feed-forward network to process the attended features.

$$H_{\text{ffn}} = \text{Feed_Forward_Network}(H_{\text{attn}}) \quad (7)$$

The processed features are then combined and normalized to produce the final feature representation for the prediction task.

To optimize the model, we employed the LAMB optimizer (Layer-wise Adaptive Moments based on Batch size), which is particularly effective in handling large batch sizes. Unlike traditional optimizers, LAMB adjusts the learning rate for each layer individually, taking into account both the gradient magnitude and the layer's weight norm. This enables the model to maintain stable training dynamics, especially in deep networks, while benefiting from faster convergence. The LAMB optimizer's advantages over traditional optimizers like Adam include improved scalability with large batch sizes (128 in our model) and more efficient training in deep architectures. The learning rate was carefully selected to ensure stable and rapid convergence, while a dropout rate of 0.5 was used to mitigate overfitting.

The model was trained for 100 epochs, with Mean Squared Error (MSE) as the loss function and evaluation metric to measure the alignment between predicted and actual binding affinities. The entire workflow is depicted in Algorithm 1 and Fig 3.

Algorithm 1

1. Input: SMILES strings, S Protein sequences P , Labels Y

2. Output: Predicted interaction scores Y'

3. Data Preparation

3.1. Shuffle the dataset, D containing triplets S, P, Y

3.2. Split the data into training and test sets:

$$\begin{aligned} \text{train_smiles}, \text{test_smiles} &\leftarrow S \\ \text{train_proteins}, \text{test_proteins} &\leftarrow P \end{aligned}$$

4. Tokenization and Padding

4.1. Initialize a character-level tokenizer, $\text{Tokenizer}_{SMILES}$

4.2. Fit the tokenizer on the training SMILES strings $\{S_i\}_{i \in D}$

4.3. Convert the SMILES strings to sequences of tokens, $\text{Seq}_{SMILES} = \text{Tokenizer}_{SMILES}(S)$

4.4. Pad the sequences to a maximum length, L_{SMILES}

$$\text{Padded}_{SMILES} = \text{Pad}(\text{Seq}_{SMILES}, L_{SMILES})$$

4.5. similarly padding for proteins

$$\text{Padded}_{Protein} = \text{Pad}(\text{Seq}_{Protein}, L_{Protein})$$

5. Define two input layers S_{input} for SMILES and P_{input} for protein sequences.

6. Encode SMILES and proteins using convolutional layers followed by LSTM layers.

$$E_{SMILES} = \text{Embedding}(S_{input})$$

$$C_{SMILES} = \text{Conv}_{1D}(E_{SMILES})$$

$$H_{SMILES}^{LSTM} = \text{LSTM}(C_{SMILES})$$

7. Apply transformer blocks with multi-head attention and LSTM layers.

$$H_{SMILES}^{Trans} = \text{MultiHeadAttention}(H_{SMILES}^{LSTM})$$

$$H_{SMILES} = \text{LayerNormalization}(H_{SMILES}^{LSTM}, H_{SMILES}^{Trans})$$

$$F_{SMILES} = \text{LSTM}(H_{SMILES})$$

8. Similar steps 6 and 7 to be followed for the protein encoder.

$$F_{Proteins} = \text{LSTM}(H_{Proteins})$$

9. Concatenate the outputs of the SMILES and protein encoders.

$$F_{combined} = \text{Concatenate}(F_{SMILES}, F_{Protein})$$

10. Output the final interaction score, y' .

4. RESULTS AND DISCUSSION

4.1. EVALUATION METRICS

The performance of the model is evaluated using metrics such as accuracy, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R²), Area Under the Precision-Recall Curve (AUPR), and Concordance Index (CI). These metrics provide a comprehensive assessment of the model's ability to predict drug-target binding affinities accurately.

