EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

3D Convolutional Neural Networks for Predicting Protein Structure for Improved Drug Recommendation

Pokkuluri Kiran Sree[1,*], SSSN Usha Devi N[2]

[1]Dept of C.S.E, Shri Vishnu Engineering College for Women(A), Bhimavaram, India
[2]Dept of C.S.E, University College of Engineering, JNTU Kakinada

Abstract

INTRODUCTION: Protein structure prediction is critical for recommendation personalized medicine and drug discovery. This paper introduces a robust approach using 3D Convolution Neural Networks (3D CNN's) to improve the accuracy of the structure of protein structure thus contributing for the drug recommendation system.
OBJECTIVES: In contrast to conventional techniques, 3D CNNs are able to identify complicated folding patterns and comprehend the subtle interactions between amino acids because they are able to capture spatial dependencies inside protein structures.
METHODS: Data sets are collected from Protein Data Bank, including experimental protein structures and the drugs that interact with them, are used to train the model. With the efficient processing of three-dimensional data, the 3D CNNs exhibit enhanced capability in identifying minute structural details that are crucial for drug binding. This drug recommendation system novel method makes it easier to find potential drugs that interact well with particular protein structures.
RESULTS: The performance of the proposed classifier is compared with the existing baseline methods with various parameters accuracy, precision, recall, F1 score, mean squared error (MSE) and area under the receiver operating characteristic curve (AUC-ROC).
CONCLUSION: Deep learning and 3D structural insights work together to create a new generation of tailored and focused therapeutic interventions by speeding up the drug development process and improving the accuracy of pharmacological recommendations.

Keywords: Deep Learning, CNN (Convolution Neural Networks), Protein Structure

Received on 23 December 2023, accepted on 02 April 2024, published on 08 April 2024

doi: 10.4108/eetpht.10.5685

1. Introduction

Understanding the three-dimensional structures of proteins is crucial in the fields of drug development and personalized medicine. Proteins display a wide range of shapes that are essential to their functions [1]. Analysing these complex structures is essential to creating medications that target particular proteins with precision and efficacy, improving therapeutic results. Conventional approaches to protein structure prediction frequently struggle to capture the intricate spatial interactions that these biomolecules possess. Presenting 3D Convolutional Neural Networks (3D CNNs), a ground-breaking method that could revolutionize our capacity to predict protein structures with previously unheard-of precision. We have developed a unique method to identify the protein structure which will be helpful in drug recommendation named as 3D-CNN [2].

Drug discovery requires a fundamental understanding of protein structures through certain interactions with other molecules, including medications, proteins carry out their biological roles [3]. A protein's binding sites are determined by the three-dimensional arrangement of its atoms, and a drug's potency and selectivity are determined by its capacity to bind to these sites. Therefore, developing medications that precisely target and control the activity of certain proteins implicated in illnesses requires accurate protein structure prediction [4].

*Corresponding author. Email: drkiransree@gmail.com

In terms of literal dimensions, 3D CNNs expand the range of protein structure prediction. These neural networks are capable of identifying tiny structural motifs that are crucial for drug interactions and of discerning complex folding patterns by processing three-dimensional spatial information. The network's layers are able to capture both local and global aspects of protein structures since they are trained with hierarchical representations. Because 3D CNNs are inherently capable of comprehending the three-dimensional arrangement of atoms, they are especially suitable for more accurate protein structure prediction [5].

Personalized medicine could be greatly enhanced by the use of 3D CNNs into drug recommendation systems. Through precise prediction of protein structures, these models aid in the identification of potential drugs that demonstrate advantageous interactions with particular proteins [6]. This strategy goes beyond conventional drug discovery techniques, providing a more focused and effective way to suggest medications that are suited to certain protein structures and, as a result, patient profiles [7].

3D CNNs are trained using a variety of datasets that include experimental protein structures and the pharmacological interactions associated with them. The model's performance can be improved, particularly in situations with a lack of labelled data [8], by using transfer learning approaches, in which the model is pre-trained on a sizable dataset before being fine-tuned on particular tasks. The model's robustness and prevention of overfitting are ensured by hyper parameter tweaking and cross-validation, giving confidence in the model's capacity to generalize to previously undiscovered protein structures [9].

