Journal of Nursing Research, Patient Safety and Practise
Vol: 05, No. 01, Jan-June 2025
ISSN: 2799-1210, DOI: 10.55529/jnrpsp.51.12.23

Research Paper

Affinity prediction using a neural network model for the m1 protein of group a streptococcus in customized medicine

Chaitra Gajulla Nagaraja[1*], Dr. Daisy Vanitha John[2], Pratham Tarachand Lalwani[3]
[1*,2,3] School of Allied Health Care & Sciences, Jain Deemed-To-Be University, Bangalore, Karnataka, India.

Article Info

Article History:
Received: 08 March 2025
Revised: 17 May 2025
Accepted: 24 May 2025
Published: 09 July 2025

Keywords:
Neural Network
Drug affinity
M1 Protein
GAS infection
Deep Neural Network

ABSTRACT

The M1 protein of Group A Streptococcus (GAS) is a critical virulence factor contributing to a wide range of diseases, from common pharyngitis to life-threatening invasive infections and autoimmune sequelae. Its role in immune evasion, inflammation, and tissue invasion makes it an attractive target for drug and vaccine development. Addressing M1-mediated pathogenesis is crucial to reducing the impact of GAS infections, particularly in vulnerable populations like children. This research study proposes the dry lab approach that leverages computational methods to provide valuable pre-analysis insights to optimise the wet lab experiments. This approach aims to reduce the experimental costs and time by predicting key outcomes in silico, guiding researchers in selecting the most promising candidates for further validation in the wet lab. The affinity of drug-protein interaction is central to developing effective treatments against M1 protein-associated GAS infections. High-affinity drugs can inhibit M1 functions, reduce immune-mediated damage, and ensure specificity and safety. By targeting M1 with high precision, these drugs can significantly mitigate the global burden of GAS diseases. This research study shows that using neural networks to predict drug-protein affinity to know the pros and cons of the proposed method that can significantly accelerate and enhance the development of therapeutics targeting the M1 protein of Group A Streptococcus (GAS).

Corresponding Author:
Chaitra Gajulla Nagaraja
School of Allied Health Care & Sciences, Jain Deemed-To-Be University, Bangalore, Karnataka, India.
Email: gn.chaitra@jainuniversity.ac.in

1. INTRODUCTION

Affinity refers to the strength of interaction between two molecules, such as a drug and its target protein. The affinity of a drug for its target protein is a measure of how tightly and specifically the drug binds to the protein. Traditional in vitro methods for affinity testing require extensive laboratory procedures. These include protein purification, reagent preparation, and repeated experimental runs, which can take days or weeks to complete. Additionally, in vitro experiments are expensive and may be limited in scalability when screening large numbers of drug candidates. In contrast, AI-based neural network models can predict binding affinity in a matter of seconds or minutes once trained.

For the past three decades, binding affinity prediction has been modelled using small datasets [1], [2]. There have been empirical, structure-based methods, knowledge-based methods, and quantitative structure-based methods [3], [4], [5], [6], [7], [8], [9], [10] to predict. But all these methods have drawbacks of low dataset or improper validation methods. A benchmark dataset of 144 protein-protein interactions were also used with un-redundant data of experimental data. However, there is no benchmark dataset with protein ligand affinity prediction, especially for M1 protein of Group a Streptococcus (GAS) infection in children. This study utilizes a neural network prediction model to predict the affinity of ligands for the M1 protein and its variations. This research shows the feature-based model using neural network. However, this paper is a working paper of the research as mentioned above and there is need of more datasets to improve the accuracy of the proposed model.

2. RELATED WORK

The M1 serotype of GAS is highly studied because it infected more than 18 million people, and cause 50,000 deaths as mentioned [11]. Though there is substantial research on GAS infections, it is still unclear what molecular pathways are responsible for their emergence and how to best manage them. Unlike the M1 matrix protein in influenza, this M1 protein is associated with the GAS cell wall and is encoded by the emm1 gene. It plays a pivotal role in the pathogenicity of GAS and its ability to evade the host immune response. The M1 protein is a fibrillar, surface-bound protein [12]. The genetic diversity of emm types allows GAS to adapt to host immune defences, but the M1 type is particularly linked to severe infections due to its ability to interact with host systems and immune components effectively. It facilitates colonization, adhesion, and invasion of host tissues and contributes to the severity of GAS-associated diseases. Especially, in children it affects more due to developing immunity, where they are less equipped to handle immune evasion strategies employed by M1-expressing GAS. This may be acquired due to frequent exposure by close contact in schools and day-care centres, increasing transmission risks.

