ISSN : 2319-2607 (Print); 2583-0856 (Online) IISUniv.J.Sc.Tech. Vol.12(1), 70-79 (2023)
REVIEW ARTICLE

Unveiling the Potential of Deep Learning in Protein Function Prediction: A Comprehensive Review

Anushri Vijay[1]*, Neha Tiwari[1], Amita Sharma[1], Prashanth Suravajhala[2]
[1]IIS (deemed to be University), Jaipur, India.
[2]Systems Genomics, Amrita School of Biotechnology Amrita University, Amritapuri, Kerala.

Abstract

Deep learning has revolutionized protein function prediction in the areas of gene expression, protein-protein interaction, and G-coupled receptor analysis. This paper explores the architectures, strategies, benchmark datasets, and evaluation metrics used in deep learning models for predicting gene expression, protein-protein interactions, and GPCR functions. The models leverage large-scale genomic data, sequence information, and evolutionary features to achieve remarkable accuracy in predicting protein function. The advancements in deep learning have provided valuable insights into biological systems, aiding drug discovery and therapeutic interventions. Further progress in deep learning methods holds great potential for enhancing our understanding of protein function and its role in biological processes.

Keywords: Artificial Intelligence(AI), Convolutional Neural Networks(CNN), Deep Learning(DL), Deep Neural Networks, Gene Expression, G-Protein Coupled Receptors, Nuclear Receptors, Protein Function Prediction(PFP), Protein-Protein Interaction, Recurrent Neural Networks(RNN).

Introduction

Proteins are vital components that govern various biological processes in living organisms. Understanding their functions is crucial for uncovering the intricacies of cellular mechanisms, facilitating drug discovery, and enabling targeted therapeutic interventions. In recent years, deep learning has emerged as a transformative approach for predicting protein function, enabling researchers to explore complex biological systems at a deeper level. Research focused on protein function prediction, specifically in the areas of gene expression, protein-protein interaction, and G-coupled receptor (GPCR) analysis, is of utmost importance due to the fundamental roles these processes play in biological systems.

The understanding of gene expression patterns is key to unravelling cellular mechanisms, leading to novel insights for drug discovery and therapeutic interventions (Dedrick, 2007). Protein-protein interactions are central to orchestrating complex cellular processes and signalling cascades, making their prediction critical for comprehending biological networks (Soleymani *et al.,* 2022). GPCRs, acting as crucial mediators of cellular communication, represent important targets for pharmacological interventions, and accurately predicting their functions and interactions can greatly expedite drug development (Salon, *et al.,*2011).

The necessity for research in these areas is further underscored by the vast amount of genomic data available and the advent of deep learning architectures. Deep learning has demonstrated remarkable accuracy in deciphering gene expression profiles, identifying interacting protein pairs, and effectively classifying GPCRs and predicting their ligands. The potential of deep learning in these domains offers a transformative approach to gain profound insights into complex biological systems and accelerate advancements in protein function prediction. Consequently, this research holds the potential to significantly enhance our understanding of biological systems and pave the way for novel targeted therapeutic interventions.

Although various approaches exist for predicting protein function, this paper specifically focuses on three key areas: gene expression data, protein interaction networks, and nuclear/GPCRs, which have proven to be effective in protein function prediction. Gene expression data aids in predicting protein function by identifying

*Corresponding Author Email: 12anushrivijay@gmail.com

co-expressed genes associated with specific functions. Protein interaction networks enable the prediction of protein function by identifying interacting partners of known functionally relevant proteins. Nuclear/GPCRs contribute to protein function prediction by identifying proteins that bind to specific nuclear receptors or GPCRs. The paper aims to provide a comprehensive review of cutting-edge deep learning methodologies, datasets, and challenges encountered in each of these areas, showcasing the remarkable progress achieved through the power of deep learning. Furthermore, it outlines potential future directions in this rapidly evolving field. The advancements in protein function prediction hold immense promise for advancing our understanding of biological systems and translating this knowledge into innovative therapeutic interventions.

In the following sections, we will explore the architectures, strategies, and results of deep learning models applied in gene expression prediction, protein-protein interaction analysis, and G-coupled receptor studies. By combining computational approaches with biological insights, we hope to pave the way for further breakthroughs in protein function prediction and ultimately contribute to our understanding of the intricate mechanisms governing life processes.

Artificial Intelligence Techniques in Protein Function Prediction

Artificial Intelligence (AI) techniques, particularly those rooted in deep learning, have significantly advanced protein function prediction (PFP) methodologies. In this section, we delve into the studies focusing on the essence of AI's role in enhancing predictions within gene expression data, protein interaction networks, and nuclear/GPCRs. Furthermore, we shed light on the popular evaluation metrics pivotal in assessing the performance of these deep learning-based models.