Protein structure prediction greatly benefits from machine learning, which uses algorithms to examine large amounts of biological data. Models trained on existing protein structures can be used to accurately predict new ones using supervised learning [10]. Deep learning techniques improve the extraction and representation of features, capturing intricate correlations found in amino acid sequences. This helps in the prediction of protein folding patterns, tertiary structures, and possible roles. Structural biology is being revolutionized by the integration of many data sources and sophisticated algorithms, which improve accuracy and speed up drug discovery and biological mechanism comprehension [11].

Machine learning uses prediction models to examine large biological and chemical datasets in order to recommend drugs. Based on established drug-target relationships, these models forecast possible therapeutic possibilities for particular illnesses [12]. Deep learning algorithms, for example, improve pattern identification and reveal intricate connections between medicinal effects and molecular structures. The integration of several data sources, such as chemical characteristics and genomes, enhances the precision of proposed new molecules. Pharmaceutical research [13] advances as a result of machine learning's ability to speed up drug discovery, simplify the identification of potential candidates, lower expenses, and produce novel treatments more quickly.

2. Literature Survey

Various machine learning techniques have been used in numerous research to solve problems related to the prediction of protein structure, function, and interactions. Deep learning and other supervised learning techniques have become well-known for their capacity to extract complex patterns from enormous biological datasets. Attention mechanisms, recurrent neural networks (RNNs)[14], and convolutional neural networks (CNNs) are frequently used by researchers to improve feature representation and identify long-range dependencies in amino acid sequences.

Machine learning is used in research to predict protein structures from homologous structures. Protein three-dimensional (3D) structure [15] can be predicted with amazing accuracy by deep learning algorithms like Alpha Fold. Utilizing machine learning approaches, residue-residue connections can be predicted, which facilitates the inference of spatial correlations between amino acids and tertiary structure prediction. In order to better understand protein folding, classification models are used to predict secondary structural components (such as beta sheets and alpha helices) from amino acid sequences. By connecting sequencing data with established functional annotations, machine learning is used to forecast the functions of proteins. Enzyme function prediction and Gene Ontology (GO) term prediction are two techniques [16].

Drug discovery is aided by the development of models that forecast how medications will interact with protein targets. These models help identify promising candidates and provide insight into their modes of action. Protein-protein interactions that are important for physiological activities [17] can be identified by using machine learning to predict whether two proteins will interact or not. To help models learn meaningful representations, complicated links within amino acid sequences or protein structures are captured using word embeddings and graph-based representations [18].

Protein prediction models are often assessed using a variety of benchmark datasets, including CASP [19] (Critical Assessment of Structure Prediction). Evaluation metrics offer a quantitative assessment of the prediction accuracy of models and include precision, recall, F1-score, and structural similarity indices. Challenges arise from imbalanced datasets, in which some protein groups are underrepresented. Research is still ongoing to address these imbalances and create reliable models for a variety of protein families [20].

A thorough review of the literature on machine learning-based drug recommendation systems indicates an expanding corpus of research targeted at improving treatment planning, drug discovery, and personalized medicine. Research investigates how to construct comprehensive patient profiles by integrating genomes, patient demographics, and electronic health records (EHR). Personalized pharmacological regimens are recommended by machine learning models using this data, which takes into account the unique characteristics of each patient. Proactive medication suggestions for prevention or early intervention are made possible by the use of predictive modelling tools,

which are used to forecast disease progression and identify patients at higher risk [21].

Drug-drug interactions are represented by graph-based models as a network that may have negative effects or positive interactions. These models support physicians in minimizing medication regimens and avoiding hazardous combinations. By encoding links between medications, illnesses, and biological pathways in knowledge graphs, drug interactions can be better understood and recommendations can be made with more accuracy. In order to forecast missing values and suggest medications based on comparable patient profiles, collaborative filtering approaches, such as matrix factorization algorithms, evaluate patient-drug interaction matrices. Deep learning architectures are used by neural collaborative filtering algorithms to identify complex patterns in patient data and offer tailored medication recommendations [22].

