Enhancing Antibody Antigen Interaction Efficiency Through AI Based Computational Approaches

Pavani Devi M[1], Jayasree Pinajala[2], Nitalaksheswara Rao Kolukula[3*], James Stephen Meka[4], Pavan Satish Chandaka[5]

[1]Assistant Professor, Department of Computer Science, Godavari Global University, GIET Campus, Rajamahendravaram, Andhra Pradesh, India

[2,5]Department of Computer Science and Engineering, Chaitanya Engineering College, Visakhapatnam, Andhra Pradesh, India

[3]Department of Computer Science and Engineering, School of computer Science and Engineering, GITAM University, Visakhapatnam, Andhra Pradesh, India

[4]Ambedkar Chair Professor, Andhra University, Visakhapatnam, Andhra Pradesh, India

[*]Corresponding Author

DOI: https://dx.doi.org/10.51244/IJRSI.2026.130200202

Received: 05 March 2026; Accepted: 10 March 2026; Published: 23 March 2026

ABSTRACT

Improving antibody-antigen interactions is critical for therapeutic antibody development. The artificial intelligence-powered method improves binding affinity and specificity by leveraging deep learning and structural bioinformatics. Despite these advances, significant hurdles remain, including the difficulty of simulating dynamic interactions under physiological settings, the scarcity of data for uncommon antigens, and the computational demands of structural predictions.

To address these issues, this paper combines Variational Auto Encoders (VAE), transformers, and graph-based models into a single pipeline, resulting in enhanced structure prediction and binding affinity estimation on benchmark datasets. Specifically, transformer-based models such as Alphafold and RoseTTAFold are employed to predict antibody structures, focusing particularly on variable regions like the CDR-H3 loop, while Graph Neural Networks (GNN) and Graph Transformer Networks (GTN) are used to model complex binding interfaces.

The proposed method achieves the RMSE of 0.15 and MAE of 0.10, indicating the low error rate. These findings demonstrate the system potential to produce structurally stable, high-affinity antibodies while also greatly speeding up rational therapeutic antibody design.

Keyword: Antibody-Antigen Interaction, Deep Learning, Variational Auto Encoder, Graph Neural Networks, Complementarity Determining Regions Optimization

INTRODUCTION

Enhancing antibody affinity is crucial for improving therapeutic efficacy. Testing mutations within the Complementarity-Determining Regions (CDR) helps identify those that can boost binding strength. [1]. AI-based computational approaches enhance antibody-antigen interaction efficiency by accurately predicting interactions from sequence data, advancing therapeutic development [2].

Big data and deep sequencing show intricate immunological patterns, which allows for the computer construction of precise synthetic antibodies [3]. Antibody-antigen modeling use deep learning to predict protein

interactions with excellent accuracy. This method improves the precision of antibody-antigen complex modeling used in therapeutic design [4]. A FASTIA for swiftly analyzing protein variant interactions is presented, with examples of its use in understanding single-domain antibody interactions [5]. Machine learning is used to predict the binding affinity of antibody-protein antigen complexes. Interface and surface area features are critical in improving prediction performance [6]. Machine learning is utilized to extract features associated with antibody-antigen binding affinity. Learned representations help in the knowledge and prediction of molecular interactions [7]. Sequence-to-structure prediction approaches for antibody-antigen complexes are evaluated to determine their effectiveness. The analysis identifies systematic biases that influence prediction accuracy [8].

Contribution

- This study aims to enhance antibody-antigen interaction efficiency using advanced AI-based computational approaches.
- Data from the OAS database is integrated with deep learning models such as VAE, Transformers, and Graph Neural Networks for structure prediction and interaction analysis.
- The approach enables accurate modeling and optimization of antibody binding regions, particularly CDR loops, to improve specificity and affinity.
- The framework supports rational design of therapeutic antibodies with enhanced binding capabilities through AI-driven structural insights.

This study consists of multiple sections, each offering unique insights. Section 2 provides a review of the literature, and Section 3 explains the proposed model. In Section 4, the results of the recommended processes are shown, and this model is compared to a few other contemporary approaches and Section 5 offers a conclusion about recommendations for future work.

LITERATURE SURVEY

Zhang et al. (2024) [9] used machine learning models to predict and assess the structural interactions between antibodies and antigens, which improved the accuracy of their research. Rouyan et al. (2025) [10] evaluated the application of AI and ML in antibody discovery, emphasizing their importance in expediting the identification and optimization procedures. Chaves et al. (2025) [11] discussed the problems and advances in structure-based computational design for antibody mimetics.

Yuan et al. (2023) [12] introduced Dg-affinity, a language model-based technique for predicting antigen-antibody affinity from sequences. Gallo (2025) [13] investigated the potential of AI and deep sequencing large data to revolutionize synthetic antibody design. Li et al. (2025) [14] improved antigen-antibody binding interface prediction with ESM-2 and Bi-LSTM, obtaining greater accuracy than previous techniques.

Problem statement

AI-based therapeutic antibody design faces obstacles such as poor data quality, limited model generalization, and complex antigen-antibody predictions. To overcome these issues, rigorous preprocessing produces cleaner, more uniform datasets, while models such as Variational Auto Encoders and AlphaFold increase prediction accuracy. Graph Neural Networks improve antibody-antigen interaction predictions, allowing for more dependable and tailored therapeutic designs.

PROPOSED METHOD

To improve antibody-antigen interaction efficiency, data from the Observed Antibody Space (OAS) dataset is processed, followed by sequence cleaning and feature extraction using VAE, resulting in 3D antibody structures. Advanced models like as AlphaTTAFold and GTNN are then used to forecast antibody structures and optimize interactions, with a focus on improving CDRs for increased specificity and affinity. The overall proposed work shows figure 1.