Mean Squared Error (MSE) measures the average squared difference between the actual and predicted values. It is a fundamental metric for regression problems, providing a clear indication of the model's prediction error.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad (8)$$

A lower MSE indicates better model performance, with an ideal value of 0, which would mean no error. High MSE values suggest that the model is not capturing the underlying trend in the data. The Concordance Index (CI) measures the agreement between predicted and actual rankings, commonly used in survival analysis and similar tasks. CI provides an evaluation of the ranking accuracy of predictions. CI values range from 0.5 (random chance) to 1 (perfect prediction). Values above 0.7 are generally considered good. Higher CI values indicate better model performance in terms of ranking predictions correctly.

4.2. PERFORMANCE OF THE PROPOSED FUSIONNET MODEL

The FusionNet model demonstrates strong performance in predicting drug-target interactions on both the Kiba and Davis datasets. The Kiba dataset, achieved a training loss and MSE of 0.1874, with validation loss and MSE of 0.1852, indicating good generalization without overfitting. On the Davis dataset, the model had a training loss and MSE of 0.2073, and a validation loss and MSE of 0.2296, showing slight deviations but maintaining strong performance.

The system architecture shown in Fig. 3. displays MSE over 100 epochs, where the training MSE drops sharply in the first 10 epochs, and the validation MSE stabilizes, demonstrating effective learning and good generalization. Fig. 4. depicts "Actual vs. Predicted Values" and "Residual Plots." For the Davis dataset, predicted values show more spread, and residuals indicate areas of inaccuracy, whereas, for Kiba, predictions are tightly clustered with residuals randomly distributed around zero, reflecting better accuracy. Overall, the model performs more accurately on the Kiba dataset.

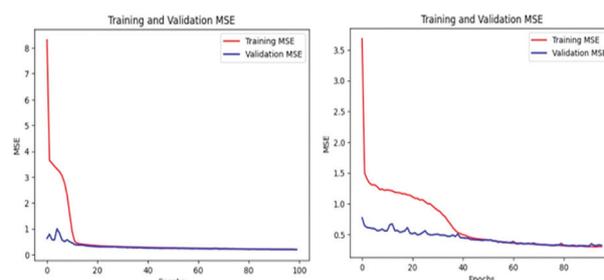


Fig. 3. Training and Validation MSE for Kiba and Davis Dataset

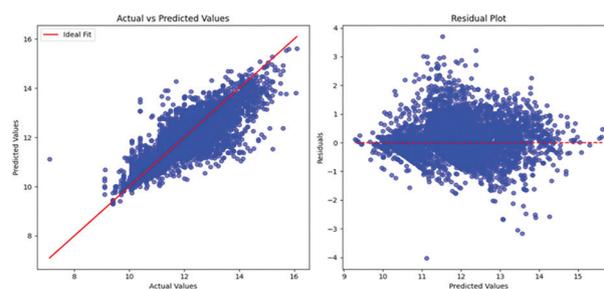


Fig. 4. Actual vs. Predicted Values and Residual Plots for Kiba dataset

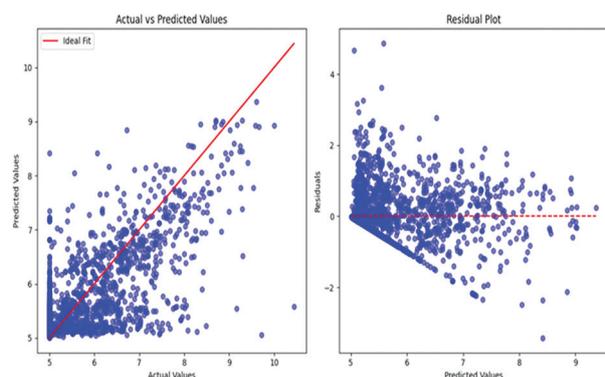


Fig. 5. Actual vs. Predicted Values and Residual Plots for Davis dataset for Davis dataset

Fig. 5 and Fig. 6 show two scatter plots of predicted vs. actual values for two models. The Kiba plot, with an R-squared of 0.710, indicates that 71% of the variance in actual values is explained by the model, showing a strong correlation and good predictive accuracy. The Davis plot, with an R-squared of 0.681, explains 68.1% of the variance, indicating a weaker correlation and lower predictive accuracy.