A state-of-the-art method for combining deep learning with computational modelling in protein structure prediction is the combination of 3D Convolutional Neural Networks (3DCNNs) and Cellular Automata (CA)[23]. 3DCNNs are able to capture complex three-dimensional correlations between amino acid sequences by extracting spatial characteristics from protein structures. Simultaneously, dynamic processes [24] of protein folding are simulated by Cellular Automata on a grid, where each cell represents a particular residue or area. A synergistic model that takes advantage of both the dynamic modelling capabilities of CA and the spatial awareness of 3DCNNs is made possible by the integration of these two components. While the CA captures the developing states of cells based on local interactions, the 3DCNN learns hierarchical characteristics. This hybrid model provides a thorough comprehension of the dynamics of protein folding, potentially improving the precision and interpretability

3. Design of the Proposed Method

The process of creating 3D Convolutional Neural Networks (3DCNNs) for medicine recommendation and protein structure prediction requires customizing architectures to process three-dimensional spatial input. In this hybrid model, convolutional and pooling layers are used by 3D Convolutional Neural Networks (3DCNNs) to process protein or drug structures and extract spatial characteristics. Cellular Automata (CA) simulate dynamic states based on local interactions by operating on a grid concurrently.

The 3DCNN and CA outputs are combined in the integration layer to capture temporal dynamics as well as spatial nuances. Predictions for drug interactions or protein structures are provided by task-specific output layers. In end-to-end training, backpropagation is used to optimize all of the model's parameters. This synergistic architecture improves medication recommendations and protein structure predictions by combining dynamic simulation with spatial feature extraction. To increase the accuracy and interpretability of the model, research is still ongoing as shown in figure 1.

3.1 Data Representation

Protein structures are represented as three-dimensional grids with each voxel representing a distinct spatial location inside the protein. Draw molecular structures as three-dimensional grids where each voxel represents a molecular characteristic, like the type of bond, atom, or pharmacophore. Within the 3D grid, encode data on evolutionary profiles, solvent accessibility, and amino acid kinds.

3.2 Architectural Design

In order to capture spatial correlations in the molecular structure and enable the network to identify important properties for drug interaction, we have included 3D convolutional layers. The network can learn hierarchical features by influencing the receptive field through choices in stride and kernel size. To down sample spatial dimensions while keeping pertinent characteristics, use 3D pooling layers.

3.3. Feature Fusion and Interaction Prediction

We have incorporated residual connections, skip connections to help train deeper designs by facilitating information transfer between levels. Combine the output of convolutional layers with multi-modal data, such as evolutionary profiles or secondary structure annotations. We have used the molecular structure-based output layer design to anticipate binding locations on target proteins. Create output layers for binary interaction labels or affinity scores in order to anticipate drug-target interactions as shown in figure 1.

3.4. Protein and Drug Recommendation System

To improve model generalization, rotate or translate protein structures into the training set. To improve performance, apply previously learned models to comparable tasks or domains. To represent dynamic processes on a grid and capture the temporal dimensions of drug-protein interactions we are using cellular automata. We have created a hybrid model by combining the features that the 3DCNN extracted with the states that were derived from the cellular automata. In order to jointly optimize the 3DCNN and Cellular Automata components, train the combined model from beginning to end.

.