One-D CNN (1-D CNN) – One-Dimensional Convolutional Neural Networks (1-D CNNs) are widely utilized in various applications such as personalized biomedical data categorization, early diagnosis, structural health monitoring, anomaly detection in power electronics, and identification of electrical motor failures. 1-D CNNs, which consist of convolutional and sub-sampling layers, have achieved notable performance levels. These networks offer the advantage of being compact, easy to set up, and capable of performing real-time computations at a low cost, making them highly suitable for practical applications (Kiranyaz *et al.,* 2021).

Feed Forward Multilayer Deep Neural Network- The Feed Forward Multilayer Deep Neural Network operates by multiplying input values with corresponding weights, summing them, and comparing the total against a predefined threshold. If the sum exceeds the threshold, the output is typically 1; otherwise, it is typically -1. This network architecture uses back-propagation to update weights, modifying each hidden layer based on the output values of the final layer. This iterative process allows the network to learn and improve its performance (Hakala *et al.,* 2022).

Multilayer Graph Convolutional Network (GCN) - Graph Convolutional Networks (GCNs) analyse-neighbouring nodes to capture their characteristics. While Convolutional Neural Networks (CNNs) are designed for regular, structured data, GCNs are a generalized version that can handle irregular and non-Euclidean structured data. Unlike CNNs, GCNs accommodate variable node connections and nodes without a specific order (Zangari *et al.,* 2021).

Bi-directional Recurrent Neural Network (Bi-RNN)- Recurrent Neural Networks (RNNs) process inputs sequentially, but they lack the ability to consider future inputs when contextualizing the present input. Bi-directional RNNs (Bi-RNNs) overcome this limitation by duplicating the RNN processing chain, allowing inputs to be evaluated in both forward and reverse temporal order. This enables Bi-RNNs to incorporate future context into the analysis (Huang *et al.,* 2021).

Residual Neural Network (ResNet) - The ResNet technique enhances the accuracy and performance of Deep Neural Networks (DNNs) by adding extra layers. The concept behind this approach is that as the network deepens, the layers can capture more intricate features over time. For example, in image recognition, the initial layers may learn to detect edges, followed by layers that recognize textures, and ultimately layers that identify objects (Suh *et al.,* 2021).

Deep Conditional Random Field- Conditional Random Field (CRF) is a powerful statistical modeling tool used for pattern recognition, particularly in text sequence classification. CRF inference enables precise boundary delineation and detailed segmentation, enhancing the accuracy of pixel-level label predictions. It serves as a valuable solution to address the limitations of CNNs in pixel-level labeling tasks (Arnab *et al.,* 2018).

Maxout Neural Network- Max-out layer is a neural network layer where the activation function selects the maximum value among the inputs. It provides a flexible

nonlinearity that can approximate any activation function by choosing from a set of linear pieces. Max-out networks have been employed for acoustic modeling in speech recognition tasks, showing promising results in various conditions (Wan *et al.*, 2019).

Fig. 1 provides an overview of deep learning (DL) in three categories. The first category encompasses convolutional deep neural networks, which utilize distinct neural network types and are effective in tasks like image classification and object detection. The second category consists of discriminant architectures that focus on predicting labels or numbers. Finally, the third category comprises generic architectures capable of training with limited data, where both input and output are known. These categories demonstrate the diverse applications and capabilities of DL in various domains.

