Computational Biology and Chemistry with AI and ML

Naga Ramesh Palakurti

Principal Solutions Architect, Business Delivery, PA,

USA 0009-0009-9500-1869

DOI:10.37648/ijrmst.v17i01.006

[1]*Received: 12 March 2024; Accepted: 29 April 2024; Published:05 May 2024*

ABSTRACT

Deep learning, a transformative force in computational biology, has reshaped biological data analysis and interpretation terrain. This review delves into the multifaceted role of deep knowledge in this field, exploring its historical roots, inherent advantages, and persistent challenges. It investigates explicitly its application in two pivotal domains: DNA sequence classification, where it has been used to identify disease-causing mutations, and protein structure prediction from sequence data, where it has enabled the accurate determination of protein tertiary structures. Moreover, it offers a glimpse into the future trajectory of this dynamic field, sparking intrigue and excitement about the potential of deep learning.

Deep learning, a powerful tool in computational biology, can be traced back to the inception of 'threshold logic,' a fusion of algorithmic principles and mathematical frameworks devised to mimic cognitive processes. This seminal breakthrough, which introduced the concept of a threshold function that could be used to model complex decision-making processes, paved the way for deep learning to decipher intricate patterns embedded within vast biological datasets. Deep learning presents many benefits, including precise disease diagnosis, novel drug discovery, and personalized medicine facilitation. Its prowess lies in handling vast, intricate datasets while enhancing generalization capabilities, a testament to its evolution and development that we can appreciate.

INTRODUCTION

Role of Deep Learning in Computational Biology

The intersection of advanced computing techniques and biological research has propelled the field of computational biology forward, allowing for deeper insights into the complexities of living systems. This synthesis necessitates applying data analysis, theoretical frameworks, mathematical modelling, and computational simulations to explore phenomena ranging from behaviour and social interaction to intricate biological processes (1).

The genesis of machine learning in this domain can be traced back to the 1990s when neural network applications were first employed to analyze gene expression data. However, significant strides were made with the advent of deep learning algorithms, particularly artificial neural networks like convolutional (CNNs) and recurrent (RNNs) networks, designed to discern intricate patterns within complex datasets and furnish reliable predictions (2).

A monumental breakthrough in this field was the inception of the DeepBind algorithm in 2015 by researchers from Harvard and MIT. This cutting-edge neural network model, DeepBind, was able to uncover RNA-binding protein binding sites, revealing previously undiscovered regulatory elements within the genome (3). This heralded a new era, where deep learning has emerged as an indispensable tool for addressing a wide array of biological challenges, from predicting protein structures to identifying disease-associated genetic mutations (2).

AlphaFold, a product of DeepMind's expertise, is a prime example of the practical applications of deep learning in computational biology. This advanced neural network provides remarkably accurate insights into protein structures, thereby pushing the boundaries of structural biology. The use of deep learning in computational biology has yielded significant breakthroughs across diverse domains such as genomics, medical diagnostics, and drug discovery.

---

[1] *How to cite the article:* Palakurti N.R (May 2024); Computational Biology and Chemistry with AI and ML; *International Journal of Research in Medical Sciences and Technology;* Vol 17, 29-39, DOI: http://doi.org/10.37648/ijrmst.v17i01.006

This technological advancement has proven to be a game-changer in the field of genomic data analysis. It enables the precise detection of disease-causing genetic defects, paving the way for the development of personalized treatments. Furthermore, deep learning aids in predicting functional pathways for newly discovered drugs, streamlining the process of target identification and significantly reducing the need for extensive trial-and-error experimentation.

Deep learning excels in analyzing vast and intricate datasets, detecting subtle patterns that conventional statistical methods may overlook, thus enabling a deeper understanding of biological systems. Market projections indicate a substantial growth trajectory, with the global market for deep learning in life sciences projected to reach USD 34.83 billion by 2021, driven by the pursuit of personalized medicine and the imperative to enhance drug discovery efficiency (4).