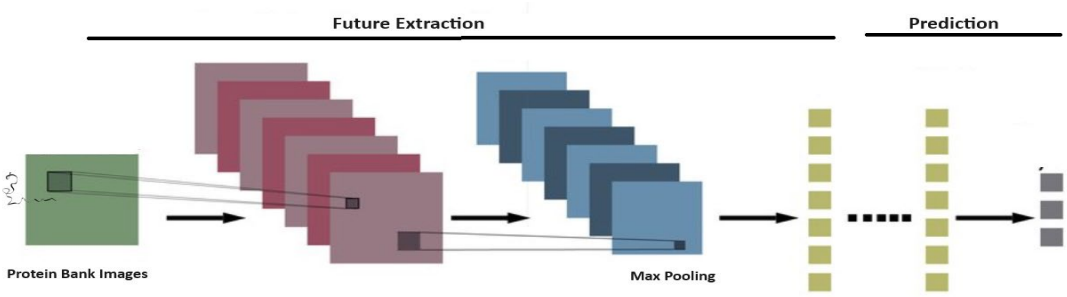

Figure 1: Design of 3D-CNN for Protein Structure Prediction

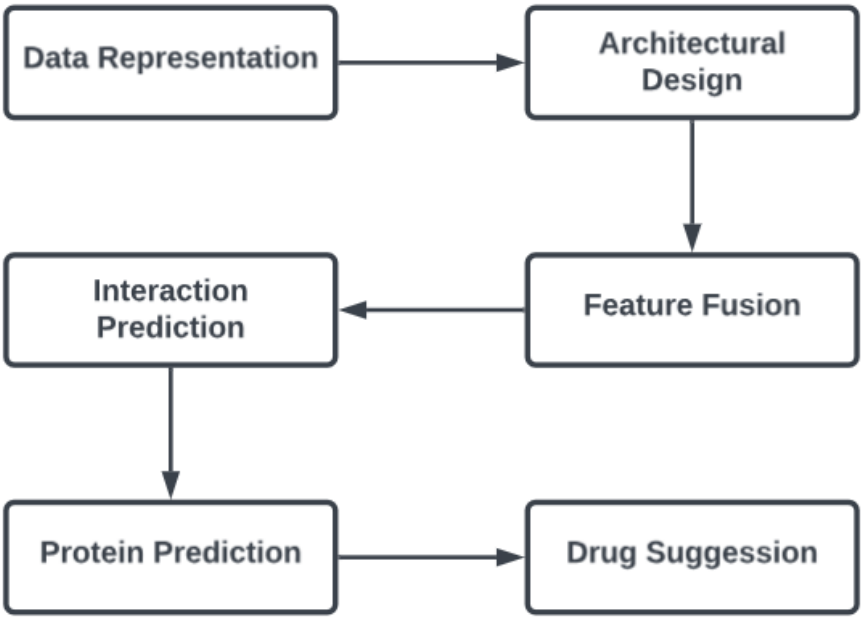

Figure 2: Implementation of 3D-CNN for Protein Structure & Drug Suggestion Prediction

4. Experimental Results and Comparisons

We have collected 7,79,274 datasets from Protein bank to predict the secondary and quaternary structure of the protein. We have taken 70% of the datasets for the training, 15% for validation and 15% for testing as shown in the table 1. We have tested our innovative work with various standard protein banks ie EGF1, TRYPSINSER, RNASE, EFHRD1, 16MHC and these are compared with the baseline methods SVM, LSTM with our developed approach 3DCNN as shown in table 2 and figure 3. Evaluation criteria are essential for evaluating the predictive ability of computational approaches in protein structure prediction. Precision is a metric that expresses the fraction of accurately predicted positive instances and quantifies the accuracy of positive predictions. Sensitivity, also known as recall, evaluates how well the approach captures all of the real positive instances in the dataset. By balancing these two measurements, the F1 Score—a harmonic mean of precision and recall—offers a thorough assessment of prediction accuracy. The Matthews Correlation Coefficient (MCC) provides a balanced assessment, especially in imbalanced datasets, by taking into account true positives, true negatives, false positives, and false negatives. Together, these measures provide information about the accuracy, sensitivity, and general efficacy of protein structure prediction techniques, which helps scientists choose and improve computational models for precise three-dimensional protein structure predictions.

Table 1: Bifurcation of the Datasets

Training Set	Validation Set	Test Set
70%	15%	15%
4,18,784	1,80,245	1,80,245

Table 2: Comparison of the Base Line methods with 3DCNN on various Protein Bank Sites

Site	Method	Precision	Recall	F1 Score	MCC
EGF1	3DCNN	0.906	0.970	0.975	0.915
	SVM	0.837	0.789	0.911	0.834
	LSTM	0.898	0.853	0.884	0.871
TRYPSINSER	3DCNN	0.902	0.974	0.999	0.904
	SVM	0.885	0.712	0.866	0.825
	LSTM	0.816	0.813	0.882	0.889
RNASE	3DCNN	0.989	0.969	0.901	0.912
	SVM	0.904	0.898	0.858	0.898
	LSTM	0.829	0.689	0.873	0.807
EFHRD1	3DCNN	0.992	0.947	0.905	0.959
	SVM	0.878	0.826	0.879	0.822
	LSTM	0.801	0.671	0.867	0.883