Targeting the M1 protein of GAS is crucial due to its association with epidemics and severe invasive diseases [13]. Effective intervention could help prevent complications such as rheumatic fever and streptococcal toxic shock syndrome (STSS) [14]. Although GAS remains largely sensitive to penicillin [15], rising resistance to other antibiotics [16] and the absence of a vaccine make M1 protein an important complementary therapeutic target. Modulating M1-induced inflammatory responses may help reduce tissue damage and systemic complications. Given the high morbidity and mortality associated with GAS infections, particularly in developing countries like India, targeting the M1 protein could significantly alleviate the global disease burden.

2.1. Estimating Binding Affinity

Predicting protein-ligand binding affinity using neural networks is undoubtedly a challenging task, requiring precise methodologies to achieve accurate predictions. One crucial factor in this process is the inclusion of 3D atomic coordinates. [17] Discussed various models used for this purpose and their corresponding Root Mean Squared Error (RMSE) scores, highlighting the effectiveness of such approaches. Employing these computational methods helps to significantly reduce the time spent on virtual screening [18], [19]. Over the years, protein-ligand binding affinity prediction has evolved from traditional physics- and chemistry-based wet lab experiments [20] to empirical scoring methods [21]. In recent decades,

machine learning [22] and deep learning approaches [23] have further revolutionized this field, providing more efficient and accurate predictions.

Comprehensive survey of models is explained by Kim et.al [23]. These models are descriptor based scoring functions [24]. Using Convolutional Neural Network (CNN) the first affinity predict model was considering voxels in grid to understand the protein-ligand complexity, but has showed inefficient as it has no much information carried in its voxel corresponding to pharmacophoric properties [25], [26], [27].

Traditionally there have been structural based methods that uses the 3D structure of the both ligand and protein. Molecular docking [28] is the automated method that predicts the binding poses. These poses are based on physics, empirical and knowledge-based method [29], [30]. Deep Learning based applications are available [31]. DEAttentionDTA [32] model utilizes dynamic word embedding and self-attention mechanism and showed enhancements better than GraphscoreDTA, DeepDTAF. It has showed the different datasets to deal with neural network.

2.2. Protein Sequences

Protein sequences are linear chains of amino acids that define the structure and function of a protein [33]. They are encoded by the corresponding gene sequences in the DNA and translated into polypeptides through the processes of transcription and translation. Each amino acid in the sequence is represented by a standard single-letter or three-letter code, with the sequence determining the protein's three-dimensional structure, interactions, and biological activity. Understanding protein sequence is fundamental in drug discovery, as it enables the identification of molecular targets and design of effective therapeutics.

For example: Imatinib [34] is a tyrosine kinase inhibitor used primarily to treat Chronic Myeloid Leukaemia (CML) showed the significance of protein sequence knowledge in drug discovery. Venetoclax [35] is a selective inhibitor of the B-Cell lymphoma 2 that plays an important role in apoptosis. These examples emphasize the importance of protein sequence in drug discovery. Protein sequence is starting point of any drug discovery. However, protein sequence alone is not enough for drug discovery as it provides the important information about the target like affinity. Additional features like protein structure, dynamics, post-translation modifications and cellular interactions are essential for developing effective drugs [36].

2.3. Simplified Molecular Input Line Entry (SMILE) Sequences

This research has adopted a technique of using the SMILES that is used as feature for designing the neural network. SMILES represent chemical structures as linear text strings using atomic symbols and bonding rules. It helps in virtual screening of drug molecules. Helps in predicting affinity to protein targets [37]. These sequences are used in many AI driven drug design [38], QSAR models [39], [40], they predicted improved accuracy better than traditional methods. Protein-ligand interaction predictions [38] using SMILES outperformed the traditional molecular docking techniques in predicting. Deep Neural Network based methods [41] have used SMILE to find the drug-target binding affinity. Hence it is a powerful tool for modern drug discovery, accelerating drug development with AI and computational methods.

2.4. Auto Dock

Autodock [42] is a protein-ligand affinity calculating tool. It uses the physics-based docking simulation. They use 3D structure of the protein, molecular force fields and rigid/flexible docking simulations for predicting binding conformation and binding energy. However, it slows as it is computationally expensive.

2.5. Nucleotide Sequence

In this research nucleotide sequence is also used in the dataset collection to train the NN model. It provides the genetic information that tailors treatment to individual patients based on their genetic makeup. This is crucial in avoiding adverse effect that can be potential risk to individuals due to generalised medicine.