Fig. 7 displays two precision-recall curves for two models. The precision-recall curves for kiba, have an average precision (AP) of 0.80, showing a model that maintains high precision across various recall levels but drops off as recall approaches 1.0. The precision-recall curves for the Davis curve, have an AP of 0.86, indicating better overall performance, with higher precision maintained over a broader range of recall values.

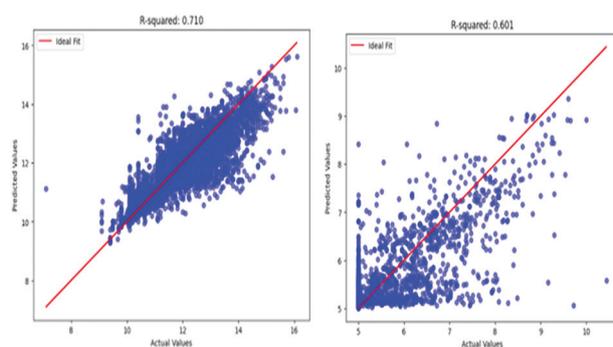


Fig. 6. Actual vs. Predicted Values for Kiba and Davis dataset

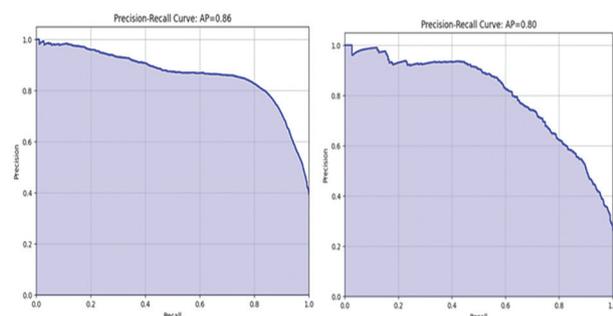


Fig. 7. Precision-recall curves for the Kiba and Davis dataset

Fig. 8 shows an AUC (Area Under the Curve) of 0.9 for the Davis dataset, indicating a high level of model performance with a good balance between sensitivity and specificity. The Fig. 13 shows an AUC of 0.91, which is slightly better, suggesting even better performance. Both models significantly outperform random guessing (represented by the red dashed line).

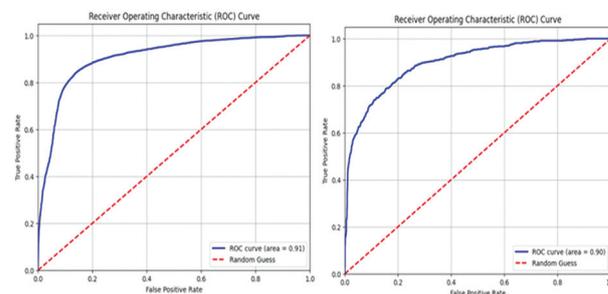


Fig. 8. ROC curve for Kiba and Davis dataset

The Protein Attention Map and SMILES Attention Map as in Fig. 9 provide valuable insights into the model's decision-making process by visualizing how attention is distributed across protein sequences and molecular structures, respectively. In the Protein Attention Map, each vertical strip represents specific positions in the protein sequence, with varying color intensities indicating the attention values assigned by the model. This helps capture important local and long-range dependencies essential for accurate drug-target binding predictions.

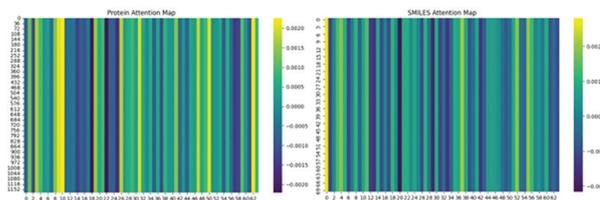


Fig. 9. Attention map for protein and smile sequences

Similarly, the SMILES Attention Map highlights attention distribution across molecular sequences represented in SMILES notation, where each vertical line corresponds to a token (atom or bond) in the molecule. The intensity reflects the importance given by the model to different parts of the molecular structure. These visualizations enhance interpretability, showing how the model focuses on crucial areas in both protein and molecular data, thereby improving transparency in the prediction process.