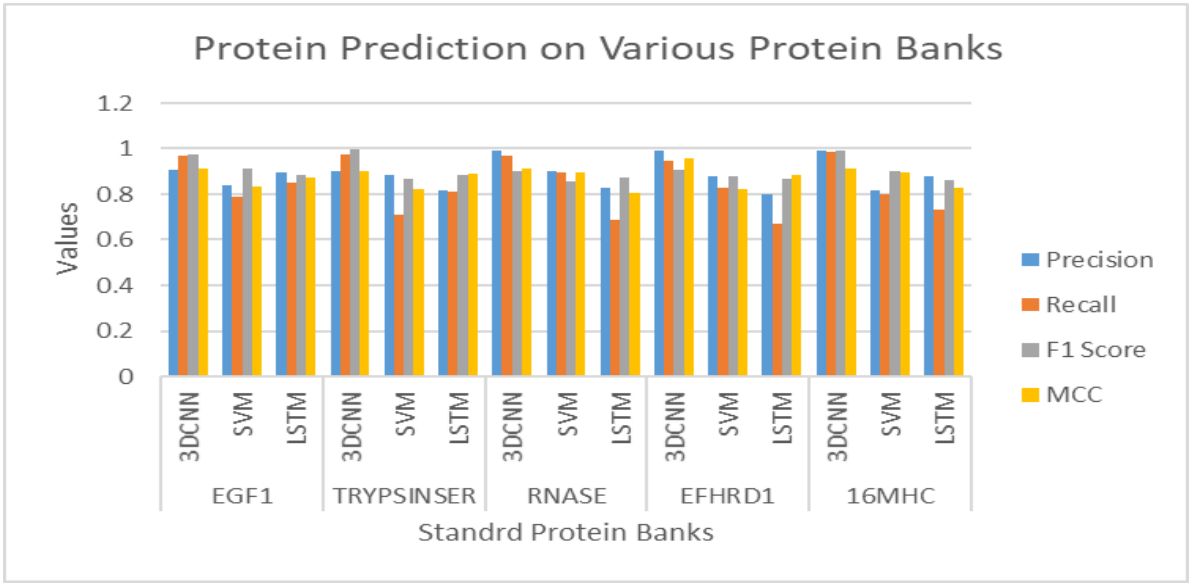

Figure 3: Robustness of the 3DCNN on various Protein Banks

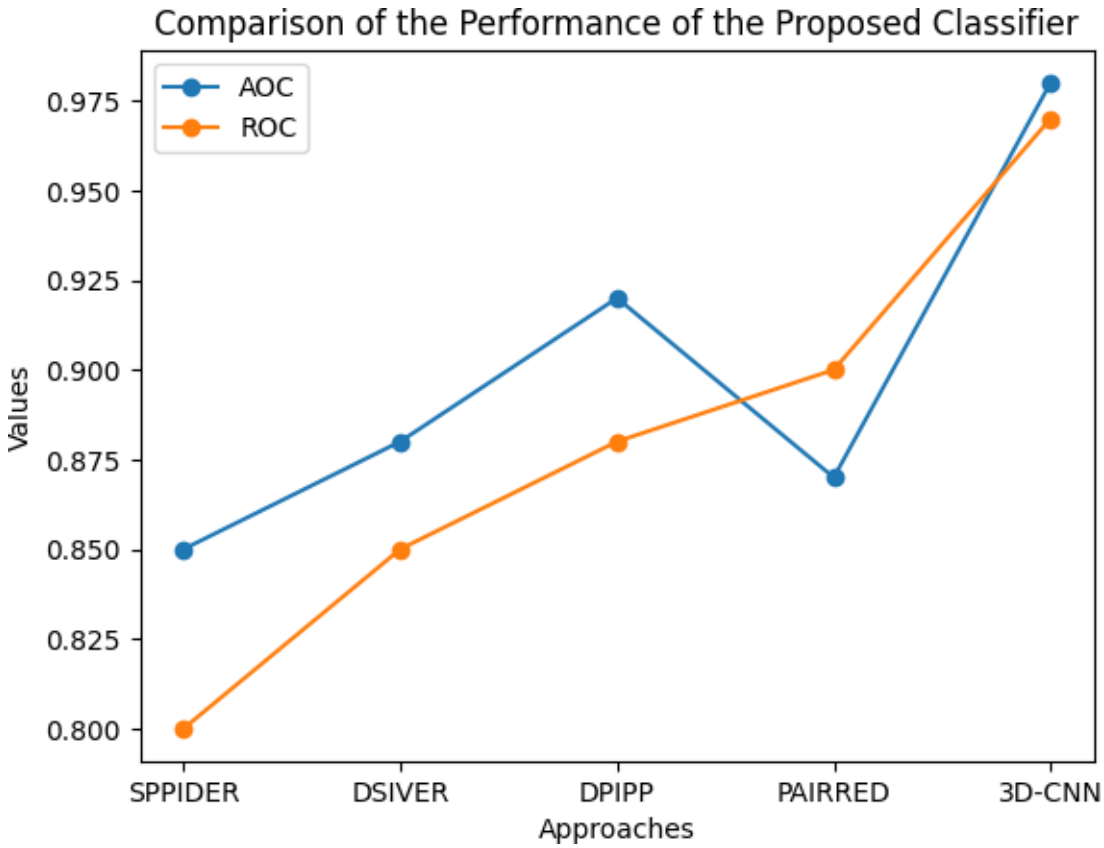

Figure 4: AOC, ROC comparison with baseline methods

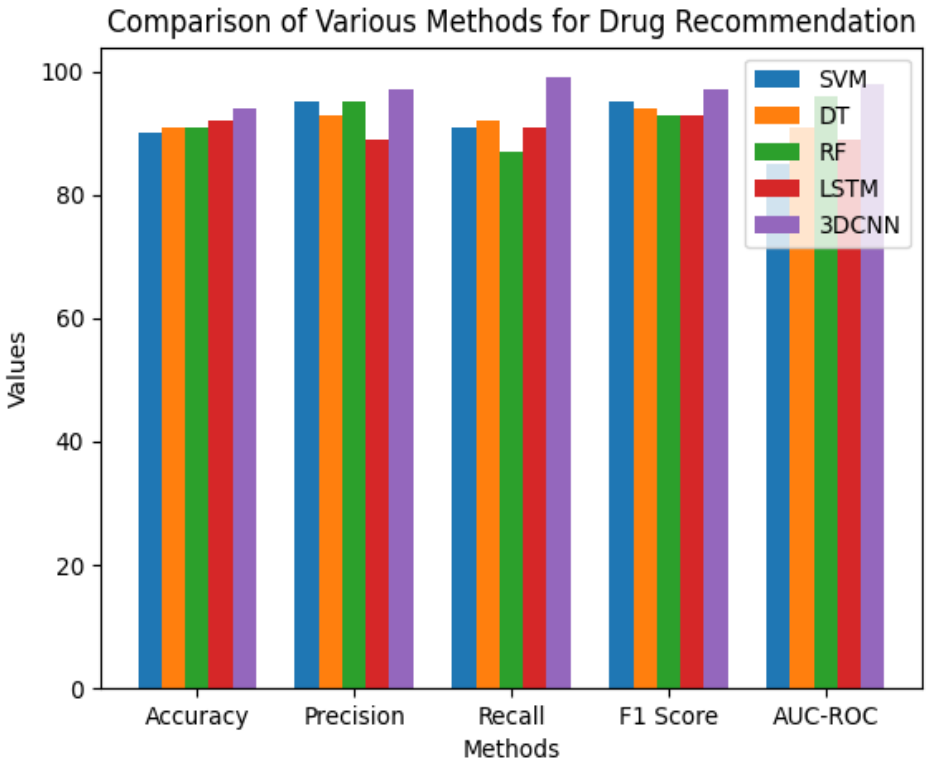

Figure 4: Drug Recommendation Comparison with existing methods

Matthews correlation coefficient (MCC) reported by 3DCNN is considerably more when compared with SVM, LSTM on all these EGF1, TRYPSINSER, RNASE, EFHRD1,16MHC benchmarked sites. F1 Score is promising with 3DCNN which reports 0.975 which outperforms the existing literature over various sites. One of the most vital parameter is Recall which was reported as 0.975 in average as the amino acid sequence is quite difficult in the helical strand. The other methods have reported 0.82 average recall value. 3DCNN has reported an average precision of 0.98. These values reported made this work as innovative and robust for different length of amino acid sequences

The performance of computer models in protein structure prediction is evaluated using the evaluation metrics AUC (Area Under the Curve) and ROC (Receiver Operating Characteristic). ROC curves plot the true positive rate versus the false positive rate across a range of threshold values since determining whether a residue in a protein structure is in contact or not is binary. The AUC represents the model's ability to discriminate between positive and negative cases by condensing the ROC curve into a single statistic. A high AUC in this case indicates that the model can accurately anticipate interactions between protein residues. These measures are essential for benchmarking and contrasting various protein structure prediction techniques, enabling researchers to assess the precision and dependability of computational models. The AOC, ROC values of 3D-CNN are better compared with the existing literature SPPIDER, DSIVER, DPIPP and PAIRRED. The ROC, AOC performance is far better compared to the base line methods with values 0.98 and 0.97 as shown in figure 4.