One such example is Abacavir [43] hypersensitivity in HIV treatment experienced severe hypersensitivity reactions. This strongly associated with HLA-B*5701. For this genetic testing allows healthcare providers to identify at-risk individuals and prescribe alternative medications.

3. METHODOLOGY

Proposed model shown in Figure 1 mentions the steps to design the model. The Novel neural network is designed to understand the pattern of protein sequence, gene sequence, SMILE sequence and predict the affinity of the drug on the protein. These features combinations are unique compared to available datasets. Protein sequence patterns play a significant role in predicting binding affinity, particularly for protein-protein interaction and protein-ligand interactions. These patterns help identify key residues involved in binding, which is crucial for applications in drug design, virtual screening, and protein engineering. In Machine learning models using protein sequences can predict binding affinities by leveraging amino acid properties and sequence descriptors. For instance, the ISLAND [44] method employs support vector machines to make sequence-only predictions of binding affinity, demonstrating improved accuracy compared to prior methods. Advances in Machine learning have led to more effective prediction models [45]. Sequence based methods use feature-based techniques such as embedding of amino acids properties and pre-trained representations to predict binding affinities with throughput. In drug discovery, sequence-based drug-affinity (DTA) model predicts binding strength based on protein sequence and SMILES representation of ligands. These methods are computationally efficient and applicable to large datasets. Deep learning models for protein-ligand affinity integrate sequence features and graph neural networks to enhance predictive accuracy. These models capture sequence-based interactions critical for binding prediction [46].

Figure 1. Process of Proposed Model of NN-based Affinity Prediction

4. RESULTS AND DISCUSSION

For the training of the neural network, the gene sequence, protein sequence related to M1 and similar sequences were collected, and the drug or ligands SMILE sequences were used to estimate the affinity. This research is specifically targeted for M1 protein of GAS in children and its variations.
Following 50 protein sequences were used for training: 7JM3_1, 1AA7, 6Z5J, 3MD2, 1EA3, 5V6G, 2Z16, 5V7S, 4PUS, 5V7B, 6I3H, 5V8A, 5CQE, 5WCO, 4K5L, 4ZW5, 8T6H, 8T7P, 8T83, 1PD3, 3Q44, 4K5M, 4K5N, 4K5O, 4K5P, 4ZX3, 8SLO, 3EBG, 4J3B, 4ZQT, 6Z5L, 8SVL, 3PUU, 3Q43, 6EA1, 6EAA, 6EAB, 6EE3, 6EE4, 6EE6, 8EX3, 8EYD, 3EBH, 4ZW6, 4ZW7, 4ZW8, 4ZX4, 4ZX5, 4ZX6, 6EA2. On each protein, few ligands were checked for docking score using Autovina.

Table 1. Training and Validation Metrics (Epochs 1–100)

Epoch	Training Loss	Validation Loss	Training MAE	Validation MAE
1	1.05	1.16	0.83	0.89
2	0.8071	0.8978	0.6257	0.681
3	0.7274	0.8151	0.4619	0.6107

4	0.7	0.79	0.43	0.425
5	0.4472	0.7919	0.4178	0.4201
6	0.4582	0.8091	0.3266	0.427
7	0.478	0.4158	0.3324	0.4413
8	0.5036	0.4489	0.3428	0.3005
9	0.5333	0.4867	0.2667	0.3233
10	0.3162	0.5279	0.283	0.3488
11	0.3515	0.5717	0.3012	0.3763
12	0.3887	0.6175	0.2309	0.2454
13	0.4274	0.6651	0.2519	0.2757
14	0.4673	0.294	0.2738	0.3072
15	0.2582	0.344	0.2066	0.3395
16	0.3	0.395	0.23	0.2125
17	0.3425	0.4468	0.254	0.2462
18	0.3857	0.4993	0.1886	0.2803
19	0.4294	0.5524	0.2135	0.315
20	0.2236	0.606	0.2389	0.1901
21	0.2682	0.24	0.1746	0.2255
22	0.3132	0.2945	0.2006	0.2612
23	0.3585	0.3494	0.2268	0.2972
24	0.4041	0.4045	0.1633	0.1735
25	0.2	0.46	0.19	0.21
26	0.2461	0.5157	0.2169	0.2467
27	0.2925	0.5717	0.154	0.2836
28	0.339	0.2079	0.1812	0.1606
29	0.3857	0.2643	0.2086	0.1978
30	0.1826	0.3208	0.1461	0.2352
31	0.2296	0.3776	0.1737	0.2727
32	0.2768	0.4345	0.2014	0.1503
33	0.3241	0.4915	0.1393	0.188
34	0.3715	0.5486	0.1672	0.2258
35	0.169	0.1859	0.1952	0.2637
36	0.2167	0.2433	0.1333	0.1417
37	0.2644	0.3008	0.1615	0.1797
38	0.3122	0.3584	0.1898	0.2179
39	0.3601	0.4161	0.1281	0.2561
40	0.1581	0.4739	0.1565	0.1344
41	0.2062	0.5318	0.1849	0.1727
42	0.2543	0.1697	0.1234	0.2112
43	0.3025	0.2277	0.152	0.2496
44	0.3508	0.2858	0.1806	0.1281
45	0.1491	0.344	0.1193	0.1667
46	0.1974	0.4022	0.148	0.2053
47	0.2459	0.4605	0.1767	0.244
48	0.2943	0.5188	0.1155	0.1227
49	0.3429	0.1571	0.1443	0.1614
50	0.1414	0.2156	0.1731	0.2002
51	0.19	0.274	0.112	0.239
52	0.2387	0.3325	0.1409	0.1179
53	0.2874	0.3911	0.1699	0.1568