Admittedly, deep learning in computational biology is not without its challenges. The requirement for a substantial amount of training data for optimal accuracy and the intricacy of interpreting results that may deviate from traditional biological models are significant obstacles. However, these challenges are not insurmountable, and they only serve to underscore the potential for further innovation and advancement in this field.

Despite the challenges, the potential of deep learning to revolutionize computational biology and improve healthcare outcomes is undeniable. This article delves into the applications of deep learning in computational biology, explores associated challenges and prospects, and highlights notable developments that are shaping the future of personalized medicine and drug discovery.

A Concise History and Evolution of Deep Learning

The term "deep learning" was introduced to the machine learning realm by Rina Dechter in 1986 and to artificial neural networks by Igor Aizenberg and colleagues in 2000, focusing on Boolean threshold neurons (5). The roots of deep learning trace back to 1943 when Walter Pitts and Warren McCulloch, with their unwavering determination, devised a computer model inspired by human brain neural networks, pioneering "threshold logic" to simulate cognitive processes. Despite occasional setbacks during the Artificial Intelligence Winters (5), as depicted in Table 1, Warren McCulloch and Walter Pitts laid the foundation for neural networks with threshold logic, utilizing mathematical principles and algorithms (6). In 1958, Frank Rosenblatt introduced the perceptron, a two-layer neural network for pattern recognition, also proposing additional layers, practically implemented in 1975.

Table 1 An evolutionary and historical overview of deep learning is shown in the table.

1873	A. Bain	The earliest models of neural networks, called Neural Groupings, were introduced and were inspired by the Hebbian Learning Rule.
1943	McCulloch & Pitts	The MCP Model was introduced, which is considered the precursor to Artificial Neural Models.
1949	D. Hebb	Considered as the father of neural networks, he introduced the Hebbian Learning Rule, which formed the basis for modern neural networks.
1958	F. Rosenblatt	The first perceptron, which closely resembles modern perceptrons, was introduced.
1969	Minsky and Papert	Publish Perceptrons, which criticizes the perceptron and limits the potential of neural networks
1974	P. Werbos	Introduced Backpropagation
1980	T. Kohonen	Introduced Self Organizing Map
	K. Fukushima	Neocogitron was introduced, which served as inspiration for Convolutional Neural Networks.
1982	J. Hopheld	The Hopfield Network was introduced
1985	Hilton & Sejnowski	The Hopfield Network was introduced

1986	P.Smolensky M. I. Jordan	Introduced Harmonium, which is later known as Restricted Boltzmann Machine Defined and introduced Recurrent Neural Network
1990	Y. LeCun	LeNet was introduced, demonstrating the practical potential of deep neural networks.
1997	Schuster & Paliwal Hochreiter& Schmidhuber	Introduced Bidirectional Recurrent Neural Network Introduced LS'TM, solved the problem of vanishing gradient in recurrent neural networks
2006	G. Hinton	Deep Belief Networks were introduced, along with the layer-wise pretraining technique, which marked the beginning of the current deep learning era.
2009	Salakhutdinov& Hinton	Deep Boltzmann Machines were introduced.
2012	G. Hinton	Dropout, an efficient method for training neural networks, was introduced.
2012	Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton	Convolutional neural network (CNN) for Image classification
2014	Ian Goodfellow, Yoshua Bengio, and Aaron Courville	Generative adversarial network (GAN) for image generation
2020s		Deep learning is still developing and being applied to more and more activities, such as financial trading, medical diagnosis, and self-driving automobiles.

In 1980, Kunihiko Fukushima introduced Neoconitron, a hierarchical, multilayered artificial neural network beneficial for handwriting and pattern recognition tasks. By 1989, algorithms for deep neural networks existed, albeit with impractical training times, often spanning days. 1992, Juyang Weng presented Cresceptron, automating 3D object recognition in cluttered scenes.

The mid-2000s saw the dramatic rise of "deep learning" after Geoffrey Hinton and Ruslan Salakhutdinov's seminal paper demonstrated the groundbreaking efficacy of training neural networks with multiple layers gradually. In 2009, at the NIPS Workshop on Deep Learning for Speech Recognition, it was revealed that pre-training neural networks could be skipped when working with extensive datasets, a discovery that significantly reduced error rates. By 2012, artificial pattern recognition algorithms achieved a milestone, reaching human-level performance on certain tasks, and Google's deep learning algorithm garnered attention for its ability to identify cat features.

