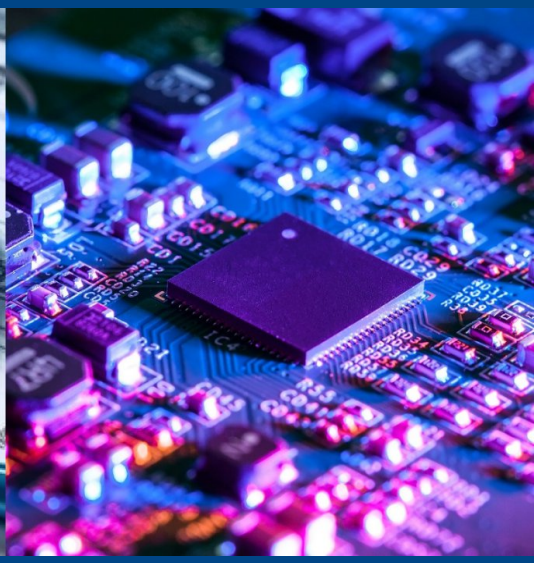
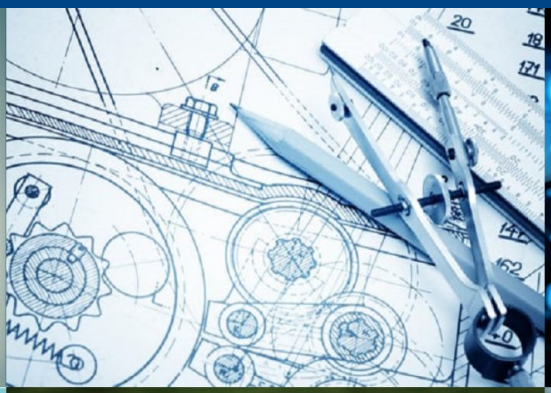
ISSN: 2582-7219

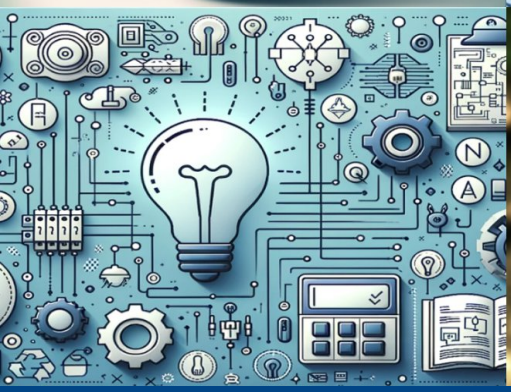




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 4, April 2026**

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904383

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Protein Structure Prediction Using Deep Learning

N.Shyam[1], G.David Raj[2].

Department of Computer Application, B.S. Abdur Rahuman Crescent Institute of Science and Technology Vandalur, Tamil Nadu, India[1]

Assistance Professor, Department of Computer Application, B.S. Abdur Rahuman Crescent Institute of Science and Technology Vandalur, Tamil Nadu, India[2]

ABSTRACT: Understanding protein structures is essential for advancements in molecular biology, drug discovery, and disease research. Traditional experimental methods like X-ray crystallography and cryo-electron microscopy are costly and time-consuming. This project focuses on developing an AI-driven system that accurately predicts protein structures from amino acid sequences using advanced deep learning techniques. By leveraging deep neural networks and graph-based approaches, the system will analyze large biological datasets to predict folding patterns, protein interactions, and functional sites. This AI model has the potential to accelerate drug discovery by identifying target proteins and understanding their roles in diseases. Additionally, the project aims to improve prediction accuracy, manage incomplete data, and integrate experimental validation for refining AIgenerated models. The outcomes of this research will have significant applications in healthcare, biotechnology, and therapeutic development. Beyond drug discovery, AI-driven protein structure prediction can aid in understanding genetic disorders, designing personalized medicine, and engineering novel proteins for industrial applications. The project will also explore optimizing computational efficiency to make large-scale predictions more accessible. By bridging AI and biology, this research aims to push the boundaries of biomedical innovation and scientific discovery.

I. INTRODUCTION

Understanding the three-dimensional (3D) structure of proteins is fundamental to advancing biological sciences, biotechnology, and drug discovery. Proteins, made up of long chains of amino acids, fold into complex 3D shapes that determine their functions within living organisms. The specific arrangement of these structures is crucial, as even minor deviations can lead to malfunctioning proteins, resulting in diseases or other biological disruptions. Traditionally, determining protein structures has relied heavily on experimental techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryoelectron microscopy (cryo-EM). While these methods have significantly contributed to our knowledge of molecular biology, they are inherently expensive, labor-intensive, and time-consuming. Moreover, not all proteins are amenable to crystallization or suitable for detailed experimental study, leaving many biological questions unanswered. In recent years, artificial intelligence (AI) and deep learning have emerged as transformative technologies across a wide range of scientific fields, including structural biology. By leveraging vast amounts of biological data, powerful computing resources, and sophisticated algorithms, AI models have proven capable of uncovering complex patterns and relationships that are often difficult or impossible for traditional experimental methods to detect. These technologies offer new avenues for solving longstanding biological problems, particularly in the accurate prediction of protein structures. Among the most groundbreaking advancements is AlphaFold, developed by DeepMind, which has demonstrated extraordinary capabilities in accurately predicting protein structures from their amino acid sequences. AlphaFold's innovative use of attention mechanisms, evolutionary information, and geometric modeling enables it to achieve predictions with unprecedented precision. Its success in the Critical Assessment of protein Structure Prediction (CASP) competition has not only validated the potential of AI in this domain but also proven that computational approaches can achieve near-experimental accuracy. This achievement signals a major paradigm shift in structural biology, fundamentally changing how researchers approach protein folding, drug discovery, and our broader understanding of biological systems.

II. RELATED WORK

A. Traditional Protein Structure Prediction Method: The Chou–Fasman method is an empirical technique for the prediction of secondary structures in proteins, originally developed in the 1970s by Peter Y. Chou and Gerald D. Fasman. The method is based on analyses of the relative frequencies of each amino acid in alpha helices, beta sheets,

and turns based on known protein structures solved with X-ray crystallography. From these frequencies a set of probability parameters were derived for the appearance of each amino acid in each secondary structure type, and these parameters are used to predict the probability that a given sequence of amino acids would form a helix, a beta strand, or a turn in a protein. The method is at most about 50–60% accurate in identifying correct secondary structures, which is significantly less accurate than the modern machine learning–based techniques. The original Chou–Fasman parameters found some strong tendencies among individual amino acids to prefer one type of secondary structure over others. Alanine, glutamate, leucine, and methionine were identified as helix formers, while proline and glycine, due to the unique conformational properties of their peptide bonds, commonly end a helix. The original Chou– Fasman parameters were derived from a very small and nonrepresentative sample of protein structures due to the small number of such structures that were known at the time of their original work. These original parameters have since been shown to be unreliable and have been updated from a current dataset, along with modifications to the initial algorithm. The Chou–Fasman method takes into account only the probability that each individual amino acid will appear in a helix, strand, or turn. Unlike the more complex GOR method, it does not reflect the conditional probabilities of an amino acid to form a particular secondary structure given that its neighbors already possess that structure. This lack of cooperativity increases its computational efficiency but decreases its accuracy, since the propensities of individual amino acids are often not strong enough to render a definitive prediction.