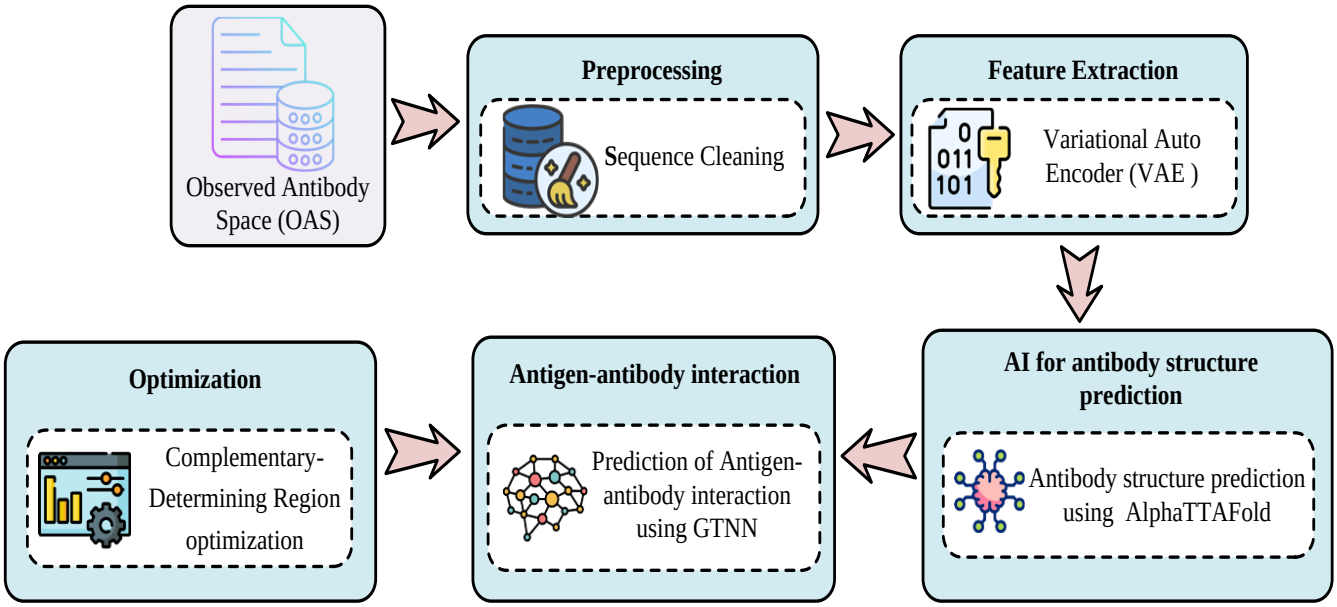

Figure 1. The overall proposed diagram

Data Collection

The Observed Antibody Space (OAS) database [15] contains an extensive collection of annotated antibody sequencing data from various immunological repertoires. It enables AI-driven modeling to optimize antibody-antigen interactions by giving real-world sequence patterns and variability information.

Preprocessing

In the preprocessing stage, the input data is cleaned to remove confusing sequences and non-canonical amino acids.

Sequence cleaning: Sequence cleaning [16] ensures that AI models receive high-quality input by eliminating confusing sequences and non-standard amino acids that may cause errors. Alignment arranges sequences to highlight conserved sections, whereas trimming ensures that all sequences have consistent lengths for reliable comparison. These steps reduce noise, improve model learning, and increase prediction reliability. Clean, well-aligned data is essential for analyzing antibody-antigen interactions computationally. The mathematical representation of sequence cleaning is shows equation 1.

$$S^{'} = T(A(C(S))) \quad (1)$$

where S represent the input, C(S) represent the Cleaning function that removes non-canonical amino acids and noisy sequences, T(A(C(S))) is represent the Trimming function that standardizes sequence lengths after alignment and $S^{'}$ is represent the final output of the preprocessing. After preprocessing the sequencing data is passed into the feature extraction.

Feature Extraction using Variational Auto Encoder (VAE)

VAEs extract significant information by encoding antibody sequence and structural data into a structured latent space that groups functionally similar patterns together [17]. This allows for the identification of sequence motifs and structural properties that improve antigen binding effectiveness. VAEs enable AI-driven antibody design for better antigen contact by capturing both sequencing and 3D information. The process of VAE shows equation 2.

$$L_{VAE} = \mathrm{E}_{q\phi(z|x)}[\log p_{\theta}(x|z)] - D_{KL}(q_{\phi}(z|x) \| p(z)) \quad (2)$$

where x is a input data, z is a latent vector represent feature extraction, $q_\phi(z \mid x)$ is a encoder, $p_\theta(x \mid z)$ is a decoder, p(z) represent the prior over the latent space and D_{KL} is represent Kullback-Leibler divergence ensuring regularization of the latent space.

AI for antibody structure prediction

Deep learning methods are used to properly predict the 3D structures of antibodies from their amino acid sequences. Understanding antibody-antigen interactions helps to speed up the discovery of medicines and design effective therapies.