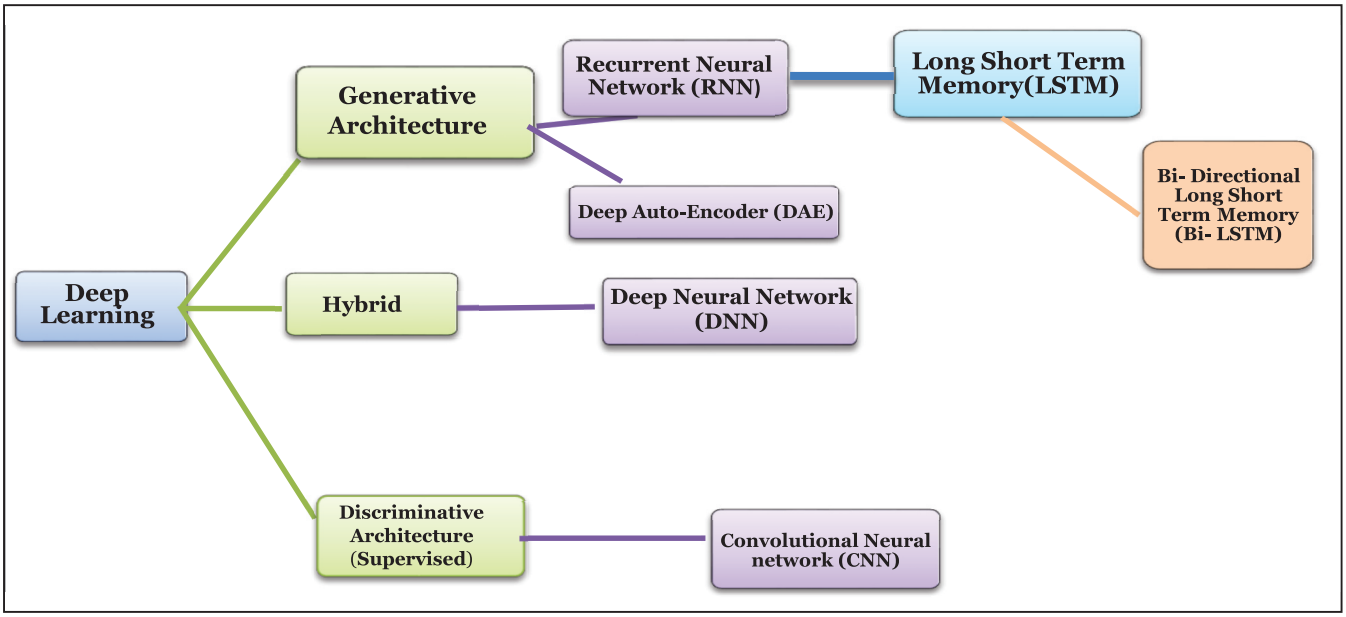

Fig.1. An overview of Deep Learning Technique

Evaluation Metrics for PFP based on Deep Learning

Evaluation metrics are crucial for assessing the performance and effectiveness of protein function prediction models based on deep learning. This section provides a brief overview of commonly used evaluation metrics in the field, highlighting their relevance in assessing the accuracy and reliability of predictions.In this study, various evaluation metrics were employed, including accuracy, sensitivity, specificity, precision, recall, receiver operator characteristic (ROC), area under the curve for precision-recall (AUC-PR), average precision (AP), and F1-score. Additionally, the ranking loss statistic, originating from information retrieval, was utilized to train models for ranking items (You *et al.*, 2021). Coverage, on the other hand, is a measure commonly used in unit testing to assess how many lines of code and execution pathways are covered by at least one test case.

The formulas for some of these evaluation metrics are as follows:

- Accuracy: (TP + TN) / (TP + TN + FP + FN)
- Sensitivity (True Positive Rate): TP / (TP + FN)
- Specificity (True Negative Rate): TN / (TN + FP)
- Precision: TP / (TP + FP)
- Recall (Sensitivity): TP / (TP + FN)
- F1-score: 2 * (Precision * Recall) / (Precision + Recall)

Note: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

These evaluation metrics play a crucial role in assessing the performance of deep learning-based protein function prediction models. The selection of appropriate metrics depends on the nature of the task, dataset characteristics, and specific goals of the prediction problem.

Methodology

Literature search and database

We conducted an extensive literature search to investigate the Protein Function Prediction Using Deep Learning. A combination of keywords "Deep Learning", "Protein Function Prediction","Gene Expression Data", "Protein Interaction Network", "Nuclear/G-protein

Coupled receptor (GPCR)" were entered into the Google Scholar; PubMed/MEDLINE; Scopus; Web of Science; Elsevier and IEEE Xplore. The search was limited to publications in the English language. The purpose of this systematic review was to gather original and review articles that provide insights into Protein Function Prediction parameters using Deep Learning.

Inclusion and Exclusion criteria

For this systematic review, we applied inclusion and exclusion criteria to select relevant articles. We prioritized highly reviewed articles and abstracts that focused on protein function prediction using deep learning. Our inclusion criteria comprise: (a) studies dedicated to protein function prediction using deep learning; (b) data suitable for prediction of protein function; (c) studies investigating protein function prediction using gene expression data, the Protein Interaction Network, and the Nuclear/G-protein Coupled Receptor (GPCR) through a deep learning approach; (d) preference for recent or high-quality publications when multiple sources covered the same or overlapping data. Our exclusion criteria are studies exclusively focusing on protein function and structure. Fig. 2 illustrates Systematic Review process that was used to analyse protein function prediction using deep learning.