Accuracy in drug recommendations guarantees total correctness and reflects the system's dependability. Accuracy emphasises the necessity of prescribed medications, which is essential to prevent false positives. Recall highlights the system's capacity to identify all pertinent medications, avoiding false negative results. The F1 Score offers a fair assessment that takes recall and precision into account. The model's discriminatory power, which is essential for differentiating between relevant and irrelevant medication suggestions, is assessed using AUC-ROC. By guaranteeing relevant and correct medication recommendations, these measures together improve the efficacy, safety, and dependability of drug recommendation systems, assisting medical practitioners in making well-informed selections and improving patient outcomes as in figure 4.

3D-CNN reports an average accuracy of 98.2% for the drug suggestions which is very promising. The performance of the proposed classifier is tested with the standard methods SVM, DT,RF and LSTM over various parameters like precision, accuracy, F1 Score, recall and AU-ROC.

5. Conclusion

We have successfully implemented 3D Convolutional Neural Networks (3D CNNs) to predict protein structure for drug recommendation. 3D CNNs show improved prediction powers by utilising the spatial correlations found in protein structures, allowing for more accurate representations of molecular interactions. This ultimately results in enhanced drug recommendation algorithms. A thorough assessment of the model's performance is ensured by the inclusion of evaluation metrics such as accuracy, precision, recall, F1 Score, and AUC-ROC. By providing more precise and successful therapeutic interventions based on a deeper comprehension of molecular structures and interactions, the development of 3D CNNs for protein structure prediction has the potential to transform personalised medicine in addition to improving drug recommendations. 3D-CNN augmented with CA has provided a novel base for predicting the protein and then recommendation of the drugs. The accuracy of protein prediction is 98.26% and drug recommendation is 96.36%

References

[1] Hu, ShanShan, Chenglin Zhang, Peng Chen, Pengying Gu, Jun Zhang, and Bing Wang. "Predicting drug-target interactions from drug structure and protein sequence using novel convolutional neural networks." BMC bioinformatics 20 (2019): 1-12.

[2] Jiménez, José, Stefan Doerr, Gerard Martínez-Rosell, Alexander S. Rose, and Gianni De Fabritiis. "DeepSite: protein-binding site predictor using 3D-convolutional neural networks." Bioinformatics 33, no. 19 (2017): 3036-3042.

[3] Pu, Limeng, Rajiv Gandhi Govindaraj, Jeffrey Mitchell Lemoine, Hsiao-Chun Wu, and Michal Brylinski. "DeepDrug3D: classification of ligand-binding pockets in proteins with a convolutional neural network." PLoS computational biology 15, no. 2 (2019): e1006718.

[4] Wang, Xiao, Genki Terashi, Charles W. Christoffer, Mengmeng Zhu, and Daisuke Kihara. "Protein docking model evaluation by 3D deep convolutional neural networks." Bioinformatics 36, no. 7 (2020): 2113-2118.

[5] Sato, Rin, and Takashi Ishida. "Protein model accuracy estimation based on local structure quality assessment using 3D convolutional neural network." PloS one 14, no. 9 (2019): e0221347.

[6] Xu, Yinqiu, Hequan Yao, and Kejiang Lin. "An overview of neural networks for drug discovery and the inputs used." Expert opinion on drug discovery 13, no. 12 (2018): 1091-1102.

[7] Pokkuluri, Kiran Sree, and SSSN Usha Devi Nedunuri. "A novel cellular automata classifier for covid-19 prediction." Journal of Health Sciences 10, no. 1 (2020): 34-38.

[8] Vaz, Joel Markus, and S. Balaji. "Convolutional neural networks (CNNs): Concepts and applications in pharmacogenomics." Molecular diversity 25, no. 3 (2021): 1569-1584.