AlphaFold: AlphaFold allows for the accurate modeling of 3D antibody structures directly from amino acid sequence. It identifies critical binding areas, helping in AI-powered antibody engineering and optimization. AlphaFold highlight how artificial intelligence can speed up therapeutic antibody analysis by simplifying structure prediction. The process can be representing in equation 3.

$$\hat{S} = f_{AI}(A) \qquad (3)$$

where $\hat{S}$ is represented as predicted 3D structure of antibody , A is an amino acid sequence of the antibody and f_{AI} is the AI-based predictive function

RoseTTAFold: RoseTTAFold uses deep learning to predict 3D structures by processing sequence and structural information using multi-track neural networks, with a focus on antibody variable regions such as CDR. Equation 4 is RoseTTAFold process:

$$Q(e_{x,y}) = Soft\max(V_{x,y}) \qquad (4)$$

where $e_{x,y}$ is predict the distance between x and y, $V_{x,y}$ is the neural output for x and y and $Soft\max$ is convert logits into a probability distribution over distance bins

Transformer-based models for AlphaTTAFold: AlphaTTAFold is a transformer-based method for accurately predicting antibody structures. AlphaTTAFold is derived from AlphaFold [18] and RoseTTAFold [19], and it focuses on improving predictions in variable regions like CDR-H3. It uses attention mechanisms to model complex spatial relationships and long-range interactions in antibody sequences. This results in faster, more accurate predictions, which are important for therapeutic antibody design. The antibody structure prediction shows figure 2.

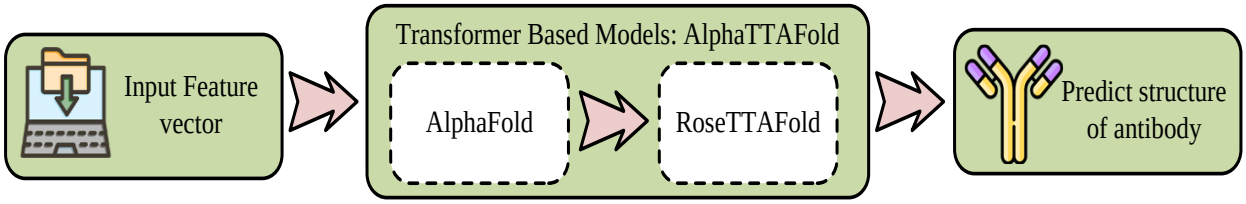

Figure2. Antibody structure prediction

Prediction of Antibody-antigen interaction

The antibody-antigen connection is critical for immunological defense, diagnosis and therapy, with new techniques enhancing antibody performance [20]. Predicting these interactions is critical to immunology and medication design. Deep learning now allows for precise modeling of their complicated structural interactions.

Graph Neural Network: Graph Neural Networks (GNNs) represent antibody-antigen complexes as graphs, with residues as nodes and interactions as edges. This allows them to capture complex spatial and biochemical

relationships for accurate binding affinity prediction. The equation 5 representing how a Graph Neural Network (GNN) updates node features in antibody-antigen interaction is follow:

$$h_v^{(t+1)} = \sigma\left(\sum_{u \in N(v)} W.h_u^t + b \right) \quad (5)$$

where $h_v^{(t+1)}$ is a updated feature of node v at layer t+1, N(v) is a neighbor node of v, W is a learnable weight and b is a bias term σ is a activation function.

Graph Transformer Network: AI-based approaches model antibody-antigen interactions as graphs, with Graph Transformer Networks capturing complex patterns using attention mechanisms. These models predict binding scores, improving interaction efficiency and aiding therapeutic antibody design. The equation 6 is antibody-antigen interactions using GTN:

$$h_i^{'} = \sum_{j \in \mathrm{N}(i)} \alpha ij W h_j \quad (6)$$

where $h_i^{'}$ is the updated feature for node i, N(i) is the set of neighbor of node i, h_j is the input feature of neighbor node j, W is a learnable weight matrix and αij is the attention coefficient between node i and j.

Graph-based deep learning model for Graph TransNeural Networks (GTNNs): Graph-based deep learning models are critical to analyzing complicated biological systems such as protein-protein interactions. Models like Graph Neural Networks (GNNs) [21] and Graph Transformer Networks (GTNs) [22] detect crucial molecule structural patterns. Graph TransNeural Networks (GTNNs) extend these methods by utilizing self-attention to detect critical regions such as paratopes and epitopes in antibody-antigen interactions. This results in more accurate predictions and promotes the production of effective therapeutic antibodies. The final process equation 7 is follow:

$$h_v^{(t+1)} = \sigma\left(\sum_{u \in N(v)} \alpha_{vu} W.h_u^t + b \right) \quad (7)$$

After predict the antibody-antigen interaction the output of interaction A_i is fed to the CDR for training. The prediction of antibody-antigen interaction shows figure 3.