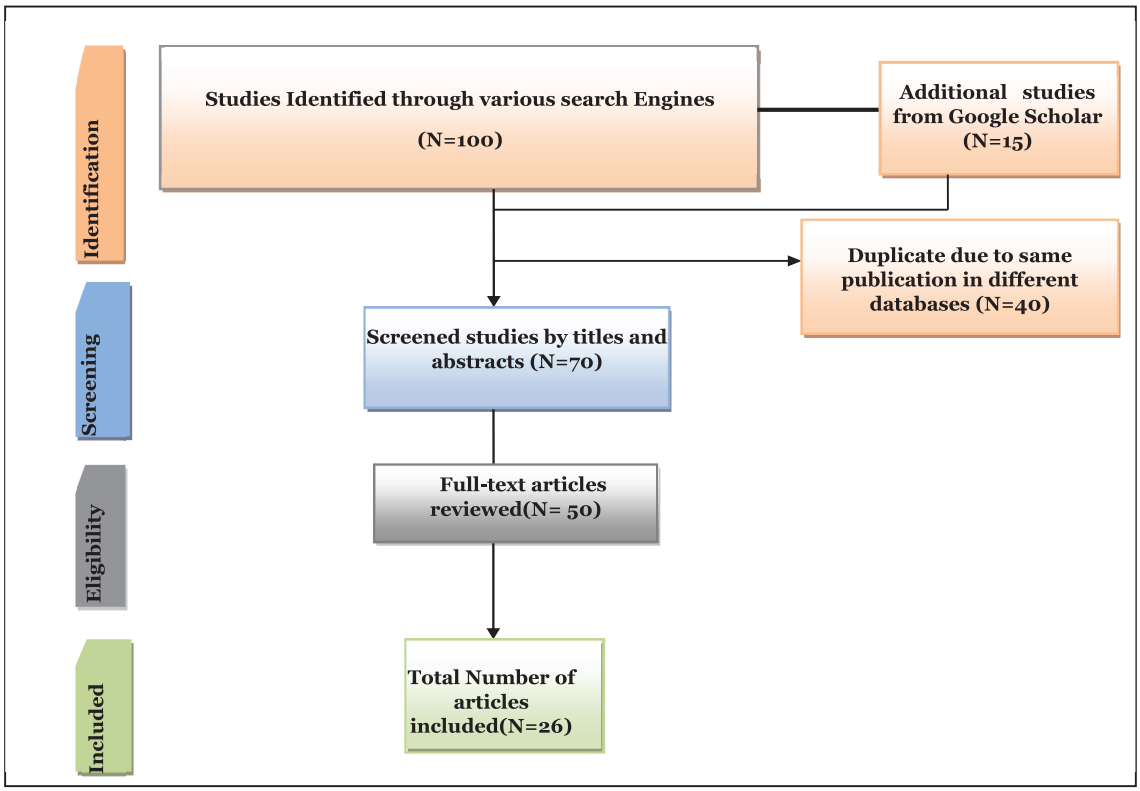

Fig.2. Flowchart for the Systematic Review of protein function prediction methods using DL approaches.

AI Technique for Predicting PF Based on Sequence and Structure

Artificial intelligence (AI) techniques have been successfully employed to predict protein function (PF) based on sequence and structure information. In gene expression prediction, deep learning models, such as convolutional neural networks (CNNs), have been utilized to analyze DNA and RNA sequences and capture relevant features for accurate expression level classification (Sapoval *et al.*, 2022). For protein-protein interaction (PPI) prediction, AI techniques such as graph neural networks (GNNs) have been effective in modeling the complex structural information of interacting proteins, enabling the identification of interacting protein pairs(Zhou *et al.*, 2022).Additionally, in G-coupled receptor (GPCR) analysis, deep learning approaches have been employed to predict GPCR functions and ligands by integrating sequence data, evolutionary information, and physicochemical properties (Yadav *et al.*, 2022). These AI techniques have demonstrated their efficacy in predicting PF in each of the three areas: gene expression, PPI, and GPCR analysis.

By studying gene expression data, protein interaction networks, and nuclear/GPCRs individually, we aim to

provide a comprehensive understanding of the predictive models, techniques, and challenges associated with each domain. Through this detailed analysis, we strive to contribute to the advancement of protein function prediction and its practical applications in various biological and biomedical contexts. We believe that these three approaches can be used to improve the accuracy and robustness of protein function prediction models. We also believe that these approaches can be used to predict the functions of proteins that are not well-studied or that are difficult to predict using other methods.

Gene Expression Data

Gene expression data plays a vital role in understanding the functional characteristics of genes and their involvement in various biological processes. In the context of protein function prediction, the analysis of gene expression data provides valuable insights into the transcriptional activity and regulatory mechanisms underlying protein functions. This section focuses on the utilization of gene expression data for protein function prediction, highlighting relevant studies and their contributions to the field.