54	0.3361	0.4497	0.1089	0.1957
55	0.1348	0.5083	0.1379	0.2346
56	0.1836	0.147	0.1669	0.1136
57	0.2325	0.2057	0.106	0.1526
58	0.2813	0.2644	0.135	0.1916
59	0.3302	0.3232	0.1642	0.2307
60	0.1291	0.382	0.1033	0.1097
61	0.178	0.4408	0.1324	0.1488
62	0.227	0.4997	0.1616	0.188
63	0.276	0.1386	0.1008	0.2271
64	0.325	0.1975	0.13	0.1062
65	0.124	0.2564	0.1592	0.1454
66	0.1731	0.3154	0.0985	0.1846
67	0.2222	0.3744	0.1277	0.2238
68	0.2713	0.4334	0.157	0.1031
69	0.3204	0.4924	0.0963	0.1423
70	0.1195	0.1315	0.1256	0.1816
71	0.1687	0.1905	0.1549	0.2209
72	0.2179	0.2496	0.0943	0.1002
73	0.267	0.3087	0.1236	0.1395
74	0.3162	0.3679	0.153	0.1788
75	0.1155	0.427	0.0924	0.2181
76	0.1647	0.4862	0.1218	0.0975
77	0.214	0.1254	0.1512	0.1369
78	0.2632	0.1846	0.0906	0.1762
79	0.3125	0.2438	0.12	0.2156
80	0.1118	0.303	0.1494	0.095
81	0.1611	0.3622	0.0889	0.1344
82	0.2104	0.4215	0.1183	0.1739
83	0.2598	0.4807	0.1478	0.2133
84	0.3091	0.12	0.0873	0.0927
85	0.1085	0.1793	0.1168	0.1322
86	0.1578	0.2386	0.1463	0.1717
87	0.2072	0.2979	0.0858	0.2111
88	0.2566	0.3573	0.1153	0.0906
89	0.306	0.4166	0.1448	0.1301
90	0.1054	0.476	0.0843	0.1696
91	0.1548	0.1153	0.1139	0.2091
92	0.2043	0.1747	0.1434	0.0886
93	0.2537	0.2341	0.083	0.1281
94	0.3031	0.2935	0.1125	0.1677
95	0.1026	0.3529	0.1421	0.2072
96	0.1521	0.4123	0.0816	0.0868
97	0.2015	0.4717	0.1112	0.1263
98	0.251	0.1111	0.1408	0.1659
99	0.3005	0.1706	0.0804	0.2054
100	0.1	0.23	0.11	0.085