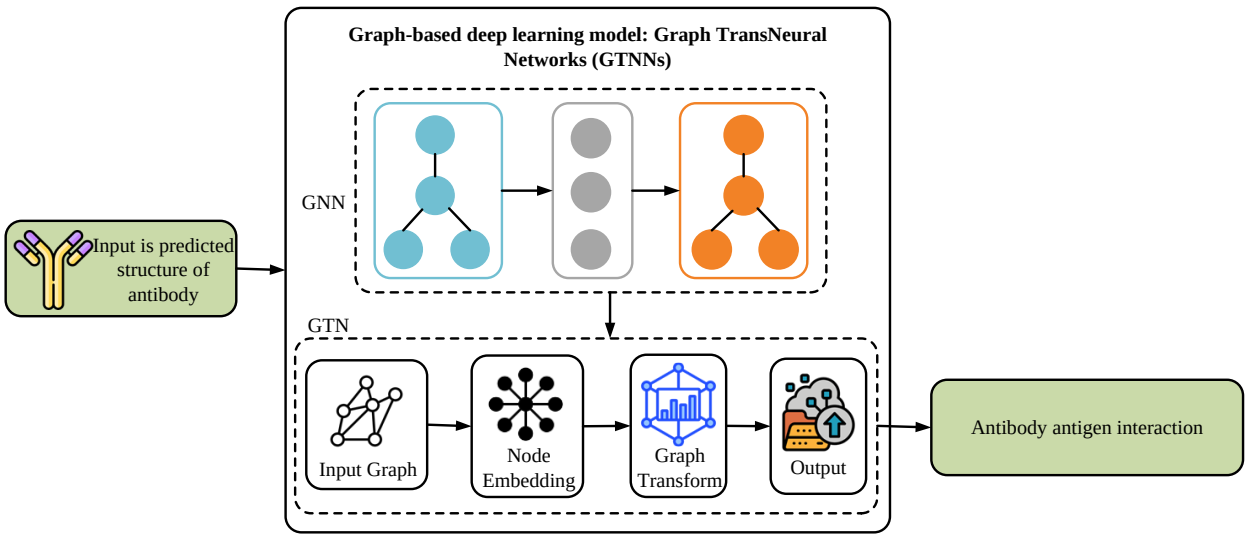

Figure3. Prediction of antibody-antigen interaction

Complementarity Determining Regions (CDR) for training Graph TransNeural Networks (GTNN)

Complementarity Determining Region (CDR) [23] is antibody components that directly connect to an antigen, making them essential for immunological detection. CDR is used to represent interaction patterns between antibodies and antigens while training Graph TransNeural Network (GTNN). Using CDR focused data allows GTNN to learn biologically relevant characteristics for better prediction performance. The optimal solution is calculated with error can be represented in equation 8.

$$RMSE_{error} = \frac{1}{n}\sum_{h=1}^{n}[C_r - A_i] \quad (8)$$

where C_r denoted the final output and A_i is the predicted antibody-antigen interaction.

RESULTS AND DISCUSSIONS

AI-driven models significantly improved antibody-antigen interaction efficiency, with enhanced prediction accuracy and optimized binding through deep learning techniques. Evaluation metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and AUC confirmed improved predictive accuracy and structural optimization of antibody binding sites. The system features an Intel Core i3, 32GB RAM, 1TB SSD and high-speed internet, running on Windows 11 with Python 3.12 and libraries like TensorFlow and PyTorch. Google Cloud Storage and Compute handle data storage and cloud services, while PostgreSQL serves as the database. Pandapower is recommended for power system analysis due to its computational capabilities.

Comparison Analysis

Table1. Dataset comparison of OAS

Methods	**RMSE**	**MAE**
IgG4 [24]	0.42	0.48
SAbDab [25]	0.23	0.35
Proposed OAS	0.15	0.10

The table 1 shows the performance of the different methods based on RMSE and MAE. Lower values indicate better performance of the OAE method achieves the best results with the lowest RMSE (0.15) and MAE (0.10). This shows the proposed method significantly outperforms existing approaches like AbDb, sdAb, IgG4 and SAbDab.

Table2. Error Comparison of AlPhaFold

Method	RMSE	MAE
DeepFoldNA	0.29	0.28
DeepAb	0.22	0.20
Proposed	0.10	0.12

The table 2 shows the performance of three methods using RMSE and MAE values. The proposed method shows the best performance with RMSE of 0.10 and MAE of 0.12. DeepAb follows with RMSE of 0.22 and MAE of 0.20, while DeepFoldNA has the highest errors (RMSE: 0.29, MAE: 0.28).

Table3. Error Comparison of GTNN

Method	RMSE	MAE
CNN	0.33	0.36
GAN	0.25	0.24
Proposed	0.11	0.16

The table 3 shows the performance of three methods using RMSE and MAE values. The proposed method shows the best performance with RMSE of 0.11 and MAE of 0.16. GAN follows with RMSE of 0.25 and MAE of 0.24, while CNN has the highest errors (RMSE: 0.33, MAE: 0.36).