Table 1. Prediction of Gene Expression Data using DL

S.No.	Study	AI Technique	Accuracy (%)	Features	Dataset	Limitations	Improvements
1.	Majji *et al.*, 2023	Deep Recurrent Neural Network-based chronological Horse Herd Political Optimization	Prediction Error (PE) = 45.6% minimal Root Mean Square Error (RMSE)=46.7%	88%	Gene expression profiles, Protein sequence	Gene expression and subcellular localization datasets	Limited to protein subcellular localization
2.	Alharbi and Vakanski, 2023	Recent neural network architectures—such as graph and transformer networks	-	gene expression analysis, covering feature engineering techniques	RNA-Seq methods	graph and transformer networks have not been used in ML methods of Cancer data classification	Multimodal fusion boost classification can be improved
3.	Martiny *et al.* 2021	Deep Neural Network	88%	Gene expression profiles, Protein sequence	Gene expression and subcellular localization datasets	Limited to protein subcellular localization	Integration of protein sequence information
4.	Bardak *et al.*, 2021	Deep Neural Network	88%	Gene expression profiles, Protein sequence	Gene expression and protein-protein interaction datasets	Limited to protein-protein interactions	Integration of additional features
5.	Aggarwal and Hasija, 2022	1D CNN + DNN	-	combining with PPI data; gene expression data	sequence data, genomic expression, 3D structures, and data on protein-protein interaction	-	Integration of gene expression profile with current SOTA(state-of-the-art) protein sequence representations

Dash(-) means not available.

From the table1, it is evident that several studies have explored the use of deep learning algorithms for predicting gene expression data in the context of protein function prediction. Different deep learning techniques, such as Deep Neural Networks (DNNs) and Recurrent Neural Networks (RNNs), have been applied in these studies. Understanding the biological mechanisms behind the predictions is crucial for gaining insights into protein function. Future improvements could focus on developing more interpretable models or integrating domain knowledge to enhance biological interpretability. Overall, while deep learning has shown promise in predicting gene expression data for protein function, there is a need for further exploration and advancement. Addressing the limitations and incorporating improvements suggested in the table 1 can lead to more accurate and biologically meaningful predictions, facilitating a deeper understanding of protein function and its implications in biological processes.

Protein Interaction Network

Protein interaction networks provide valuable insights into the complex web of interactions between proteins, offering a wealth of information for predicting protein function. Deep learning approaches have been applied to leverage protein interaction network data for accurate protein function prediction. This section focuses on the utilization of protein interaction networks in protein function prediction, highlighting relevant studies and their contributions to the field. Table 2 provides a concise overview of the approaches, accuracy percentages, features used, datasets employed, and limitations of each study. It highlights the potential for future improvements, such as exploring transfer learning, multi-view learning, and attention mechanisms, to enhance accuracy and address the limitations identified.

Across the studies from Table 2, it is evident that deep learning techniques, such as Ensemble of Deep Autoencoders, Maxout neural networks and DNNs, have proven effective in leveraging protein interaction network data for protein function prediction. Integration of additional features, such as protein sequence and structure information, has contributed to improved accuracy. However, challenges including data sparsity, class imbalance, computational complexity, and interpretability remain, and addressing these challenges is crucial for further advancements in the field.

Nuclear/G-protein Coupled receptor(GPCR)

Nuclear/G-protein coupled receptors (GPCRs) play a crucial role in cellular signalling pathways and have attracted significant attention in research. However, a substantial number of GPCRs lack experimental structures, necessitating the use of computational methods to predict their structures and interactions with ligands. Deep learning approaches have emerged as valuable tools in predicting protein function, including for nuclear/GPCRs, by harnessing the wealth of information present in their sequences and structural properties. By leveraging these computational methods, researchers aim to enhance our understanding of cellular signalling, accelerate the discovery of novel therapeutic targets, and expedite the development of potential drug candidates for GPCRs.

The table 3 presents two key studies that employ deep learning approaches for predicting protein function specifically for nuclear/G-protein coupled receptors (GPCRs). These studies focus on different aspects of nuclear/GPCRs and leverage various deep learning algorithms and datasets to achieve their objectives.The analysis highlights some common limitations in the field. These include the availability of high-quality training data, interpretability of deep learning models, and the need to consider broader functional aspects beyond ligand binding prediction. Improvements can be made by integrating additional structural information or protein dynamics, exploring transfer learning techniques to leverage related tasks, and considering other features or data types to capture a comprehensive understanding of nuclear/GPCR functions.

In conclusion, while the studies demonstrate the potential of deep learning in predicting nuclear/GPCRs for protein function, there is a need to address limitations and further enhance the accuracy and biological relevance of predictions. By incorporating improvements and overcoming challenges, deep learning models can contribute to a deeper understanding of nuclear/GPCR functions and their role in cellular signalling pathways.