2014, Google acquired DeepMind, a UK-based AI startup, for £400 million. In 2015, Facebook integrated DeepFace, a deep-learning technology that enables automatic photo tagging and individual identification. DeepFace utilized deep networks with 120 million parameters, showcasing remarkable face recognition capabilities. Finally, in 2016, Google DeepMind's AlphaGo algorithm mastered the game of Go, defeating professional player Lee Sedol in a highly publicized tournament in Seoul.

CONCESSIONS AND LIMITATIONS OF EMPLOYING DEEP LEARNING IN COMPUTATIONAL BIOLOGY

The swift progression in genomics and imaging technologies has ushered in a deluge of molecular and cellular profiling data from diverse global sources. This surge in biological data volume and velocity challenges conventional analysis methods. The accumulation of biomedical data has garnered significant attention from industry and academia, underscoring its potential for biological and healthcare research applications. Machine learning has emerged as a prevalent and effective methodology to distil valuable insights from this vast expanse of bioinformatics data. Leveraging training data, machine learning algorithms uncover underlying patterns, construct models, and make predictions. Applying deep learning to computational biology harbours the potential to revolutionize biology and medicine, yet it also poses challenges. An overview of the use of deep learning in computational biology is illustrated in Figure 1.

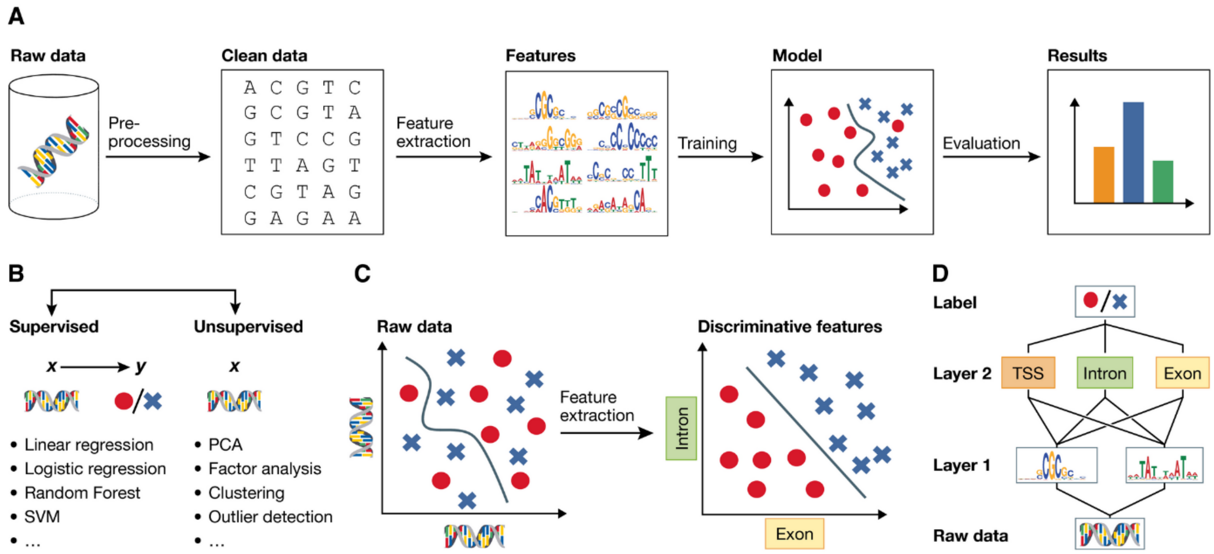

Fig 1: Computational biology using deep learning

Advantages of Deep Learning Utilization

Deep learning technology stands to catalyze disease diagnosis and prediction. It offers more meaningful, data-driven approaches for developing diagnostic tools to identify pathological samples accurately. Additionally, it streamlines the screening of large datasets and reduces costs for drug discovery applications by pinpointing drug targets and predicting drug responses. Deep learning can also enhance drug repositioning based on transcriptomic data, expediting drug discovery processes.