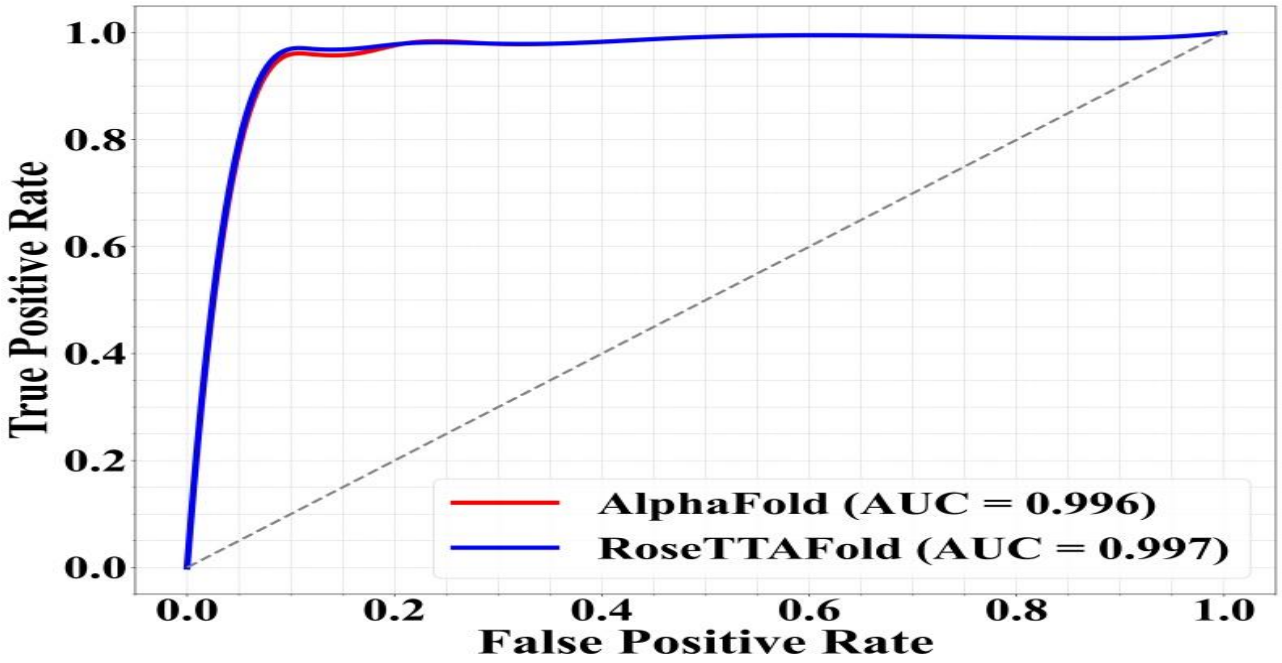

Figure 4. ROC curve for antibody structure prediction model

Figure 4 shows the Receiver Operating Characteristic (ROC) curves for two antibody structure prediction models: AlphaFold and RoseTTAFold. Both models perform exceptionally well, with Area Under Curve (AUC) values of 0.996 and 0.997, respectively. RoseTTAFold slightly outperforms AlphaFold based on AUC.

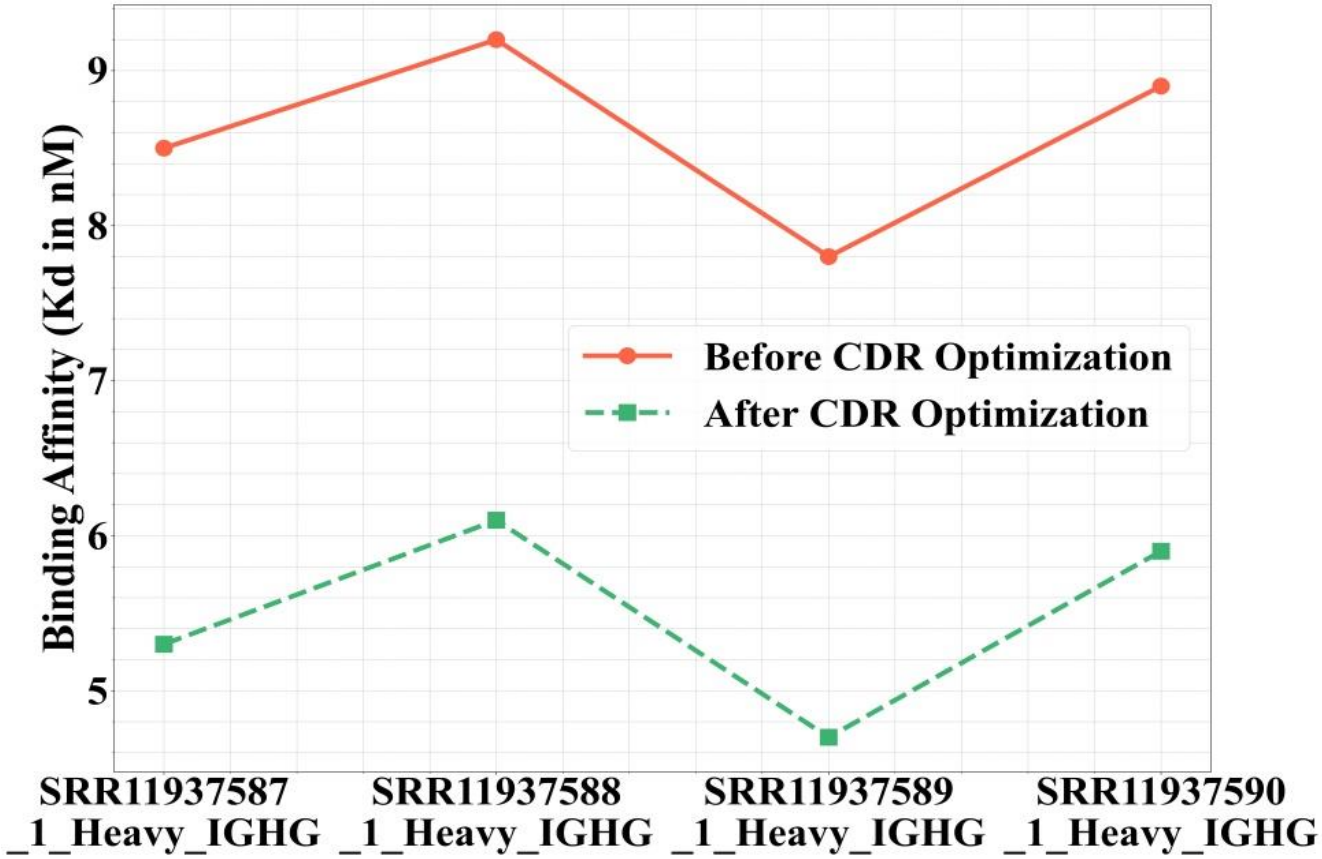

Figure 5.CDR Optimization

Figure 5 shows the binding affinity (Kd in nM) of antibodies before and after CDR optimization. The red line shows higher Kd values before optimization, indicating weaker binding. After CDR optimization (green dashed line), all samples show improved (lower) Kd values, reflecting stronger binding affinity.