4.2. COMPARATIVE ANALYSIS OF VARIOUS APPROACHES

In this paper, the proposed FusionNet model has been compared to previous state-of-the-art approaches like SimBoost[21], GANsDTA[22], DeepCDA[23], and DeepDTA[11] using mean squared error (MSE), root mean squared error (RMSE), concordance index (CI),

area under the precision-recall curve (AUPR), and rm2 evaluation metrics, as shown in Table 2. In the case of the Davis dataset, the proposed model has obtained an MSE of 0.20 and an rm2 value of 0.681. This result showcases a clear improvement over existing methods by offering the lowest MSE and the highest rm2, indicating its ability to make more accurate predictions. In more detail, compared to the SimBoost, GANsDTA,

DeepCDA, and DeepDTA models with MSE values of 0.28, 0.27, 0.24, and 0.26, respectively, FusionNet demonstrates a higher prediction accuracy due to its ability to capture both local and long-range dependencies within drug-target sequences. Additionally, the rm2 values of these methods lag, further highlighting the efficacy of our approach.

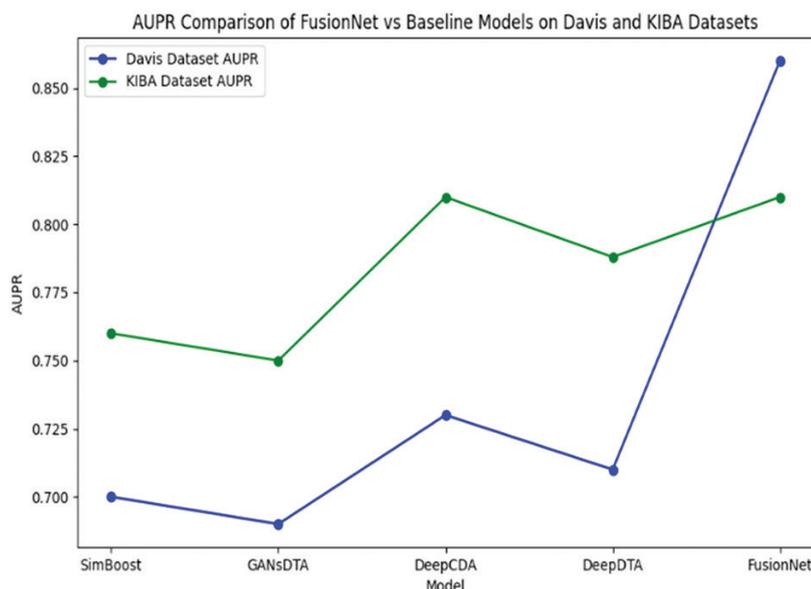


Fig 10. AUPR comparison graph of FusionNet's Vs baseline models on the Davis and KIBA datasets

Similarly, on the KIBA dataset, FusionNet outperforms existing methods, achieving an MSE of 0.18 and an rm2 of 0.71. The Transformer component allows FusionNet to model complex drug-target interactions, leading to more biologically meaningful predictions. For instance, while SimBoost, GANsDTA, DeepCDA, and DeepDTA returned MSE values of 0.22, 0.22, 0.17, and 0.19 respectively, FusionNet's multi-head attention mechanism captures intricate patterns across SMILES and protein sequences, improving generalization. Other metrics, such as root mean squared error and concordance

index, reflect this model's robust performance across datasets. Figure 10 provides a side-by-side AUPR comparison graph, highlighting FusionNet's performance against baseline models on Davis and KIBA datasets.

The biological implication is substantial, as more accurate drug-target affinity predictions enhance early-stage drug discovery by efficiently identifying potential drug candidates. In summary, FusionNet's superior accuracy and generalizability make it highly promising for accelerating therapeutic discovery.