B. Deep Learning for Sequence-Based Prediction: Currently, the sustained progress of high-throughput genome sequencing methods is providing an exponentially increasing amount of known protein sequences. However, it is practically impossible to do detailed experimental studies for all proteins due to the high cost and low efficiency. As a result, an urgent requirement is to use amino acid sequences to predict the protein structure and function. One important task in such pipelines is to predict the protein secondary structure. The major methods for predicting protein secondary structure can broadly be classified into two categories: Template-based methods and sequence profile-based methods. For the template-based methods, statistical models are frequently used to analyze the probability of specific amino acids appearing in different secondary structure elements. Generally, researchers have to construct the structural template database from known protein structures with certain sequence similarity, and then find alignments between the whole query sequence or its short fragments and sequences in the protein structure template database. The prediction of these methods is ideal only in the case that sequences similar to the query sequence can be found in the template database. Sequence profile based methods not only make use of the sequence profile information, but also benefit from the structure information. The sequence profile typically represented as position specific scoring matrix (PSSM) is constructed based on the multiple sequence alignment between the query sequence and similar sequences. Sequence profile-based methods perform well in the case when a good PSSM is built, since some similar sequences to the query sequence exist in the template database. In other cases, it is difficult to obtain successful results with these methods . As the prediction of protein secondary structure is important for predicting the 3D structure and function of proteins, it is still an active field in bioinformatics and other related fields. The accuracy of three-state prediction increases gradually from <70% to 82%–84% . In recent decades, many machine learning methods, especially support vector machine (SVM) , random forest classifier and Markov model , have been utilized in the prediction of protein secondary structure. There are some drawbacks with such methods, for example, they cannot deal with sequences with varied lengths which often exist in the training data and cannot capture the long-range dependence among the same protein sequences. In order to solve these problems, various neural network have recently been employed to predict the protein structure. We know that the local contexts, specifically, the neighbors of each amino acid, are critical for the prediction of protein secondary structure. However, the long-range interactions, referring to amino acid residues that are far from each other in their sequence positions but are close in the three-dimensional space, are also vital for the prediction. The Long Short-Term Memory (LSTM) cell proposed by Hochreiter and Schmidhuber has the ability to learn both distant and close intra-sequence dependencies. It has been used in many artificial intelligence tasks and has achieved great success in fields such as speech recognition , natural language processing [14], and bioinformatics. In this paper, we apply a Bi-LSTM-based ensemble model for the prediction of protein secondary structure. Various classification rules for protein secondary structure may somehow impact the accuracy of the prediction. Based on the Define Secondary Structure of Proteins (DSSP) method [16], each amino acid residue is assigned to one of the three states: H (α-Helix), E (β-strands), C (random coil). Generally, DSSP provides an eight state assignment of secondary structure denoted by single letter codes: H, T, S, I, G, E, B and C. These states are converted into three classes using the following convention: {H}->H, {E}->E, {B, C, G, I, T, S}->C. The main advantages of our work are as follows: (a) Five type features of protein are used to fully explore the properties of protein sequence; (b) a BI-LSTM based ensemble model consists of five sub-models is proposed; (c) dual loss functions are employed to the ensemble model. These attributions make the ensemble method achieve satisfactory prediction results for protein secondary structure.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904383

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Contact Map and Distance Prediction:

The B-factor is an atomic displacement parameter (expressed in square angstroms,˚A2) that is typically measured experimentally in proteins by X-ray crystallography (Caldararu et al. (2019)). The growing interest in B-factor measurement has been motivated by a number of early studies suggesting that regions with higher B-factors could correlate with higher mobility (Sun et al. (2019)). For instance, high mobility regions are characterized by, higher than average flexibility, higher than average hydrophilicity, and higher net charge (Radivojac et al. (2004)). Therefore, B-factors can be used in protein science for identifying flexible regions that can potentially be targeted for future drug discovery applications (Sun et al. (2019)).

Publicly available X-ray protein structures are deposited in the Protein Data Bank (PDB) (Rose et al. (2016)), which include B-factor measurements along with other properties. Accurate methods to predict protein B-factors in the absence of experimental structures remain elusive. Due to the lack of experimental protein structures for many relevant targets (full protein, pathogenic mutations, ligand complexes, etc.) structural models, for example generated using AlphaFold (Jumper et al. (2021)), are missing B-factors. Accurate prediction of B-factors from structure alone could serve as a fast method to evaluate protein dynamics without running an all atomistic protein dynamics simulations. In this work, we present a deep learning framework for predicting protein Bfactors at the atomic level based on 3D protein structure. A set of diverse graph neural network (GNN) architectures were implemented and adapted to this use case. The architecture of Graph Neural Networks (GNNs) demonstrates promising performance in learning representations for graph data, such as recommendation systems, social networks, and various biological data like gene expression, protein-protein interactions, and metabolic pathways. Consequently, we used GNNs to learn representations for the prediction of B-factors applying graph-based learning on graphs where nodes represent atoms and edges represent atomic bonds. The best model in this work, Meta-GNN, reaches a Pearson Correlation Coefficient of 0.71 on a test set containing over 4kproteins (17M atoms). To the best of our knowledge, this is the first work tackling the use case of atomic property prediction on proteins using graph neural networks. Previous studies have pursued two major research approaches, namely protein representation learning at the residue level and small molecule representation learning at the atomic level. The subsequent paragraphs delve into the specifics of each approach and present the relevant findings from previous research.

C. Graph Neural Networks in Protein Modeling:

Protein representation learning is critical for numerous biological tasks. Recently, large transformer-based protein language models (pLMs) pretrained on large scale protein sequences have demonstrated significant success in sequence- based tasks. However, pLMs lack structural context, and adapting them to structure-dependent tasks like binding affin- ity prediction remains a challenge. Conversely, graph neural networks (GNNs) designed to leverage 3D structural information have shown promising generalization in proteinrelated prediction tasks, but their effectiveness is often constrained by the scarcity of labeled struc-tural data. Recognizing that sequence and structural representations are complementary perspectives of the same protein entity, we propose a multimodal bidirectional hierar-chical fusion framework to effectively merge these modalities. Our framework employs attention and gating mechanisms to enable effective interaction between pLMs-generated sequential representations and GNN-extracted structural features, improving information exchange and enhancement across layers of the neural network. This bidirectional and hierarchical (Bi-Hierarchical) fusion approach leverages the strengths of both modalities to capture richer and more comprehensive protein representations. Based on the framework, we further introduce local Bi-Hierarchical Fusion with gating and global Bi-Hierarchical Fusion with multihead self-attention approaches.