Conclusion

In this paper, we have delved into the domain of protein function prediction using deep learning techniques, with a specific focus on gene expression data, protein-protein interaction networks, and nuclear/G-protein coupled receptors (GPCRs). We reviewed key studies published until 2023, showcasing the application of deep learning algorithms in each area. From the analysis of the literature, it is evident that deep learning holds immense promise in predicting protein function based on gene expression data. Various models, such as DeepBind and DeepLoc, have demonstrated their effectiveness in

Table 2. Prediction of Protein Interaction Network using DL

S. No.	References	AI technique and Model Name	Performance	Features	Dataset	Limitations	Improvement
1.	Hu *et al.*, 2022	combination of CNNs and RNNs, Multilayer Perceptron (MLP)	-	Primary sequences	Primary sequences, domain composition, secondary structures and 3D structures	-	-
2.	Jha *et al.*, 2022	combines graph neural network (GNN) and language model (LM)	Accuracy = 98.13%, F-score = 98.73%, MCC = 95.20%	Nodes of protein graphs(molecular protein graph (amino acids/residues as nodes) from a PDB file containing structural information)	Amino acids/residues as nodes	-	Multi-scale representation learningand intrinsic-extrinsic convolution and pooling for learning on 3D protein structures
3.	Jamasb *et al.*, 2021	Ensemble of Deep Autoencoders and AutoPPI model	Accuracy and AUC = 0.97	Protein sequence information, evolutionary conservation scores, physicochemical properties, structural characteristics, and residue interactions.	Experimentally validated protein-protein interaction sites with corresponding non-interacting residues.	There is need for large-scale annotated datasets, challenges in capturing complex protein-protein interactions, and potential biases in the training data.	Advancements in model architectures, optimization techniques, and interpretability methods can further enhance the accuracy and interpretability of predicted PPI sites.
4.	Barot *et al.*, 2021	Maxout neural networks and NetQuilt model	-	Homology-informed network similarity, which leverages evolutionary relationships among proteins across multiple species	Protein sequences and their corresponding functional annotations across multiple species.	Availability and coverage of functional annotations for all proteins, potential biases in the training data, and the challenges in capturing complex relationships amon g proteins in different species.	Incorporating additional biological data, such as gene expression profiles or protein structure information, to enhance the accuracy and coverage of predictions.
5.	Zhang *et al.*, 2018	Deep neural networks (DNNs)	Average accuracy=94.34%, Precision 95.62%, Recall = 92.96%, Specificity=95.74%, MCC = 88.73%, F1 = 94.27% and AUC = 98.24%	Protein sequence information, structural characteristics, physicochemical properties, evolutionary conservation scores, and functional annotations as well as high-quality non-interacting pairs as a feature to distinguish true interactions from false positives.	Known protein-protein interactions along with a set of high-quality non-interacting protein pairs.	Availability of reliable negative examples, potential biases in the selection of non-interacting pairs, and challenges in capturing the full complexity of protein-protein interactions	Incorporation of additional features, such as subcellular localization information or gene expression profiles, t o improve the predictive performance. Furthermore, advancements in machine learning algorithms, feature engineering techniques, and integration of multi -omics data can further enhance the accuracy and reliability of predicted PPIs.

Table 3. Prediction of Nuclear/G-protein Coupled receptor (GPCR) using DL

S. No.	References	AI technique and model name	Prediction	Performance	Dataset and Features	Limitations	Improvements
1	Mollaei and Barati Farimani, 2023	Decision Tree, Random Forest, and XGBoost	predicting the most probable transition pathway between activation states of GPCRs based on structural features of the receptors	Accuracy- a)XGBoost=97.37% b) Random Forest=94.48% c)Decision Tree- 90.45%	PDB structures of the GPCRs	-	-
2	Jiang *et al.,* 2022	XGBoost and TrAdaBoost algorithm	GPCR	Accuracy=0.26% higher than similar methods	Experimentally validated GPCR interactions along with corresponding non-interacting pairs, Amino acid sequence	Availability of comprehensive and high-quality interaction data, potential biases in the training dataset, and challenges in capturing the full complexity of GPCR interactions, particularly in the presence of various signalling pathways and ligand-specific effects	Incorporating additional data sources, such as ligand information, signalling pathway data, or protein-protein interaction networks, to improve the accuracy and coverage of predictions.
3	Yadav *et al.,* 2022	3D CNN, GNN, and XGBoost models	GPCR	Accuracy- 95.85%	PDB structures for GPCRs	-	-
4	Wu *et al.,* 2017	Deep Conditional random fields and dCRF-TM model	GPCR	Accuracy=79%	Uniprot database, DB1, DB2, and DB3 datasets. DB1 derived from TMPDB, DB2 derived from PDBTM, DB3 derived from TOPDB	Availability of high-quality and diverse training data, challenges in accurately capturing the complex transmembrane topology patterns, and potential biases in the training dataset.	Incorporating additional information, such as evolutionary profiles, secondary structure predictions, or inter-residue contacts, to improve the accuracy and reliability of predictions.

capturing sequence patterns and predicting subcellular localization. Similarly, in the realm of protein-protein interaction prediction, DeepPPI has shown potential in deciphering complex protein interactions from gene expression profiles.Additionally, in the context of nuclear/GPCRs, models like dCRF-TM modelhave shed light on predicting ligand binding and overall protein function. However, challenges in data availability, interpretability, and capturing broader functional aspects remain areas for improvement.