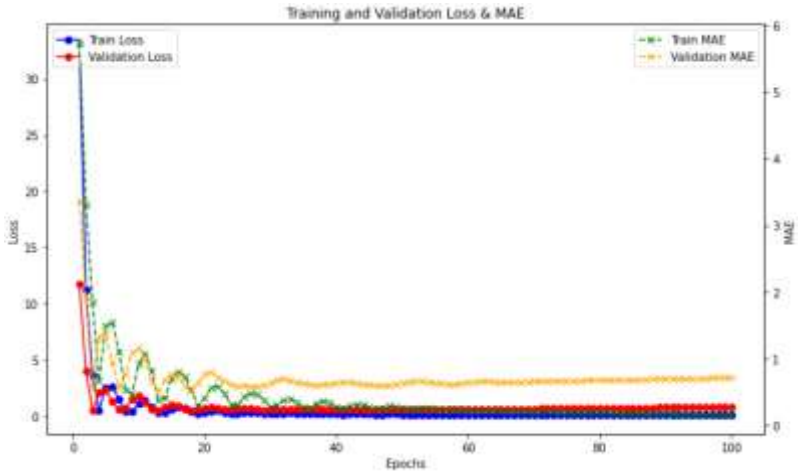

Figure 2. Training and Validation Loss & MAE over Epochs Analysis

The approach involves several important steps in data pre-processing, model architecture, and evaluation for predicting docking scores using nucleotide and SMILES sequences. Data pre-processing includes collecting nucleotide sequences, SMILES, gene sequences, and docking scores, followed by cleaning the dataset by removing rows with missing values. Each character in the nucleotide and SMILES sequences is converted into its ASCII value, and sequences are padded to ensure consistent lengths. The dataset is then split into training and testing subsets (80% and 20%, respectively). The proposed model architecture features separate input layers for nucleotide and SMILES sequences, followed by embedding layers to convert sequences into dense vector representations. Dense layers independently process each input, and the outputs are concatenated, flattened, and passed through additional dense layers. The final output layer consists of a single neuron with a linear activation to predict the docking score. During model training, the Adam optimizer and Mean Squared Error (MSE) loss function are used, with performance monitored across 100 epochs. Evaluation metrics include Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2), with results showing moderate accuracy but a negative R^2 value of -2.5704, suggesting the model is underfitting. This underperformance likely arises from the limited sample size that is 100 samples, and further improvement can be achieved by increasing the dataset size and refining the model architecture or pre-processing steps.

The graph in Figure 2, Table 1 represents the training and validation loss along with the Mean Absolute Error (MAE) over 100 epochs. The training and validation loss, along with the Mean Absolute Error (MAE), show a sharp decline in the initial epochs, indicating that the model is learning quickly. After approximately 10-20 epochs, both loss and MAE stabilize, suggesting that the model has reached a steady state. The training and validation curves remain closely aligned, which indicates minimal overfitting and good generalization. The validation MAE fluctuates slightly but remains relatively stable, signifying consistent performance. Overall, the model demonstrates effective learning, though further fine-tuning of hyperparameters or regularization techniques may help in optimizing performance further

5. CONCLUSION

This study presents a novel approach for predicting docking scores of protein-ligand interactions using a neural network trained on ligand SMILES representations, protein sequences, and nucleotide sequences, along with predefined docking scores obtained from Autodock Vina. By leveraging machine learning, this approach aims to accelerate the virtual screening process, reducing computational costs and enabling faster identification of potential drug candidates. Additionally, the study provides a structured framework for dataset generation, which can be expanded and refined in future research to enhance model performance and generalizability.

The proposed methodology offers significant potential for improving prediction accuracy by capturing complex molecular interactions more effectively than traditional docking techniques. It also paves the way for scalable and automated drug discovery pipelines, facilitating large-scale screening efforts

with minimal human intervention. Furthermore, the integration of such predictive models into pharmaceutical workflows can reduce experimental failures and optimize the selection of promising drug candidates for further validation.

This research serves as a foundation for future studies to explore hybrid approaches that combine deep learning with traditional docking methods, potentially leading to more interpretable and accurate results. The insights gained from this work can also be extended to various biomedical applications, such as personalized medicine and protein engineering. Ultimately, this study contributes to the advancement of computational drug discovery by offering an efficient and scalable alternative to conventional virtual screening methods.

The approach improves prediction accuracy by capturing complex molecular interactions more effectively than traditional methods, with potential refinements such as incorporating physicochemical properties and 3D structural features. It supports scalability and automation, making it suitable for large-scale drug discovery pipelines with minimal human intervention. The framework has cross-disciplinary applications, including personalized medicine, protein engineering, and computational biology. It also lays the foundation for hybrid approaches that combine deep learning with traditional docking methods for more precise results. Ultimately, this method reduces experimental failures by prioritizing the most promising drug candidates, enhancing the efficiency and effectiveness of drug discovery efforts.

Acknowledgments

We extend our heartfelt gratitude to my ex-colleague, Dr. Purabi Sarkar, and Mr. Akash S (Biopol Biosciences Pvt. Ltd) for their initial motivation and valuable insights in shaping the focus of our research.