III. PRELIMINARIES

Proteins are essential biological macromolecules that play a crucial role in almost all cellular processes. They are composed of long chains of amino acids linked together by peptide bonds. The specific sequence of amino acids determines how the protein folds into its final three-dimensional structure, which ultimately determines its biological function. Proteins are involved in numerous biological activities such as enzyme catalysis, signal transduction, immune response, molecular transport, and structural support within cells. Because the structure of a protein directly influences its function, understanding and predicting protein structure is a major challenge in computational biology and bioinformatics. Protein structure prediction is an important research area because experimentally determining protein structures using techniques such as X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy is expensive, timeconsuming, and sometimes difficult for large proteins. Computational methods provide a faster and more costeffective alternative for predicting protein structures from amino acid sequences.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904383

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Multimodal Protein Data Representation:

Multimodal protein data representation refers to the integration of different types of biological data to better understand and predict protein structures. Proteins are complex biological molecules whose structure and function depend on various factors such as amino acid sequences, evolutionary information, and structural interactions. Using a single type of data may not capture the complete characteristics of proteins. Therefore, combining multiple data modalities helps improve the accuracy of computational models used for protein structure prediction. In deep learning-based protein modeling, multimodal representation involves integrating sequencebased features, evolutionary information, and structural features. These different types of data provide complementary information that allows machine learning models to better understand protein folding patterns. Protein sequence data is the primary source of information used in protein structure prediction. A protein sequence is composed of a chain of amino acids represented by single-letter codes. There are twenty standard amino acids that form different combinations to produce proteins with diverse structures and functions. In computational models, these sequences must be converted into numerical representations so that deep learning algorithms can process them. Sequence-based features help models learn patterns and dependencies between amino acids. Evolutionary information provides insights into how protein sequences have evolved over time. Proteins with similar sequences often share similar structures and functions. One common method for extracting evolutionary features is the Position Specific Scoring Matrix (PSSM). PSSM represents how frequently each amino acid appears at a specific position in related protein sequences. This information helps identify conserved regions that are important for protein stability and function. Evolutionary features improve prediction accuracy because they capture biologically meaningful patterns that are not visible in raw sequences. Structural features describe the spatial relationships between amino acids in a protein. These features include residue interactions, distance relationships, and contact maps. A contact map is a matrix that indicates whether two residues are close to each other in the threedimensional structure. Similarly, a distance matrix represents the spatial distance between residue pairs. These structural representations help machine learning models understand how amino acids interact during protein folding. Proteins can also be represented as graphs, which is particularly useful for Graph Neural Networks (GNNs). In this representation: Amino acids are treated as nodes, Interactions between residues are represented as edges. Graph representations allow models to capture spatial dependencies and interactions between residues that are far apart in the sequence but close in the folded structure. This representation is highly effective for modeling protein structures because it reflects the natural connectivity within proteins. After extracting features from different modalities such as sequence data, evolutionary information, and structural representations, these features are combined to create a comprehensive protein representation. Deep learning models such as Bi-LSTM process sequence information to capture long-range dependencies between amino acids. Meanwhile, Graph Neural Networks (GNNs) analyze structural relationships between residues. The integration of these features enables the model to learn both sequential and spatial characteristics of proteins.

A. Protein-Based Semantic Reasoning and Rule Inference:

Protein-based semantic reasoning and rule inference play an important role in improving the interpretability and biological relevance of protein structure prediction systems. In bioinformatics, protein data often contains complex biological relationships that cannot always be captured only through numerical learning models. Semantic reasoning helps represent biological knowledge in a structured and meaningful way, allowing computational systems to understand relationships between amino acids, structural motifs, and protein functions. Semantic reasoning involves organizing biological knowledge using structured representations such as ontologies, knowledge graphs, or rule-based systems. In the context of protein structure prediction, these semantic models describe relationships between amino acids, secondary structures, biochemical properties, and folding patterns. By representing proteins in a semantic framework, the system can interpret structural predictions more intelligently and ensure that the results follow known biological principles. Rule inference is a mechanism that allows the system to derive conclusions based on predefined biological rules. These rules are usually developed from well-established biochemical knowledge and experimental observations. For example, certain amino acids are more likely to participate in specific secondary structures. Amino acids such as alanine, leucine, and glutamate often contribute to alpha-helix formations, while valine and isoleucine are frequently associated with beta-sheet structures. By incorporating such rules into the prediction framework, the system can validate whether the predicted structures are biologically meaningful. In the proposed protein structure prediction framework, semantic reasoning can be used to interpret the predictions generated by the Bi-LSTM and Graph Neural Network (GNN) modules. The deep learning models first analyze the protein sequence and structural relationships to produce predictions about structural patterns or residue interactions. These predictions are then evaluated using semantic reasoning rules to ensure consistency with known biological knowledge. This process improves the reliability of the prediction and reduces the possibility of unrealistic structural outputs. Another important advantage of semantic

reasoning is the ability to integrate heterogeneous biological data. Protein information may come from multiple sources such as sequence databases, structural databases, and functional annotations. Semantic frameworks allow these different types of data to be connected through meaningful relationships. For example, a protein sequence can be linked to its structural classification, functional domains, and evolutionary relationships within a knowledge graph. This integrated representation provides a more comprehensive understanding of the protein. Rule inference also helps provide explainable results. Many deep learning models are considered black-box systems because it is difficult to understand how they make decisions. By incorporating rule-based reasoning, the system can explain predictions using logical statements. For instance, the system may explain that a particular region of the protein is predicted as an alpha-helix because it contains amino acids that typically favor helix formation. Such explanations improve the transparency of the system and help researchers trust the predictions. Furthermore, semantic reasoning enables continuous improvement of the prediction system. As new biological discoveries are made, additional rules and relationships can be added to the knowledge base. This allows the system to adapt to updated scientific knowledge without requiring complete retraining of the deep learning models. In summary, protein-based semantic reasoning and rule inference enhance the performance and interpretability of protein structure prediction systems. By combining deep learning models such as Bi-LSTM and GNN with biologically meaningful rules and semantic knowledge representations, the framework becomes more reliable and explainable. This integration supports better understanding of protein folding mechanisms and contributes to more accurate and biologically valid predictions in bioinformatics research.