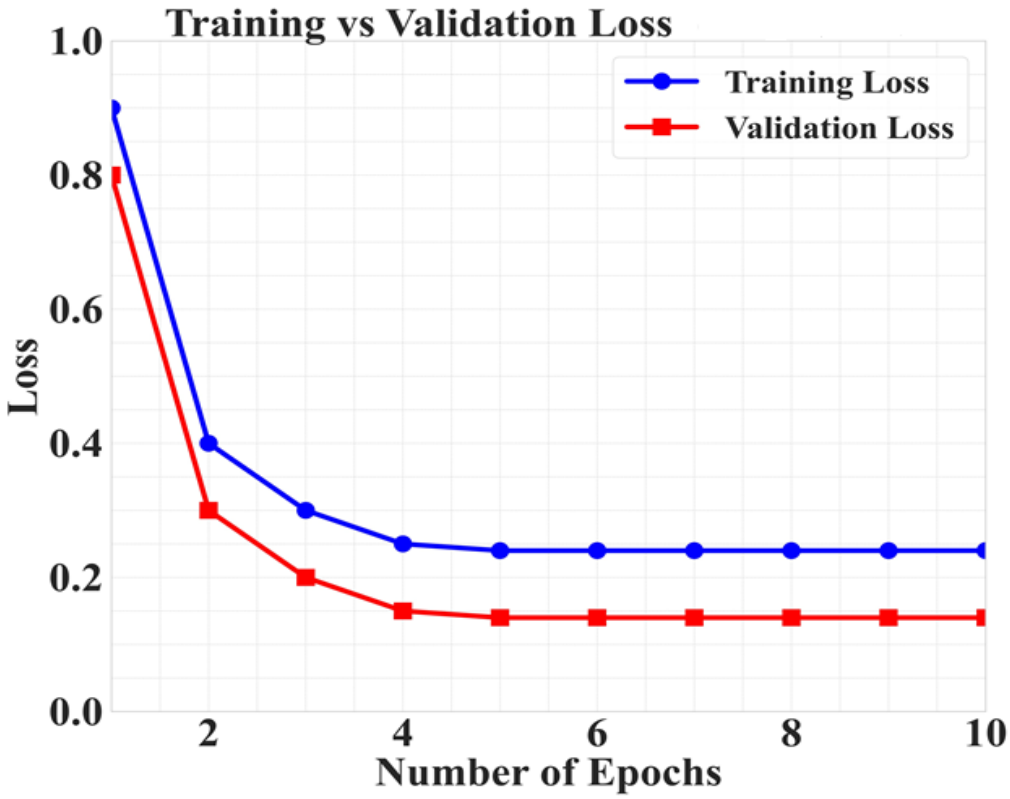

Figure 6. Training vs. validation Loss

The figure 6 shows that training loss drops from about 0.9 to 0.24 and validation loss drops from about 0.7 to 0.14 over 10 epochs. Both losses stabilize after epoch 4, indicating convergence. Validation loss stays lower than training loss throughout, suggesting good generalization and no over fitting.

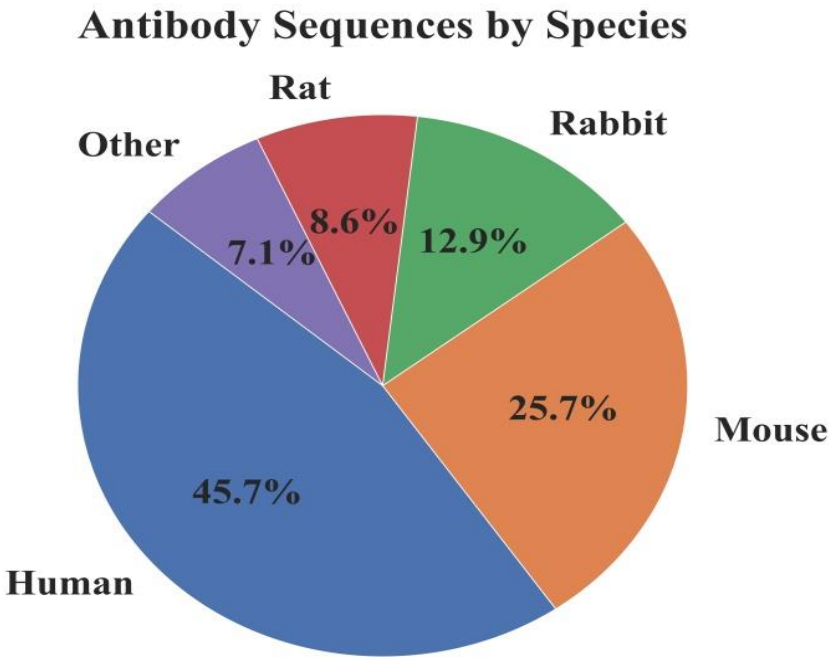

Figure 7. Antibody sequence by species

Figure 7 shows the most antibody sequences come from Humans (45.7%), followed by Mouse (25.7%), and Rabbit (12.9%). Smaller contributions are from Rat (8.6%) and other species (7.1%).