Funding Information

The authors sincerely acknowledge Jain Deemed-to-be University for the financial support provided under the Minor Research Project Scheme (Reference No. JU/MRP/SAHS/1/2022), which made this research possible.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Chaitra Gajulla Nagaraja	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Daisy Vanitha John									✓	✓		✓		
Pratham Tarachand Lalwani						✓	✓	✓						

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

Informed Consent

There is no individual data included, hence, there is no need to obtain informed consent from anyone.

Ethical Approval

The research is not related to animal use so there is no need of relevant national regulations and institutional policies for the care and use of animals.

Conflict of Interest Statement

Authors declare that there is no conflict of Interest.

Data Availability

Protein sequences and SMILE sequence samples used for training are available in the Protein Bank.

REFERENCES

[1] N. Horton and M. Lewis, 'Calculation of the free energy of association for protein complexes', Protein Sci., vol. 1, no. 1, pp. 169-181, Jan. 1992. doi.org/10.1002/pro.5560010117

[2] P. L. Kastritis and A. M. J. J. Bonvin, 'On the binding affinity of macromolecular interactions: daring to ask why proteins interact', J. R. Soc. Interface, vol. 10, no. 79, p. 20120835, Feb. 2013. doi.org/10.1098/rsif.2012.0835

[3] J. Audie and S. Scarlata, 'A novel empirical free energy function that explains and predicts protein-protein binding affinities', Biophys. Chem., vol. 129, no. 2-3, pp. 198-211, Sep. 2007. doi.org/10.1016/j.bpc.2007.05.021

[4] L. Jiang, Y. Gao, F. Mao, Z. Liu, and L. Lai, 'Potential of mean force for protein-protein interaction studies', Proteins, vol. 46, no. 2, pp. 190-196, Feb. 2002. doi.org/10.1002/prot.10031

[5] X. H. Ma, C. X. Wang, C. H. Li, and W. Z. Chen, 'A fast empirical approach to binding free energy calculations based on protein interface information', Protein Eng. Des. Sel., vol. 15, no. 8, pp. 677-681, Aug. 2002. doi.org/10.1093/protein/15.8.677

[6] Y. Su, A. Zhou, X. Xia, W. Li, and Z. Sun, 'Quantitative prediction of protein-protein binding affinity with a potential of mean force considering volume correction', Protein Sci., vol. 18, no. 12, pp. 2550-2558, Dec. 2009. doi.org/10.1002/pro.257

[7] T. Vreven, H. Hwang, B. G. Pierce, and Z. Weng, 'Prediction of protein-protein binding free energies', Protein Sci., vol. 21, no. 3, pp. 396-404, Mar. 2012. doi.org/10.1002/pro.2027

[8] C. Zhang, S. Liu, Q. Zhu, and Y. Zhou, 'A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes', J. Med. Chem., vol. 48, no. 7, pp. 2325-2335, Apr. 2005. doi.org/10.1021/jm049314d

[9] F. Tian, Y. Lv, and L. Yang, 'Structure-based prediction of protein-protein binding affinity with consideration of allosteric effect', Amino Acids, vol. 43, no. 2, pp. 531-543, Aug. 2012. doi.org/10.1007/s00726-011-1101-1

[10] P. Zhou, C. Wang, F. Tian, Y. Ren, C. Yang, and J. Huang, 'Biomacromolecular quantitative structure-activity relationship (BioQSAR): a proof-of-concept study on the modeling, prediction and interpretation of protein-protein binding affinity', J. Comput. Aided Mol. Des., vol. 27, no. 1, pp. 67-78, Jan. 2013. doi.org/10.1007/s10822-012-9625-3

[11] A. Thacharodi et al., 'the burden of group A Streptococcus (GAS) infections: The challenge continues in the twenty-first century', iScience, vol. 28, no. 1, p. 111677, Jan. 2025. doi.org/10.1016/j.isci.2024.111677

[12] M. Matsumoto et al., 'Variation in M protein production among Streptococcus pyogenes strains according to emm genotype', Microbiol. Immunol. vol. 55, no. 6, pp. 379-387, Jun. 2011. doi.org/10.1111/j.1348-0421.2011.00329.x

[13] P. J. B. Davies et al., 'Increase of severe pulmonary infections in adults caused by M1UK Streptococcus pyogenes, central Scotland, UK', Emerg. Infect. Dis., vol. 29, no. 8, pp. 1638-1642, Aug. 2023. doi.org/10.3201/eid2908.230569

[14] S. Zhang et al., 'Streptococcal M1 protein-provoked CXC chemokine formation, neutrophil recruitment and lung damage are regulated by Rho-kinase signaling', J. Innate Immun., vol. 4, no. 4, pp. 399-408, Mar. 2012. doi.org/10.1159/000336182

[15] D. Yu, D. Guo, Y. Zheng, and Y. Yang, 'A review of penicillin binding protein and group A Streptococcus with reduced-β-lactam susceptibility', Front. Cell. Infect. Microbiol., vol. 13, p. 1117160, Mar. 2023. doi.org/10.3389/fcimb.2023.1117160

[16] Institute of Medicine (US) Forum on Microbial Threats, Forum on Microbial Threats. Washington, D.C., DC: National Academies Press, 2009.