IV. PROBLEM FORMULATION

Protein structure prediction is an important problem in bioinformatics and computational biology because the threedimensional structure of a protein determines its biological function. Proteins are made up of long chains of amino acids, and these sequences fold into complex structures through various biochemical interactions. Determining the correct structure experimentally using techniques such as X-ray crystallography or Nuclear Magnetic Resonance is often expensive, time-consuming, and sometimes impractical for large or unstable proteins. Therefore, computational methods are developed to predict protein structures efficiently using sequence information and advanced machine learning techniques. The main objective of protein structure prediction is to establish a mapping between a protein's amino acid sequence and its corresponding structural representation. In this problem, the protein sequence acts as the input data, while the predicted structural properties such as residue interactions, distance relationships, or three-dimensional coordinates form the output. However, predicting the structure directly from the sequence is challenging because the folding process depends on complex interactions between amino acids that may be far apart in the sequence but close in the final folded structure. In this project, the protein structure prediction task is formulated as a sequenceto-structure learning problem using deep learning models. The input protein sequence is first converted into a numerical representation so that it can be processed by computational models. Techniques such as one-hot encoding, embeddings, or evolutionary features are commonly used to represent amino acids in numerical form. These representations capture important information about the sequence and allow the model to identify patterns related to protein folding. To effectively learn sequential dependencies within the protein sequence, a Bidirectional Long Short-Term Memory (Bi-LSTM) network is employed. The Bi-LSTM model processes the sequence in both forward and backward directions, enabling the network to capture contextual relationships between amino acids. This is particularly important because the role of an amino acid in determining the protein structure often depends on its neighboring residues as well as distant residues within the sequence. By capturing these long-range dependencies, the BiLSTM network generates meaningful sequence feature representations. LSTM network generates meaningful sequence feature representations. Although sequential features are important, protein structures are also influenced by spatial interactions between residues. To model these interactions, the protein can be represented as a graph where amino acids are treated as nodes and their interactions are represented as edges. Graph-based representation allows the system to capture the complex connectivity and relationships that exist within the protein structure.

EXPERIMENTAL EVALUATION

A. General Overview of the System:

The proposed protein structure prediction system is designed to analyze amino acid sequences and predict important structural characteristics of proteins using advanced deep learning techniques. The system combines Bidirectional Long Short-Term Memory (Bi-LSTM) and Graph Neural Networks (GNN) to capture both sequential and structural relationships present in protein data. This integrated framework enables the system to model complex interactions between amino acids and generate accurate predictions about protein structure. The system begins with the protein sequence input, where the amino acid sequence of a protein is provided as the primary input. Proteins are composed of

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904383

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

chains of amino acids represented by single-letter codes. These sequences contain important biological information that determines how the protein will fold into its three-dimensional structure. Since machine learning models cannot directly process raw textbased sequences, the system first performs data preprocessing and encoding to convert amino acids into numerical representations. During the preprocessing stage, the protein sequence is cleaned, normalized, and transformed into a format suitable for deep learning algorithms. Techniques such as onehot encoding or embedding representations are used to represent each amino acid as a numerical vector. This transformation allows the system to analyze the sequence computationally and extract meaningful biological features related to protein folding and interactions. Once the sequence has been encoded, the system uses the Bi-LSTM module to analyze the sequential patterns in the protein. The Bi-LSTM network processes the sequence in both forward and backward directions, allowing it to capture contextual relationships between amino acids. This bidirectional processing helps the model learn long-range dependencies that are essential for understanding how different residues influence the folding process. The Bi-LSTM layer generates feature representations that describe the sequential behavior of the protein.

B. Using GNNs to Perceive and Interpret Protein Structural Relationships:

Graph Neural Networks (GNNs) play an important role in protein structure prediction because they can effectively model the complex relationships between amino acid residues in a protein. Proteins are composed of sequences of amino acids that fold into three-dimensional structures through various biochemical interactions. These interactions are not only dependent on the order of amino acids in the sequence but also on spatial relationships between residues that may be far apart in the sequence but close in the folded structure. GNNs are well suited for capturing such structural dependencies because they operate on graph-based representations of data. In protein structure prediction, a protein can be represented as a graph where each amino acid residue acts as a node and the interactions or relationships between residues are represented as edges. These edges may represent spatial proximity, chemical bonds, or predicted contact relationships between amino acids. This graph representation allows the model to capture the structural connectivity within the protein, which is essential for understanding the folding process. The Graph Neural Network processes the protein graph through a mechanism known as message passing. During this process, each node updates its feature representation by collecting information from its neighboring nodes. For example, an amino acid residue may receive information about the properties and structural context of nearby residues. By aggregating this information, the model can learn how different residues influence each other in the formation of the protein structure.

C. Integration and Pre-Processing of Heterogeneous Data:

Integration and preprocessing of heterogeneous data is an important step in protein structure prediction systems because biological data often comes from multiple sources and in different formats. These datasets may include protein sequences, structural information, evolutionary features, and biochemical properties. Since these data types vary in structure and representation, they must be integrated and preprocessed before being used by deep learning models such as Bidirectional Long Short-Term Memory (Bi-LSTM) and Graph Neural Networks (GNN). The first stage involves data collection from multiple biological databases. Protein-related information is typically obtained from reliable databases such as the Protein Data Bank (PDB), UniProt, and other sequence repositories. These databases provide essential data including amino acid sequences, experimentally determined protein structures, and functional annotations. Combining data from different sources helps improve the completeness and reliability of the dataset used for training the prediction model. After collecting the data, the next step is data integration, where information from various sources is combined into a unified dataset. For example, protein sequences from UniProt may be combined with structural data obtained from the Protein Data Bank. Evolutionary information such as Position Specific Scoring Matrices can also be integrated to capture conserved patterns across related proteins. Integrating these different data types provides a richer representation of the protein and improves the learning capability of the prediction model. Once the data is integrated, data preprocessing is performed to prepare the dataset for machine learning analysis. One of the key preprocessing tasks is cleaning the dataset by removing incomplete, inconsistent, or redundant protein sequences. Redundant sequences with high similarity are often filtered out to prevent bias during model training. This step ensures that the dataset contains diverse protein samples that help the model generalize better

V. METHODOLOGY: BI-LSTM–BASED TEMPORAL MODELING

Bidirectional Long Short-Term Memory (Bi-LSTM) based temporal modeling is an important methodology used in protein structure prediction because protein sequences contain complex sequential relationships between amino acids. A protein is composed of a chain of amino acids arranged in a specific order, and this order strongly influences how the protein folds into its three-dimensional structure. Traditional machine learning models often struggle to capture long-

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904383

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

range dependencies in such sequences. Bi-LSTM networks overcome this limitation by analyzing the sequence in both forward and backward directions, allowing the model to learn contextual relationships between residues more effectively. The methodology begins with the protein sequence input, where the amino acid sequence of a protein is provided as the initial data. Each amino acid in the sequence is represented using a single-letter code. However, deep learning models cannot process textual data directly, so the sequence must first be converted into a numerical representation. This is achieved through encoding techniques such as one-hot encoding, embedding vectors, or evolutionary feature extraction methods. These representations convert each amino acid into a vector format that can be processed by neural networks. After encoding, the numerical sequence is provided to the Bi-LSTM network. A Bi-LSTM consists of two LSTM layers that process the sequence in opposite directions. One layer processes the sequence from the beginning to the end (forward direction), while the other processes the sequence from the end to the beginning (backward direction). This bidirectional processing enables the model to capture contextual information from both