Future Scope

As we move forward, several exciting avenues offer opportunities for advancing the field of protein function prediction using deep learning:

Multi-modal data integration: Integrating diverse omics data, such as epigenetic modifications, histone modifications, and chromatin accessibility, with gene expression data can provide a holistic view of gene regulation and enhance prediction accuracy.

Interpretability enhancement: Developing more interpretable deep learning models or incorporating domain knowledge can unravel the biological mechanisms underlying predictions and enable more biologically meaningful insights.

Transfer learning and domain adaptation: Exploring transfer learning techniques, where pre-trained models are fine-tuned on specific datasets, can leverage knowledge from related tasks and improve prediction performance, especially in cases with limited data availability.

Structural information incorporation: Integration of structural features and protein dynamics information can enhance the accuracy and relevance of predictions for nuclear/GPCRs and other membrane proteins.

Utilizing explainable deep learning methods for protein function prediction to enhance accuracy and provide interpretability of predicted protein functions.

Benchmarking and standardized evaluation: Establishing benchmark datasets and standardized evaluation metrics will enable fair comparison of different deep learning models and encourage the development of more effective algorithms.

By addressing these future directions, the field of protein function prediction can make substantial strides in deciphering the complex relationships between gene expression, protein interactions, and nuclear/GPCR functions. This deeper understanding will have broad implications in biology, medicine, and drug discovery, opening up new avenues for research and applications in the years to come.

Conflict of Interest

The Authors declare no conflict of interest.

References

Aggarwal, D., Hasija, Y. 2022. A Review of Deep Learning Techniques for Protein Function Prediction 2021. 2nd International Conference for Emerging Technology (INCET). https://doi.org/10.48550/arxiv.2211.09705.

Alharbi, F., Vakanski, A. 2023. Machine Learning Methods for Cancer Classification Using Gene Expression Data: A Review. *Bioengineering (Basel, Switzerland)*, *10*(2), 173. https://doi.org/10.3390/bioengineering10020173.

Arnab, A., Zheng, S., Jayasumana, S., Romera-Paredes, B., Larsson, M., Kirillov, A., Savchynskyy, B., Rother, C., Kahl, F., Torr, P. H. S. 2018. Conditional Random Fields Meet Deep Neural Networks for Semantic Segmentation: Combining Probabilistic Graphical Models with Deep Learning for Structured Prediction. IEEE Signal Process. Mag. 35(1), 37–52. https://doi.org/10.1109/msp.2017.2762355.

Bardak, B. Tan, M. 2021. DeepGREP: A deep convolutional neural network for predicting gene-regulating effects of small molecules. 2021. *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Melbourne, Australia, 2021, pp. 1-8, doi: 10.1109/CIBCB49929.2021.9562920.

Barot, M., Gligorijević, V., Cho, K., Bonneau, R. 2021. Net Quilt: deep multispecies network-based protein function prediction using homology-informed network similarity. Bioinformatics. 37(16), 2414–2422.

Dedrick, R.L. 2007. Understanding Gene Expression Patterns in Immune-Mediated Disorders. J. Immunotoxicol. 4(3), 201–207. https://doi.org/10.1080/15476910701385562.

Hakala, K., Kaewphan, S., Bjorne, J., Mehryary, F., Moen, H., Tolvanen, M., Ginter, F. 2022. Neural Network and Random Forest Models in Protein Function Prediction. IEEE/ACM Trans. Comput. Biol. Bioinform.19(3), 1772–1781.

Huang, G., Shen, Q., Zhang, G., Wang, P., Yu, Z. G. 2021. LSTMCNNsucc: A Bidirectional LSTM and CNN-Based Deep Learning Method for Predicting Lysine Succinylation Sites. BioMed Res. Int. 2021, 9923112. https://doi.org/10.1155/2021/9923112

Hu, X., Feng, C., Ling, T., Chen, M. 2022. Deep learning frameworks for protein-protein interaction prediction. Comput. Struct. Biotechnol. J. 20, 3223-3233 https://doi.org/10.1016/j.csbj.2022.06.025

Jamasb, A. R., Day, B., Cangea, C., Liò, P., Blundell, T. L. 2021. Deep Learning for Protein–Protein Interaction Site Prediction. Methods Mol. Biol. 2631, 263–288.