[17] M. Volkov et al., 'On the frustration to predict binding affinities from protein-ligand structures with deep neural networks', J. Med. Chem., vol. 65, no. 11, pp. 7946-7958, Jun. 2022. doi.org/10.1021/acs.jmedchem.2c00487

[18] D. L. Mobley and M. K. Gilson, 'Predicting binding free energies: Frontiers and benchmarks', Annu. Rev. Biophys., vol. 46, no. 1, pp. 531-558, May 2017. doi.org/10.1146/annurev-biophys-070816-

033654

[19] J. Lyu et al., 'Ultra-large library docking for discovering new chemotypes', Nature, vol. 566, no. 7743, pp. 224-229, Feb. 2019. doi.org/10.1038/s41586-019-0917-9

[20] C. Gorgulla et al., 'An open-source drug discovery platform enables ultra-large virtual screens', Nature, vol. 580, no. 7805, pp. 663-668, Apr. 2020. doi.org/10.1038/s41586-020-2117-z

[21] P. Kollman, 'Free energy calculations: Applications to chemical and biochemical phenomena', Chem. Rev., vol. 93, no. 7, pp. 2395-2417, Nov. 1993. doi.org/10.1021/cr00023a004

[22] I. A. Guedes, F. S. S. Pereira, and L. E. Dardenne, 'Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges', Front. Pharmacol., vol. 9, p. 1089, Sep. 2018. doi.org/10.3389/fphar.2018.01089

[23] Q. U. Ain, A. Aleksandrova, F. D. Roessler, and P. J. Ballester, 'Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening', Wiley Interdiscip. Rev. Comput. Mol. Sci., vol. 5, no. 6, pp. 405-424, Nov. 2015. doi.org/10.1002/wcms.1225

[24] K. Kato, T. Honma, and K. Fukuzawa, 'Intermolecular interaction among Remdesivir, RNA and RNA-dependent RNA polymerase of SARS-CoV-2 analyzed by fragment molecular orbital calculation', J. Mol. Graph. Model., vol. 100, no. 107695, p. 107695, Nov. 2020. doi.org/10.1016/j.jmgm.2020.107695

[25] S. Chakrabarty and A. Warshel, 'Capturing the energetics of water insertion in biological systems: the water flooding approach', Proteins, vol. 81, no. 1, pp. 93-106, Jan. 2013. doi.org/10.1002/prot.24165

[26] R. Wang, Y. Lu, and S. Wang, 'Comparative evaluation of 11 scoring functions for molecular docking', J. Med. Chem., vol. 46, no. 12, pp. 2287-2303, Jun. 2003. doi.org/10.1021/jm0203783

[27] Z. Hu, G. Lu, and D. Xu, 'FVC: A new framework towards deep video compression in feature space', in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021. doi.org/10.1109/CVPR46437.2021.00155

[28] C. Li, Q. Lv, N. Li, Y. Wang, D. Sun, and Y. Qiao, 'A novel deep framework for dynamic malware detection based on API sequence intrinsic features', Comput. Secur., vol. 116, no. 102686, p. 102686, May 2022. doi.org/10.1016/j.cose.2022.102686

[29] Y. Chen, Y. Kalantidis, J. Li, S. Yan, and J. Feng, 'A2-Nets: Double Attention Networks', Adv. Neural Inf. Process. Syst, vol. 31, 2018.