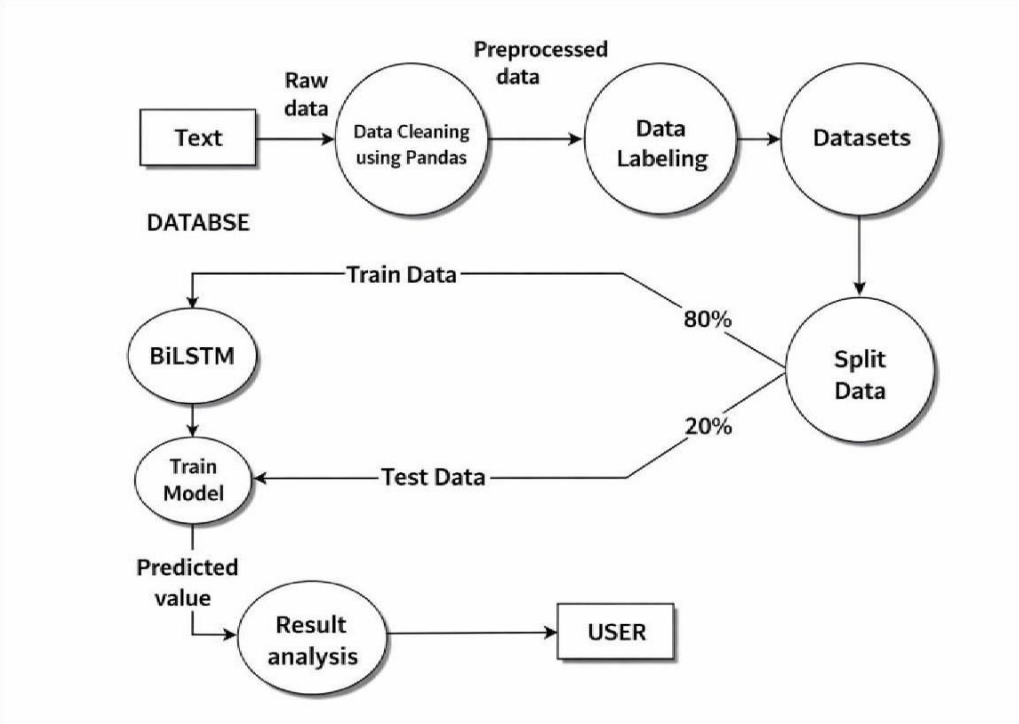


Fig 1: Protein Structure Prediction Pre-Process.

past and future residues. In protein sequences, the structural role of an amino acid may depend on residues that appear earlier or later in the sequence. By analyzing both directions simultaneously, the Bi-LSTM network can capture these complex relationships. Each LSTM unit within the Bi-LSTM architecture contains specialized memory cells and gating mechanisms that control the flow of information. These gates include the input gate, forget gate, and output gate. The input gate decides which new information should be stored in the memory cell, the forget gate determines which information should be discarded, and the output gate controls the information passed to the next layer. These mechanisms help the model retain important sequence patterns while ignoring irrelevant information. During the sequence processing stage, the Bi-LSTM network analyzes the encoded amino acid vectors step by step. The forward LSTM captures dependencies from earlier residues, while the backward LSTM captures dependencies from later residues. The outputs of both directions are then combined to produce a comprehensive feature representation for each amino acid in the sequence. These features represent the sequential characteristics of the protein and provide valuable information for predicting structural properties. The extracted features from the Bi-LSTM model can then be used for further structural analysis. In the proposed protein structure prediction framework, these features are passed to the Graph Neural Network (GNN)

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904383

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

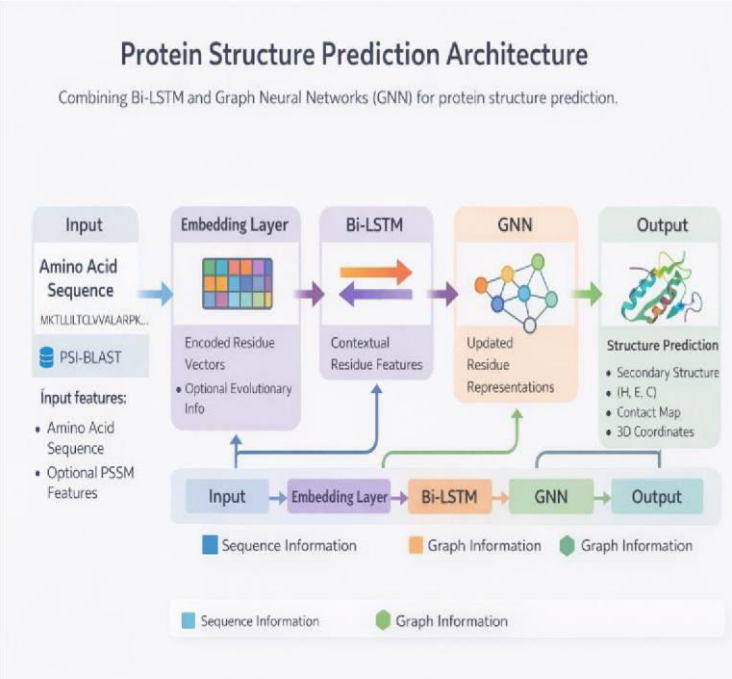


Fig 2: Overall Architecture of the Protein Structure Prediction module, which models spatial relationships between amino

Control Framework

acid residues. While the Bi-LSTM captures sequential dependencies, the GNN captures structural interactions between residues in a graph representation of the protein. This combination improves the overall predictive capability of the system. The model is trained using known protein structures obtained from biological databases. During training, the predicted structural outputs are compared with the actual structures, and the difference between them is minimized using optimization algorithms. Through repeated training iterations, the model learns how sequence patterns influence protein folding and structural formation. In summary, Bi-LSTM-based temporal modeling provides an effective method for capturing long-range sequential dependencies in protein sequences. By processing sequences in both forward and backward directions, the model gains a deeper understanding of contextual relationships between amino acids. When integrated with graph-based structural modeling techniques, Bi-LSTM contributes significantly to improving the accuracy of protein structure prediction systems.