Table 2. Comparative analysis of various approaches

Datasets	Methods	MSE	RMSE	CI	AUPR	rm2
Davis	SimBoost	0.28	-	0.83	0.70	0.644
	GANsDTA	0.27	-	0.70	0.69	0.653
	DeepCDA	0.24	-	0.89	0.73	0.649
	DeepDTA	0.26	-	0.87	0.71	0.67
	Proposed (FusionNet)	0.20	0.55	0.89	0.86	0.681
KIBA	SimBoost	0.22	-	0.83	0.76	0.629
	GANsDTA	0.22	-	0.86	0.75	0.675
	DeepCDA	0.17	-	0.88	0.81	0.682
	DeepDTA	0.19	-	0.86	0.788	0.673
	Proposed (FusionNet)	0.18	0.40	0.88	0.81	0.71

5. CONCLUSION

This work presents a new deep-learning framework, called FusionNet, which substantially improves the prediction of drug-target binding affinity. By combining Convolutional Neural Networks, Long Short-Term Memory networks, and transformers, it effectively represents local patterns, sequential dependencies, and long-range interactions in biological sequences. The model's novelty lies in its hybrid architecture, which leverages the strengths of each component to achieve enhanced robustness and accuracy in prediction. It also integrates the Layer-wise Adaptive Moments optimizer to improve training efficiency and performance, making the model scalable and reliable even for large datasets. Extensive tests were conducted with FusionNet on both the Davis and KIBA datasets, yielding state-of-the-art results. FusionNet achieved an MSE of 0.20 and an rm2 of 0.681 on the Davis dataset, and an MSE of 0.18 and an rm2 of 0.71 on the KIBA dataset. These results highlight FusionNet's superior performance, surpassing existing methods like SimBoost, GANsDTA, and DeepDTA. Unlike traditional methods that often focus on either local or global context, FusionNet addresses both simultaneously.

Biologically, this enhanced accuracy could streamline drug discovery pipelines by enabling more precise identification of potential drug-target pairs earlier in the development process, potentially reducing costs and timeframes. Additionally, the LAMB optimizer not only accelerates the training process but also enhances stability and performance on large datasets, underscoring FusionNet's scalability and robustness. Key takeaways include the hybrid model's capability to capture both local and global features, leading to improved predictive accuracy, and the model's scalability for larger datasets, which is essential for real-world applications. However, a notable limitation is FusionNet's reliance on sequence-based features without considering structural or genomic data, which might offer further insights into drug-target interactions. Future work could address this by expanding the dataset to include more diverse biological sequences and incorporating additional data such as 3D protein structures or genomic information. Scaling the model to larger datasets or alternative binding data types could present challenges that need further exploration. Another avenue could involve the use of transfer learning to leverage pre-trained models for related tasks. Finally, enhancing interpretability, such as linking attention to biologically relevant protein or molecular regions, could provide deeper insights into the biological mechanisms underlying drug-target interactions, increasing the model's value in biomedical research. This will not only establish FusionNet as a robust predictor but also enrich our understanding of molecular biology.

ACKNOWLEDGMENTS

The authors would like to thank all anonymous reviewers for their constructive advice.