[30] H. Gao, Z. Meng, X. Liang, and L. Ji, "A high-order multi-view hashing approach for cross-modal retrieval," Pattern Recognit., vol. 100, Art. no. 107121, 2020,

[31] D. C. Huang, 'Supervised Attention-Based Fusion Network for RGB-D Salient Object Detection', IEEE Trans. Image Process, vol. 29, pp. 1449-1460, 2020. doi.org/10.1109/TIP.2019.2959253

[32] J. Hu, L. Shen, and G. Sun, 'Squeeze-and-Excitation Networks', in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018. doi.org/10.1109/CVPR.2018.00745

[33] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, 'Grad-CAM: Visual explanations from deep networks via gradient-based localization', in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017. doi.org/10.1109/ICCV.2017.74

[34] Y. Chen et al., 'Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution', in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019. doi.org/10.1109/ICCV.2019.00353

[35] X. Wang, R. Girshick, A. Gupta, and K. He, 'Non-local Neural Networks', in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018. doi.org/10.1109/CVPR.2018.00813

[36] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, 'Feature pyramid networks for object detection', in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017. doi.org/10.1109/CVPR.2017.106

[37] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep residual learning for image recognition', in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. doi.org/10.1109/CVPR.2016.90

[38] M. Zheng et al., 'Weakly supervised contrastive learning', in 2021 IEEE/CVF International Conference

on Computer Vision (ICCV), Montreal, QC, Canada, 2021. doi.org/10.1109/ICCV48922.2021.00989
[39] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, 'Densely connected convolutional networks', in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017. doi.org/10.1109/CVPR.2017.243
[40] M. Tan and Q. Le, 'EfficientNet: Rethinking model scaling for convolutional Neural Networks', in Proceedings of the 36th International Conference on Machine Learning, 09--15 Jun 2019, vol. 97, pp. 6105-6114.
[41] A. Dosovitskiy et al., 'An image is worth 16x16 words: Transformers for image recognition at scale', 02-Oct-2020.
[42] A. Vaswani, 'Attention Is All You Need', Adv. Neural Inf. Process. Syst, vol. 30, 2017.
[43] M. Peters et al., 'Deep contextualized word representations', in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), New Orleans, Louisiana, 2018. doi.org/10.18653/v1/N18-1202
[44] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, in Proceedings of the 2019 Conference of the North, Minneapolis, Minnesota, 2019.
[45] Z. Zhang, Z. Yang, H. Liu, Y. Deng, and D. Zhang, 'BINDNets: Binding Free Energy Estimation Using Deep Learning with Interaction Fingerprints and Distance Map', J. Chem. Inf. Model, vol. 61, no. 11, pp. 5706-5716, 2021.
[46] S. M. Moosavi, M. J. Morency, and G. Bouchard, 'Scoring Functions for Protein-Ligand Interactions Using Deep Learning', J. Chem. Inf. Model, vol. 60, no. 12, pp. 5629-5640, 2020.

How to Cite: Chaitra Gajulla Nagaraja, Dr. Daisy Vanitha John, Pratham Tarachand Lalwani. (2025). Affinity prediction using a neural network model for the m1 protein of group a streptococcus in customized medicine. Journal of Nursing Research, Patient Safety and Practise (JNRPSP), 5(1), 12-23. https://doi.org/10.55529/jnrpsp.51.12.23

BIOGRAPHIES OF AUTHORS

	**Chaitra Gajulla Nagaraja** is a Computer Science Engineer with a Master's in Computer Science and Engineering from Christ University. She currently serves as an Assistant Professor at Jain Deemed-To-Be University, specializing in AI, computer vision, and healthcare IT. With international research experience in visual SLAM and a strong publication record, she actively contributes to interdisciplinary innovation. Her interests include drug discovery, large language models, and human-computer interaction. She can be contacted at Email: gn.chaitra@jainuniversity.ac.in
	**Dr. Daisy Vanitha John** is an accomplished microbiologist and immunologist with a Ph.D. in Biomedical Sciences from the Tuberculosis Research Centre, India. She currently serves as an Assistant Professor at the Jain School of Allied Healthcare and Sciences, Bangalore, and has held previous research and faculty positions in India, Malaysia, and the USA. Her expertise spans molecular microbiology, immunopathogenesis of infectious diseases, and advanced diagnostic methods including metagenomics and flow cytometry. Dr. John has received multiple national and international research grants and has over 25 peer-reviewed publications, in addition to book chapters and invited talks. She can be contacted at Email: daisyvanitha.j@jainuniversity.ac.in

Pratham Tarachand Lalwani is an undergraduate student pursuing a BSc in Virology and Immunology at Jain Deemed-To-Be University. He has hands-on experience in molecular biology techniques such as PCR, ELISA, and Immunofluorescence, gained during his internship at Aster Clinical Labs. Pratham has participated in workshops at premier institutions like IISc Bangalore and actively engages in awareness programs. His interests lie in research, innovation, and contributing to advancements in healthcare and diagnostics. Email: plalwani589@gmail.com