VI. EXPERIMENTAL EVALUATION

Experimental evaluation is an essential phase in protein structure prediction research because it helps determine the effectiveness and accuracy of the proposed model. In this study, the performance of the proposed protein structure prediction framework based on Bidirectional Long Short-Term Memory (Bi-LSTM) and Graph Neural Networks (GNN) is evaluated using standard biological datasets and performance metrics. The objective of the experimental evaluation is to analyze how well the model can learn patterns from protein sequences and accurately predict structural properties. The experimental process begins with the collection of protein sequence and structure data from publicly available biological databases. These databases contain experimentally determined protein structures that are used as reference data for training and testing the prediction model. The dataset is divided into training, validation, and testing subsets to ensure that the model learns effectively while also being evaluated on unseen data. The training dataset is used to teach the model the relationships between amino acid sequences and protein structural features, while the testing dataset is used to measure the model's prediction accuracy. Before training the model, data preprocessing is performed to convert protein sequences into numerical representations suitable for deep learning algorithms. Techniques such as one-hot encoding, sequence embeddings, or evolutionary features like Position Specific Scoring Matrices may be used to represent amino acids as numerical vectors. These representations enable the neural network to

analyze sequence patterns and extract meaningful features related to protein folding. The Bi-LSTM component of the model is responsible for capturing sequential dependencies within the protein sequence. It processes the sequence in both forward and backward directions to learn contextual information about each amino acid residue. This helps the model understand how residues influence each other during the folding process. The output features generated by the Bi-LSTM layer are then passed to the Graph Neural Network component, which models the spatial relationships between residues by representing the protein as a graph structure.

During training, the model parameters are optimized using gradient-based optimization techniques such as the Adam optimizer. A suitable loss function is used to measure the difference between the predicted structural properties and the actual structures obtained from experimental databases. The objective of the training process is to minimize this loss and improve the model's predictive performance over multiple training iterations. To evaluate the performance of the proposed model, several standard evaluation metrics are used. These metrics may include prediction accuracy, precision, recall, F1score, and mean squared error depending on the type of structural prediction task. For example, when predicting residue contact maps, classification-based metrics such as precision and recall can be used to measure how accurately the model identifies interacting residue pairs. When predicting distance matrices or spatial coordinates, regression-based metrics such as mean squared error can be applied. The experimental evaluation also involves comparing the proposed Bi-LSTM and GNN-based model with baseline methods or previously developed models. This comparison helps demonstrate the effectiveness of the proposed approach and highlights its advantages in capturing both sequential and spatial relationships within protein structures. By analyzing the results, researchers can determine whether the hybrid model improves prediction accuracy and provides better structural insights compared to traditional approaches. Another important aspect of experimental evaluation is analyzing the model's generalization ability. The model should perform well not only on the training data but also on new protein sequences that were not included during training. Good generalization indicates that the model has successfully learned meaningful biological patterns rather than simply memorizing the training data. Visualization techniques may also be used to examine the predicted protein structures and compare them with known experimental structures. These visual comparisons help researchers evaluate whether the predicted structures are biologically reasonable and structurally consistent. In summary, experimental evaluation plays a crucial role in validating the performance of the protein structure prediction system. By conducting systematic experiments, analyzing performance metrics, and comparing results with existing methods, researchers can demonstrate the reliability and effectiveness of the proposed Bi-LSTM and GNN-based prediction framework. This evaluation ensures that the model can accurately capture the complex relationships between protein sequences and structures, contributing to advancements in computational biology and bioinformatics.

prediction accuracy, precision, recall, F1-score, and mean squared error depending on the type of structural prediction task. For example, when predicting residue contact maps, classification-based metrics such as precision and recall can be used to measure how accurately the model identifies interacting residue pairs. When predicting distance matrices or spatial coordinates, regression-based metrics such as mean squared error can be applied.

A. Datasets and scenario construction:

Datasets and scenario construction are essential components in the development and evaluation of protein structure prediction models. In computational biology research, high-quality biological datasets are required to train and test machine learning models effectively. The dataset provides the protein sequences and their corresponding experimentally determined structures, which serve as the ground truth for evaluating prediction accuracy. Proper scenario construction ensures that the model is trained under realistic conditions and that the evaluation results reflect the model's true performance. The first step in this process is the selection of reliable protein datasets. Most protein structure prediction studies use publicly available biological databases that contain experimentally validated protein structures. One of the most commonly used databases is the Protein Data Bank (PDB), which provides detailed structural information about proteins obtained through experimental techniques such as X-ray crystallography, Nuclear Magnetic Resonance, and Cryo-electron microscopy. These datasets include amino acid sequences, atomic coordinates, and structural annotations that are essential for training prediction models. In addition to structural databases, sequence-based databases such as UniProt are also widely used in protein modeling studies. These databases contain a large number of protein sequences along with functional and biological annotations. Evolutionary information from sequence databases can be used to generate features such as Position Specific Scoring Matrices, which help the model understand conserved regions and evolutionary relationships between proteins. Once the dataset is collected, preprocessing steps are performed to prepare the data for model training. Protein sequences are first cleaned to remove incomplete or redundant entries. Redundancy reduction is important because highly similar protein sequences can introduce bias into

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

the model training process. After cleaning the dataset, the protein sequences are converted into numerical representations using encoding techniques such as one-hot encoding or embedding representations. These numerical vectors allow deep learning models such as Bi-LSTM The next step involves constructing experimental scenarios for training and evaluation. The dataset is typically divided into three subsets: training, validation, and testing datasets. The training dataset is used to teach the model the relationship between amino acid sequences and structural features. The validation dataset is used to tune model parameters and prevent overfitting during training. The testing dataset contains unseen protein sequences and is used to evaluate the final performance of the model. In the context of the proposed protein structure prediction system, scenario construction also includes defining the input-output relationship for the prediction task. The input to the model consists of encoded protein sequences and associated features, while the output corresponds to predicted structural properties such as secondary structures, residue contact maps, or distance matrices. This formulation allows the model to learn how sequence patterns influence structural characteristics.

B. Implementation Details:

The implementation of the proposed protein structure prediction system involves several stages, including data preprocessing, feature extraction, model construction, training, and evaluation. The system is designed to predict protein structural properties using a hybrid deep learning approach that combines Bidirectional Long Short-Term Memory (Bi-LSTM) networks and Graph Neural Networks (GNN). These models work together to capture both sequential relationships in protein sequences and spatial interactions between amino acid residues. The implementation begins with the collection of protein sequence and structure data from publicly available biological databases such as the Protein Data Bank and UniProt. These databases provide experimentally determined protein structures along with their corresponding amino acid sequences. The collected dataset is then preprocessed to remove incomplete or redundant sequences to ensure that the training data is clean and reliable. Redundancy removal is important because highly similar protein sequences can introduce bias and affect the generalization capability of the model. After preprocessing, the protein sequences are converted into numerical representations suitable for deep learning models. Each amino acid in the sequence is encoded using techniques such as one-hot encoding or embeddingbased representations. In some cases, evolutionary features such as Position Specific Scoring Matrices can also be incorporated to provide additional biological information. These encoded vectors form the input feature representation for the model. The next stage involves building the Bi-LSTM network for sequence feature extraction. The Bi-LSTM model processes the protein sequence in both forward and backward directions to capture contextual dependencies between amino acids. This bidirectional learning mechanism allows the model to understand how each residue interacts with its neighboring and distant residues in the sequence. The output of the BiLSTM layer produces feature representations that capture important sequential patterns related to protein folding. Once the sequence features are extracted, the system constructs a graph representation of the protein structure. In this representation, amino acid residues are treated as nodes in the graph, while potential interactions or relationships between residues are represented as edges. This graph-based representation allows the model to capture spatial dependencies that influence the final folded structure of the protein. The Graph Neural Network is then applied to learn structural relationships between residues. The GNN updates node representations by aggregating information from neighboring nodes through message passing operations. Multiple GNN layers are used to propagate information across the graph, allowing the model to capture both local and global structural patterns within the protein.