FUNDING

This research received no external funding

6. REFERENCES

- [1] B. Rupp, "Biomolecular crystallography: principles, practice, and application to structural biology", Garland Science, 2009.
- [2] R. L. Rich, D. G. Myszka, "Survey of the year 2005 commercial optical biosensor literature", *Journal of Molecular Recognition: An Interdisciplinary Journal*, Vol. 19, No. 6, 2006, pp. 478-534.
- [3] A. Velazquez-Campoy, E. Freire, "Isothermal titration calorimetry to determine association constants for high-affinity ligands", *Nature Protocols*, Vol. 1, No. 1, 2006, pp. 186-191.
- [4] L. Breiman, "Random forests", *Machine Learning*, Vol. 45, 2001, pp. 5-32.
- [5] D. Rogers, M. Hahn, "Extended-connectivity fingerprints", *Journal of Chemical Information and Modeling*, Vol. 50, No. 5, 2010, pp. 742-754.
- [6] H. Luo, Y. Xiang, X. Fang, W. Lin, F. Wang, H. Wu, H. Wang, "BatchDTA: implicit batch alignment enhances deep learning-based drug-target affinity estimation", *Briefings in Bioinformatics*, Vol. 23, No. 4, 2022.
- [7] T. Cheng, Q. Li, Z. Zhou, Y. Wang, S. H. Bryant, "Structure-based virtual screening for drug discovery: a problem-centric review", *The AAPS Journal*, Vol. 14, 2012, pp. 133-141.
- [8] M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri, D. R. Koes, "Protein-ligand scoring with convolutional neural networks", *Journal of Chemical Information and Modeling*, Vol. 57, No. 4, pp. 942-957, 2017.
- [9] Y. You, I. Gitman, B. Ginsburg, "Large batch training of convolutional networks", arXiv:1708.03888, 2017.
- [10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, "Attention is all you need", *Advances in Neural Information Processing Systems*, Vol. 30, 2017.
- [11] H. Öztürk, A. Özgür, E. Ozkirimli, "Deepdta: deep drug-target binding affinity prediction", *Bioinformatics*, Vol. 34, No. 17, 2018, pp. i821-i829.

- [12] Q. Feng, E. Dueva, A. Cherkasov, M. Ester, "Padme: A deep learning-based framework for drug-target interaction prediction", arXiv:1807.09741, 2018.
- [13] H. Öztürk, E. Ozkirimli, A. Özgür, "Widedta: prediction of drug-target binding affinity", arXiv:1902.04166, 2019.
- [14] Y. Zeng, X. Chen, Y. Luo, X. Li, D. Peng, "Deep drug-target binding affinity prediction with multiple attention blocks", *Briefings in Bioinformatics*, Vol. 22, No. 5, 2021.
- [15] I. Elansary, W. Hamdy, A. Darwish, H. Aboul, "Bat-inspired optimizer for prediction of anti-viral cure drug of Sars-Cov-2 based on recurrent neural network", *Journal of System and Management Sciences*, Vol. 10, No. 2020, pp. 20-34.
- [16] E. K. Makowski et al. "Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space", *Nature Communications*, Vol. 13, No. 1, 2022, p. 3788.
- [17] S. Min, B. Lee, S. Yoon, "Deep learning in bioinformatics", *Briefings in Bioinformatics*, Vol. 18, No. 5, 2017, pp. 851-869.
- [18] M. I. Davis, J. P. Hunt, S. Herrgard, Y. You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, C.-J. Hsieh, "Large batch optimization for deep learning: Training Bert in 76 minutes", arXiv:1904.00962, 2019.
- [19] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization", arXiv:1412.6980, 2014.
- [20] M. I. Davis, J. P. Hunt, S. Herrgard, P. Ciceri, L. M. Wodicka, G. Pallares, M. Hocker, D. K. Treiber, P. P. Zarrinkar, "Comprehensive analysis of kinase inhibitor selectivity", *Nature Biotechnology*, Vol. 29, No. 11, 2011, pp. 1046-1051.
- [21] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, M. Ester, "SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines", *Journal of Cheminformatics*, Vol. 9, No. 1, 2017, pp. 1-14.
- [22] T. Song, X. Zhang, M. Ding, A. Rodriguez-Paton, S. Wang, G. Wang, "DeepFusion: A deep learning based multi-scale feature fusion method for predicting drug-target interactions", *Methods*, Vol. 204, 2022, pp. 269-277.
- [23] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J. B. Ghasemi, A. Masoudi-Nejad. "DeepCDA: deep cross-domain compound-protein affinity prediction through LSTM and convolutional neural networks", *Bioinformatics*, Vol. 36, No. 17, 2020, pp. 4633-4642.