Jha, K., Saha, S., Singh, H. 2022. Prediction of protein–protein interaction using graph neural networks. Sci. Rep. 12(1). 8360. https://doi.org/10.1038/s41598-022-12201-9

Jiang, T., Chen, Y., Guan, S., Hu, Z., Lu, W., Fu, Q., Ding, Y., Li, H. and Wu, H. ,2022. G protein-coupled receptor interaction prediction based on deep transfer learning. IEEE/ACM Trans. Comput. Biol. Bioinform. 19 (6), 3126-3134. doi:https://doi.org/10.1109/tcbb.2021.3128172.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D. J. 2021. 1D convolutional neural networks and applications: A survey. Mech Syst Signal Process. 151, 107398. https://doi.org/10.1016/j.ymssp.2020.107398.

Majji, R., Maram, B., Rajeswari, R. 2023. Chronological horse herd optimization-based gene selection with deep learning towards survival prediction using PAN-Cancer gene-expression data. Biomed. Signal Process Control. 84, 104696. https://doi.org/10.1016/j.bspc.2023.104696

Martiny, H. M., Armenteros, J. J. A., Johansen, A. R., Salomon, J., Nielsen, H.2021. Deep protein representations enable recombinant protein expression prediction. Comput. Biol. Chem. 95, 107596.

Mollaei, P., Barati Farimani, A. 2023. Activity Map and Transition Pathways of G Protein-Coupled Receptor revealed by Machine Learning. J. Chem. Inf. Model. 63(8), 2296–2304. https://doi.org/10.1021/acs.jcim.3c00032

Salon, J.A., Lodowski, D.T., Palczewski, K. 2011. The Significance of G Protein-Coupled Receptor Crystallography for Drug Discovery. Pharmacol. Rev. 63(4), 901–937. doi:https://doi.org/10.1124/pr.110.003350.

Sapoval, N., Aghazadeh, A., Nute, M.G., Antunes, D.A., Balaji, A., Baraniuk, R., Barberan, C.J., Dannenfelser, R., Dun, C., Edrisi, M., Elworth, R.A.L., Kille, B., Kyrillidis, A., Nakhleh, L., Wolfe, C.R., Yan, Z., Yao, V., Treangen, T.J. 2022. Current progress and open challenges for applying deep learning across the biosciences. Nat. Commun. 13(1). 1728. doi:https://doi.org/10.1038/s41467-022-29268-7.

Soleymani, F., Paquet, E., Viktor, H., Michalowski, W., Spinello, D. 2022. Protein–Protein Interaction Prediction with Deep Learning: A Comprehensive Review. Comput. Struct. Biotechnol. J. 20, 5316-5341.https://doi.org/10.1016/j.csbj.2022.08.070.

Suh, D., Lee, J. W., Choi, S., Lee, Y. 2021. Recent Applications of Deep Learning Methods on Evolution- and Contact-Based Protein Structure Prediction. Int. J. Mol. Sci. 22(11), 6032. https://doi.org/10.3390/ijms22116032.

Wan, C., Cozzetto, D., Fa, R., Jones, D.T. 2019. Using deep maxout neural networks to improve the accuracy of function prediction from protein interaction networks. PLoS One. 14(7), e0209958. https://doi.org/10.1371/journal.pone.0209958

Wu, H., Wang, K., Lu, L., Xue, Y., Lyu, Q., Jiang, M. 2017 . Deep Conditional Random Field Approach to Transmembrane Topology Prediction and Application to GPCR Three-Dimensional Structure Modeling. IEEE/ACM Trans. Comput. Biol. Bioinform. 14(5), 1106–1114.

Yadav, P., Mollaei, P., Cao, Z., Wang, Y., Barati Farimani, A. 2022. Prediction of GPCR activity using machine learning. Comput. Struct. Biotechnol. J. 20, 2564–2573. https://doi.org/10.1016/j.csbj.2022.05.016.

You, R., Yao, S., Mamitsuka, H., Zhu, S. 2021. Deep Graph GO: graph neural network for large-scale, multispecies protein function prediction. Bioinformatics. 37(Suppl_1), i262–i267. https://doi.org/10.1093/bioinformatics/btab270.

Zangari, L., Interdonato, R., Calió, A., Tagarelli, A. 2021. Graph convolutional and attention models for entity classification in multilayer networks. Appl. Netw. Sci. 6(1), 87. https://doi.org/10.1007/s41109-021-00420-4.

Zhang, L., Yu, G., Guo, M., Wang, J. 2018. Predicting protein-protein interactions using high-quality non-interacting pairs. BMC Bioinformatics. 19(Suppl 19), 525. https://doi.org/10.1186/s12859-018-2525-3

Zhou, H., Wang, W., Jin, J., Zheng, Z., Zhou, B. 2022. Graph Neural Network for Protein–Protein Interaction Prediction: A Comparative Study. Molecules. 27(18), 6135. https://doi.org/10.3390/molecules27186135.