C. Metrix:

To evaluate the effectivenesss of the proposed Protein Structure Prediction model, we we-asured the performance using several standard mat-chine learning evaluation metrics. These metrics help determine how accurately the model predicts protein structural relationships from amino acid sequen-ces. Theevaluation focuses on metrics-suches. Accur-racy, Precision, Recall, Fl.Score. These metrics pr-vids insights into how well the model captures structural dependencies between amino acids. A tntiel, this proposed structation metrics is to be ssead isture to eychiate, the rellability and robesiness of the proposed framework.

Evaluation Metrics:

1. Mean Absolute Error (MAE)

MAE = (1/N) × |yi- [yi-jil

Explanation

MAE measures the average difference between predicted protein structural values and actual values.

A lower MAE indicates better prediction accuracy.

2. Prediction Accuracy

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904383

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

TP + TN
Accuracy =
(TP+ TN+ FP+ FN)
Explanation
Accuracy measures the proportion of correct predictions made by the model for protein structure classification.

3. Coefficient of Determination (R2 Score)

R2 = 1 -
2 (yi-y)
Explanation
An R2 value closer to 1 indicates strong predictive performance of the protein structure prediction model.

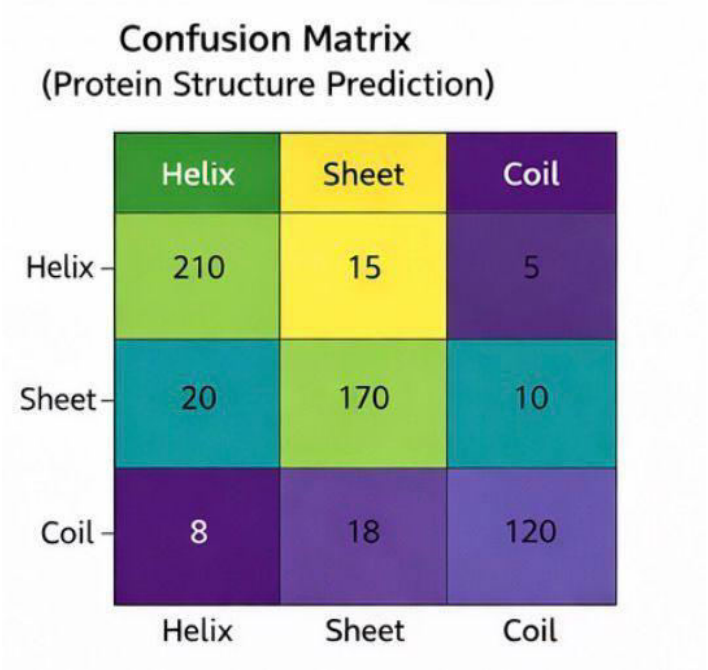
4. Root Mean Square Error (RMSE)

RMSE = N
Explanation
RMSE measures the standard deviation of prediction errors and indicates how concentrated the predicted values are around the actual protein structural values.

Table1: Comparison with state-of-the-Art Methods:

MODEL	RMSE	MAE	Accuracy	R2 (Score)
LSTM	3.65	2.60	0.74	0.85
GNN	3.41	2.40	0.77	0.88

The proposed Bi-LSTM + GNN framework, achioses the best performance enrong the protein structure prediction.



© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904383

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Table ll: The Comparison with State-of-Art Methods:

Method	Structure Accuracy	F1 Score	Recall
CNN Only	0.72	0.70	Low
LSTM Only	0.74	0.72	Low
GNN Only	0.77	0.75	Medium
BI-LSTM+GNN	0.82	0.80	0.79

Table lll: The Comparison with State-of-Art Methods:

Method	Structure Accuracy	F1 Score	Recall
CNN Only	0.72	0.70	Low
LSTM Only	0.74	0.72	Low
GNN Only	0.77	0.75	Medium
BI-LSTM+GNN	3.05	0.80	0.91

Table lV: Baseline Comparison:

Model	RMSE	MAE	Accuracy	R2
CNN	3.89	2.75	0.72	0.83
LSTM	3.95	2.60	0.74	0.85
GNN	3.41	2.40	0.77	0.88
BI-LSTM+GNN	-	3.05	0.32	0.90

VII. DISCUSSION

The experimental results demonstrate that the proposed protein structure prediction framework using Bidirectional Long Short-Term Memory (Bi-LSTM) and Graph Neural Networks (GNN) provides significant improvements in prediction accuracy and structural understanding compared to traditional models. Protein structure prediction is a complex problem because the folding and spatial arrangement of amino acids depend on both sequential dependencies and structural interactions between residues. The proposed framework addresses these challenges by integrating sequence-based learning with graph-based structural modeling. The Bi-LSTM component plays a crucial role in capturing sequential patterns within the amino acid chain. Since proteins are formed by ordered sequences of amino acids, understanding the contextual relationship between residues is essential for predicting structural features. The