[9] Pagès, Guillaume, Benoit Charmettant, and Sergei Grudinin. "Protein model quality assessment using 3D

oriented convolutional neural networks." Bioinformatics 35, no. 18 (2019): 3313-3319.
[10] Kalakoti, Yogesh, Shashank Yadav, and Durai Sundar. "Deep Neural Network-Assisted Drug Recommendation Systems for Identifying Potential Drug–Target Interactions." ACS omega 7, no. 14 (2022): 12138-12146.
[11] Sree, P. Kiran, I. Ramesh Babu, and NSSSN Usha Devi. "Investigating an Artificial Immune System to strengthen protein structure prediction and protein coding region identification using the Cellular Automata classifier." International journal of bioinformatics research and applications 5, no. 6 (2009): 647-662.
[12] Diaz, Daniel J., Anastasiya V. Kulikova, Andrew D. Ellington, and Claus O. Wilke. "Using machine learning to predict the effects and consequences of mutations in proteins." Current Opinion in Structural Biology 78 (2023): 102518.
[13] Bordin, Nicola, Christian Dallago, Michael Heinzinger, Stephanie Kim, Maria Littmann, Clemens Rauer, Martin Steinegger, Burkhard Rost, and Christine Orengo. "Novel machine learning approaches revolutionize protein knowledge." Trends in Biochemical Sciences 48, no. 4 (2023): 345-359.
[14] Stahl, Kolja, Andrea Graziadei, Therese Dau, Oliver Brock, and Juri Rappsilber. "Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning." Nature Biotechnology (2023): 1-10.
[15] Chandra, Abel, Laura Tünnermann, Tommy Löfstedt, and Regina Gratz. "Transformer-based deep learning for predicting protein properties in the life sciences." Elife 12 (2023): e82819.
[16] Szelogowski, Daniel. "Deep Learning for Protein Structure Prediction: Advancements in Structural Bioinformatics." bioRxiv (2023): 2023-04.
[17] Silpa, C., B. Sravani, D. Vinay, C. Mounika, and K. Poorvitha. "Drug Recommendation System in Medical Emergencies using Machine Learning." In 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), pp. 107-112. IEEE, 2023.
[18] Pokkuluri, Kiran Sree, and Devi Nedunuri Usha. "A secure cellular automata integrated deep learning mechanism for health informatics." Int. Arab J. Inf. Technol. 18, no. 6 (2021): 782-788.
[19] Sellamuthu, Suseela, Srinivas Aditya Vaddadi, Srinivas Venkata, Hemant Petwal, Ravi Hosur, Vishwanadham Mandala, R. Dhanapal, and Jagendra singh. "AI-based recommendation model for effective decision to maximise ROI." Soft Computing (2023): 1-10.
[20] Sellamuthu, Suseela, Srinivas Aditya Vaddadi, Srinivas Venkata, Hemant Petwal, Ravi Hosur, Vishwanadham Mandala, R. Dhanapal, and Jagendra singh. "AI-based recommendation model for effective decision to maximise ROI." Soft Computing (2023): 1-10.
[21] Murugan, Suganiya, S. R. Srividhya, S. Pradeep Kumar, and B. Rubini. "A Machine Learning Approach to Predict Skin Diseases and Treatment Recommendation System." In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1157-1163. IEEE, 2023.
[22] Al Mubasher, Hadi, Ziad Doughan, Layth Sliman, and Ali Haidar. "A Novel Neural Network-Based Recommender System for Drug Recommendation." In International Conference on Engineering Applications of Neural Networks, pp. 573-584. Cham: Springer Nature Switzerland, 2023.
[23] Fujihara, Kazuya, and Hirohito Sone. "Machine Learning Approach to Drug Treatment Strategy for Diabetes Care." Diabetes & Metabolism Journal 47, no. 3 (2023): 325-332.
[24] Sree, Pokkuluri Kiran, S. S. S. N. Usha Devi N, Phaneendra Varma Chintalapati, Gurujukota Ramesh Babu, and P. B. V. Raja Rao. "Drug Recommendations Using a Reviews and Sentiment Analysis by RNN." In International Conference on Cognitive Computing and Cyber Physical Systems, pp. 135-141. Cham: Springer Nature Switzerland, 2023.