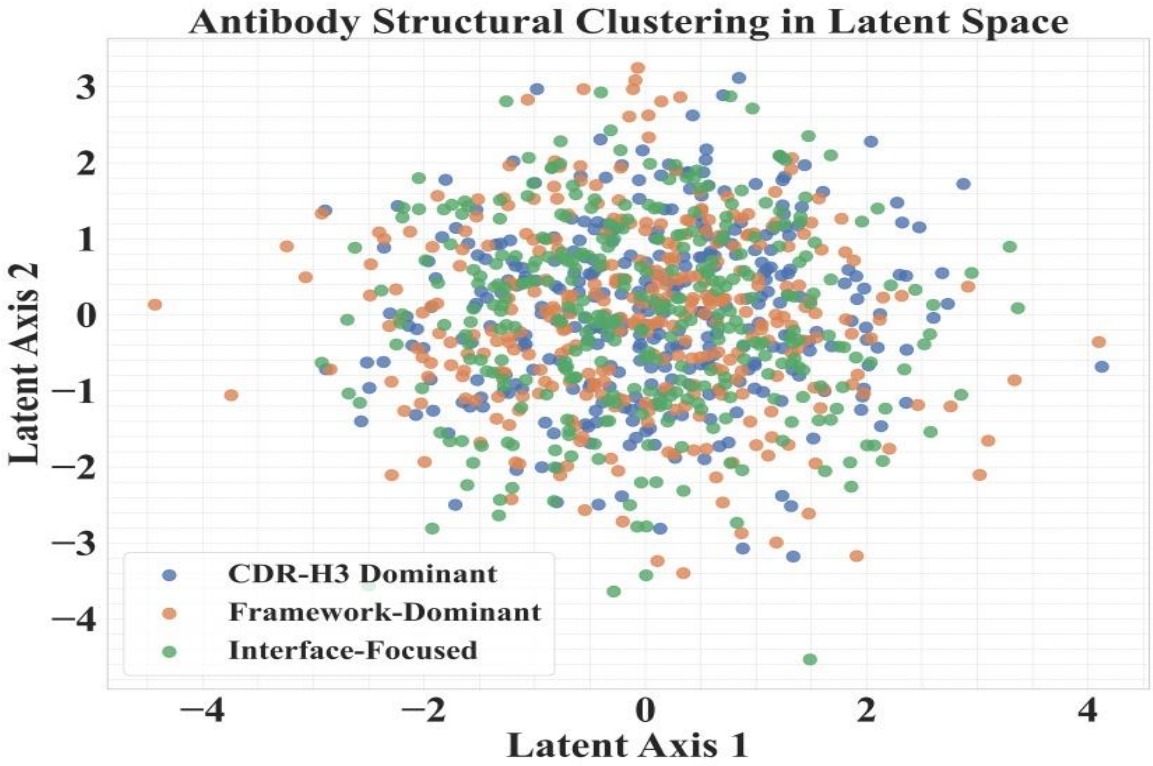

Figure 8. Antibody clustering in latent space

The figure 8 shows a clustering of antibody structures in a latent space, with axes representing different antibody features like CDR-H3 loops, framework regions, and binding interfaces. The plot likely visualizes how antibodies group based on these structural or functional characteristics.

CONCLUSIONS

In conclusion, AI-based computational techniques provide a transformative pathway for increasing antibody-antigen interaction efficiency, from data collecting to structure prediction and optimization. The merging of the VAE, AlphaTTAFold and GTNN models greatly improved prediction accuracy. As these techniques progress, the future scope will include model improvement, target antigen expansion and accelerated tailored antibody development. Finally, these advancements offer the potential to transform antibody-based medicine and immunotherapy.

REFERENCE

1. Clark, T., Subramanian, V., Jayaraman, A., Fitzpatrick, E., Gopal, R., Pentakota, N., Rurak, T., Anand, S., Viglione, A., Raman, R. and Tharakaraman, K., 2023. Enhancing antibody affinity through experimental sampling of non-deleterious CDR mutations predicted by machine learning. Communications Chemistry, 6(1), p.244. Doi: https://doi.org/ 10.1080 /19420862 .2024 .2352887
2. Xia, Y., Wang, Z., Huang, F., Xiong, Z., Wang, Y., Qiu, M. and Zhang, W., 2025. DeepInterAware: Deep Interaction Interface-Aware Network for Improving Antigen-Antibody Interaction Prediction from Sequence Data. Advanced Science, p.2412533. Doi: https:// doi.org /10.1002/advs.202412533
3. Gallo, E., 2024. The rise of big data: deep sequencing-driven computational methods are transforming the landscape of synthetic antibody design. Journal of Biomedical Science, 31(1), p.29. Doi: https://doi.org/10.1186/s12929-024-01018-5