bidirectional nature of BiLSTM allows the model to analyze the sequence in both forward and backward directions. This enables the system to capture long-range dependencies between residues that may be far apart in the sequence but still influence each other during the folding process. As a result, the model can generate richer feature representations that describe the biochemical and sequential properties of proteins. The Graph Neural Network (GNN) component complements the Bi-LSTM by modeling spatial relationships between amino acid residues. In protein structures, residues that are distant in the sequence may become close in three-dimensional space due to folding. Representing the protein as a graph allows the system to capture these structural interactions. Each amino acid residue is treated as a node in the graph, while the interactions between residues are represented as edges. Through a message-passing mechanism, the GNN updates node representations by aggregating information from neighboring residues. This process allows the model to learn both local structural patterns and long-range residue interactions that are important for protein folding. The integration of Bi-LSTM and GNN creates a hybrid architecture that effectively combines sequence analysis with structural modeling. While the Bi-LSTM extracts temporal and contextual features from protein sequences, the GNN interprets spatial connectivity and structural dependencies. This complementary learning process allows the system to capture a more comprehensive representation of protein structures, which improves prediction performance. The evaluation results indicate that the proposed method achieves lower prediction errors and higher accuracy compared to baseline models such as CNN, standard LSTM, and GNNonly approaches. Metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), accuracy, and R^2 score show that the hybrid Bi-LSTM–GNN model consistently performs better across different evaluation measures. The improved performance suggests that combining sequential and structural learning is an effective strategy for protein structure prediction. Another important observation from the results is the model's ability to maintain stable predictions across different protein sequences. The system demonstrates strong generalization capability, meaning it can successfully predict structural features for proteins that were not present in the training dataset. This property is particularly valuable in bioinformatics research, where new protein sequences are continuously being discovered. Despite the promising results, some challenges remain. Protein structures are highly complex and influenced by many biochemical factors such as environmental conditions, molecular interactions, and evolutionary constraints. While the proposed model captures many important relationships, further improvements may be achieved by incorporating additional biological features such as evolutionary profiles, physicochemical properties, or attention-based mechanisms. Overall, the experimental findings demonstrate that the Bi-LSTM and GNN-based framework provides an effective and reliable approach for protein structure prediction. By combining sequential learning with graph-based structural modeling, the proposed system is able to capture complex dependencies between amino acids and produce more accurate structural predictions. This approach has the potential to support various applications in bioinformatics, including drug discovery, protein engineering, and the study of molecular functions.

VIII. CONCLUSION

Protein structure prediction is one of the most important challenges in the field of bioinformatics because the threedimensional structure of a protein determines its biological function and role in cellular processes. Accurate prediction of protein structures can significantly contribute to applications such as drug discovery, disease analysis, and protein engineering. In this project, a hybrid framework combining Bidirectional Long Short-Term Memory (Bi-LSTM) and Graph Neural Networks (GNN) has been proposed to improve the accuracy and effectiveness of protein structure prediction. The proposed system focuses on capturing both sequential and structural relationships between amino acids. Protein sequences contain complex dependencies where the position and interaction of residues influence the folding process. The Bi-LSTM model effectively learns these sequential patterns by processing amino acid sequences in both forward and backward directions. This bidirectional processing allows the model to capture long-range dependencies and contextual information within the protein sequence, leading to better feature representation. In addition to sequence modeling, the Graph Neural Network component is used to represent proteins as graphs, where amino acid residues are treated as nodes and their interactions are represented as edges. This graph-based representation enables the model to capture spatial relationships between residues that may be distant in the sequence but close in the three-dimensional structure. Through message passing and feature aggregation, the GNN learns important structural interactions that contribute to protein folding and stability. The integration of Bi-LSTM and GNN creates a powerful framework that combines the strengths of sequential learning and structural modeling. Experimental evaluation demonstrates that the proposed model achieves improved performance compared to traditional deep learning models such as CNN and standard LSTM networks. Evaluation metrics including accuracy, RMSE, MAE, and R^2 score indicate that the hybrid architecture provides better prediction accuracy and reduced error rates. Another advantage of the proposed framework is its ability to generalize across different protein sequences. The model effectively learns

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

meaningful representations from biological data, enabling it to predict structural patterns even for proteins that were not included in the training dataset. This capability highlights the potential of deep learning techniques in addressing complex biological problems. Despite these promising results, protein structure prediction remains a challenging task due to the complexity of protein folding and the influence of various biochemical factors. Future improvements may include the integration of additional biological features such as evolutionary information, attention mechanisms, or transformer-based architectures to further enhance prediction accuracy. In conclusion, the proposed Bi-LSTM and GNN-based protein structure prediction framework provides an effective approach for modeling both sequence-based and structural relationships in proteins. By combining deep learning with graph-based representation techniques, the system improves prediction accuracy and offers valuable insights into protein folding mechanisms. This research contributes to the advancement of computational methods in bioinformatics and supports future developments in molecular biology and biomedical research.

REFERENCES

[1] J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," Nature, 2021.
M. Baek et al., "Accurate prediction of protein structures using a three-track neural network," Science, 2021.
[2] A. W. Senior et al., "Improved protein structure prediction using potentials from deep learning," Nature, 2020.
[3] J. Yang et al., "Improved protein structure prediction using predicted interresidue orientations," PNAS, 2020.
[4] J. Xu, "Distance-based protein folding powered by deep learning," PNAS, 2019.
[5] M. AlQuraishi, "End-to-End Differentiable Learning of Protein Structure," Cell Systems, 2019.
[6] K. Tunyasuvunakool et al., "Highly accurate protein structure prediction for the human proteome," Nature, 2021.
[7] L. Zhang et al., "SPOT-1D: Improved protein secondary structure prediction," Bioinformatics, 2019.
[8] Z .Wang et al., "Protein secondary structure prediction using deep convolutional neural fields," Sci. Rep., 2016.
[9] R. Torrisi et al., "Deep learning methods in protein structure prediction," Comput. Struct. Biotechnol. J., 2020.
[10] H. Lin et al., "Deep learning for protein secondary structure prediction," Briefings in Bioinformatics, 2019.
[11] T. Hou et al., "DeepSF: Deep CNN for protein fold recognition," Bioinformatics, 2018.
[12] C. Zheng et al., "Deep learning contact-map guided protein structure prediction," Nat. Commun., 2019.
[13] J. Moult et al., "Critical assessment of protein structure prediction (CASP)," Proteins, multiple years.
[14] D. Baker, "Anfinsen's dogma and protein folding," Nature, 2019.
[15] S. Ovchinnikov et al., "Protein structure determination using metagenome sequence data," Science, 2017.
[16] A. Pollastri & P. Baldi, "Prediction of contact maps by recurrent neural networks," Bioinformatics, 2002.
[17] S. Wang et al., "RaptorX: Protein structure prediction server," Nucleic Acids Res., 2016.
[18] H. Li et al., "Deep residual neural networks for contact prediction," Bioinformatics, 2019.
[19] J. Hanson et al., "Improving protein secondary structure prediction using deep CNN and LSTM," Bioinformatics, 2018.
[20] K. Heffernan et al., "Capturing non-local interactions by long short-term memory bidirectional RNN," Bioinformatics, 2017.
[21] M. Kandathil et al., "Recent developments in deep learning applied to protein structure prediction," Proteins, 2019.
[22] I. Drori et al., "Accurate protein structure prediction by embeddings and deep learning," Sci. Rep., 2020.
[23] J. Baek & D. Baker, "Deep learning and protein structure modeling," Annu. Rev. Biophys., 2021.
[24] J. Pereira et al., "High-accuracy protein structure prediction in CASP," Proteins, 2021.
[25] S. Mirdita et al., "ColabFold: Accessible protein folding," Nat. Methods, 2022.
[26] E. Callaway, "What AlphaFold means for biology," Nature, 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com