4. Yin, R. and Pierce, B.G., 2024. Evaluation of AlphaFold antibody–antigen modeling with implications for improving predictive accuracy. Protein Science, 33(1), p.e4865. Doi: https://doi.org/10.1002/pro.4865
5. Matsunaga, R. and Tsumoto, K., 2025. FASTIA: A rapid and accessible platform for protein variant interaction analysis demonstrated with a single-domain antibody. Protein Science, 34(3), p.e70065. Doi: https://doi.org/10.1002/pro.70065
6. Yang, Y.X., Wang, P. and Zhu, B.T., 2023. Binding affinity prediction for antibody–protein antigen complexes: A machine learning analysis based on interface and surface areas. Journal of Molecular Graphics and Modelling, 118, p.108364. Doi: https://doi.org/ 10.1016/j .jmgm. 2022. 108364
7. Miller, N.L., Clark, T., Raman, R. and Sasisekharan, R., 2023. Learned features of antibody-antigen binding affinity. Frontiers in Molecular Biosciences, 10, p.1112738. Doi: https://doi.org/10.3389/fmolb.2023.1112738
8. McCoy, K.M., Ackerman, M.E. and Grigoryan, G., 2024. A comparison of antibody–antigen complex sequence-to-structure prediction methods and their systematic biases. Protein Science, 33(9), p.e5127. Doi: https://doi.org/10.1002/pro.5127
9. Zhang, G., Kuang, X., Zhang, Y., Liu, Y., Su, Z., Zhang, T. and Wu, Y., 2024. Machine-learning-based structural analysis of interactions between antibodies and antigens. BioSystems, 243, p.105264. Doi: https://doi.org/10.1016/j.biosystems.2024.105264
10. Rouyan, A., Dudzic, P., Wilman, W., Satława, T., Wróbel, S. and Krawczyk, K., 2025. Applications of Artificial Intelligence and Machine Learning toward Antibody Discovery and Development. In Biopharmaceutical Informatics (pp. 77-115). CRC Press. Doi: https:// doi. org/ 10.1201/9781003300311
11. Chaves, E.J., Coêlho, D.F., Cruz, C.H., Moreira, E.G., Simões, J.C., Nascimento-Filho, M.J. and Lins, R.D., 2025. Structure-based computational design of antibody mimetics: challenges and perspectives. FEBS Open Bio, 15(2), pp.223-235. Doi: https://doi.org/10.1002/2211-5463.13855
12. Yuan, Y., Chen, Q., Mao, J., Li, G. and Pan, X., 2023. Dg-affinity: predicting antigen–antibody affinity with language models from sequences. BMC bioinformatics, 24(1), p.430. Doi: https://doi.org/10.1186/s12859-023-05562-z
13. Gallo, E., 2025. Revolutionizing synthetic antibody design: Harnessing artificial intelligence and deep sequencing big data for unprecedented advances. Molecular Biotechnology, 67(2), pp.410-424. Doi: https://doi.org/10.1007/s12033-024-01064-2
14. Li, Q., Zhao, Y., Chordia, M.D., Xia, X., Zhang, B. and Zheng, H., 2025. Enhanced prediction of antigen and antibody binding interface using ESM-2 and Bi-LSTM. Human Immunology, 86(3), p.111304. Doi: https://doi.org/10.1016/j.humimm.2025.111304
15. Olsen, T.H., Boyles, F. and Deane, C.M., 2022. Observed Antibody Space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. Protein Science, 31(1), pp.141-146. Doi: https://doi.org/10.1002/pro.4205
16. Milon, T.I., Sarkar, T., Chen, Y., Grider, J.M., Chen, F., Ji, J.Y., Jois, S.D., Kousoulas, K.G., Raghavan, V. and Xu, W., 2025. Development of the TSR-based computational method to investigate spike and monoclonal antibody interactions. Frontiers in Chemistry, 13, p.1395374. Doi: https://doi.org/10.3389/fchem.2025.1395374
17. Kim, D.N., McNaughton, A.D. and Kumar, N., 2024. Leveraging artificial intelligence to expedite antibody design and enhance antibody–antigen interactions. Bioengineering, 11(2), p.185. Doi: https://doi.org/10.3390/bioengineering11020185
18. Dewaker, V., Morya, V.K., Kim, Y.H., Park, S.T., Kim, H.S. and Koh, Y.H., 2025. Revolutionizing oncology: the role of Artificial Intelligence (AI) as an antibody design, and optimization tools. Biomarker Research, 13(1), p.52. Doi: https://doi.org/10.1186/s40364-025-00764-4
19. Baek, M., McHugh, R., Anishchenko, I., Jiang, H., Baker, D. and DiMaio, F., 2024. Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. Nature methods, 21(1), pp.117-121. Doi: https://doi.org/10.1038/s41592-023-02086-5
20. Montoya, R., Deckerman, P. and Guler, M.O., 2025. Protein Recognition Methods for Diagnostics and Therapy. BBA Advances, p.100149. Doi: https://doi.org/10.1016/j.bbadva. 2025.100149
21. Bai, G., Sun, C., Guo, Z., Wang, Y., Zeng, X., Su, Y., Zhao, Q. and Ma, B., 2023, October. Accelerating antibody discovery and design with artificial intelligence: Recent advances and prospects. In Seminars

in Cancer Biology (Vol. 95, pp. 13-24). Academic Press. Doi: https://doi.org/10.1016/j.semcancer.2023.06.005

22. Wang, F., Dai, X., Shen, L. and Chang, S., 2025. GraphEPN: A Deep Learning Framework for B-Cell Epitope Prediction Leveraging Graph Neural Networks. Applied Sciences, 15(4), p.2159. Doi: https://doi.org/10.3390/app15042159
23. Evers, A., Malhotra, S. and Sood, V.D., 2025. In Silico Approaches to Deliver Better Antibodies by Design: The Past, the Present, and the Future. In Biopharmaceutical Informatics (pp. 252-280). CRC Press. Doi: https://doi.org/10.1201/9781003300311
24. Li, B., Luo, S., Wang, W., Xu, J., Liu, D., Shameem, M., Mattila, J., Franklin, M.C., Hawkins, P.G. and Atwal, G.S., 2025, December. PROPERMAB: an integrative framework for in silico prediction of antibody developability using machine learning. In mAbs (Vol. 17, No. 1, p. 2474521). Taylor & Francis. Doi: https://doi.org/10.1080/19420862.2025.2474521
25. Meng, F., Zhou, N., Hu, G., Liu, R., Zhang, Y., Jing, M. and Hou, Q., 2024. A comprehensive overview of recent advances in generative models for antibodies. Computational and Structural Biotechnology Journal. Doi: https://doi.org/10.1016/j.csbj.2024.06.016.