Moreover, deep learning holds promise for precision medicine and personalized treatment development. By integrating patient-specific data, including gene profiling, clinical records, and lifestyle information, deep learning facilitates the customization of treatments tailored to individual patients. Its capacity to analyze vast datasets with high accuracy enables the identification of genetic markers, variations, biological markers, drug efficacy, and clinical prognosis, facilitating optimal treatment responses and disease progression identification.

Furthermore, deep learning models excel at handling large and complex datasets. They can analyze various biological or genomic data types, extracting intricate patterns and learning the most relevant features. This capability reduces the need for human intervention, which is particularly beneficial in fields like biomedicine and molecular biology, which entail navigating complex and heterogeneous data.

The scalability and transferability of deep learning models enhance their ability to handle large volumes of complex data. These models can be trained incrementally, reducing resource requirements and improving generalization to new data. Additionally, deep learning methods can identify novel patterns or relationships within data, surpassing conventional analytical approaches.

Challenges of Deep Learning Utilization

Interpretability remains a significant challenge when employing deep learning in computational biology. Deep learning models' complexity and internal representation often render them "black boxes," hindering understanding of how they predict biological mechanisms. Efforts are needed to develop interpretability techniques that elucidate the decision-making process of deep learning models, fostering trust among stakeholders.

Ethical and regulatory concerns arise regarding using deep learning in the medical field. These include patient privacy protection and the potential for biased diagnostic reports, which could lead to incorrect diagnoses or denial of treatment access. Transparent and responsible usage of deep learning models is essential to ensure accountability in biomedical research and healthcare.

Additionally, acquiring high-quality datasets poses a challenge, particularly in healthcare and biomedicine, where data availability may be limited due to privacy regulations and data heterogeneity. Integrating data from multiple sources with varying formats and standards further complicates the performance and generalization of deep learning models.

Moreover, training and utilizing deep learning models require extensive computational resources and infrastructure, which may be inaccessible to small non-profit organizations or research institutions with limited resources. Addressing these challenges is crucial to realizing the full potential of deep learning in computational biology.

The application of deep learning models has emerged as a powerful tool for analyzing and classifying DNA sequences, enabling scientists to make precise predictions regarding the structure and function of genetic material. The intricate patterns inherent in genomic data can be effectively scrutinized using artificial neural network-based algorithms, a hallmark of deep learning technology. This makes deep learning particularly well-suited for identifying crucial genetic attributes and accurately predicting the health implications associated with hereditary variations that may lead to diseases.

Role of Deep Learning in Genomic Variant Detection and Precision Medicine

Deep learning, a subset of machine learning, has not only revolutionized but transformed genomic variant detection and gene expression analysis. This transformative power of deep neural networks in understanding gene expression patterns and genetic variations has led to invaluable insights into personalized medicine, drug discovery, and disease mechanisms. The potential of deep learning to categorize variants, including disease-causing mutations, interpret splicing codes and identify long noncoding RNAs is truly inspiring.

The use of deep learning in genomic variant detection allows for predicting the organization and function of various genomic elements, such as promoters and enhancers, as well as gene expression levels. The employment of fragmentation and windowing techniques to divide the genome into optimal, non-overlapping fragments, facilitates the deep learning-based identification of genetic variations. Deep learning models' exceptional precision and accuracy in distinguishing between patients and controls and identifying individuals with multiple disorders instils confidence in their reliability. Moreover, these models have revealed genomic regions enriched with biological pathways relevant to immune responses and other vital functions, offering unparalleled insights into disease mechanisms.

Additionally, deep learning methods, including convolutional neural networks (CNNs), have been applied to forecast genetic variations associated with diseases. Deep learning models have proven more effective than traditional methods in predicting the functional implications of genomic variations, particularly in noncoding regions. Furthermore, deep learning enables the prognosis of gene expression levels influenced by single-nucleotide polymorphisms (SNPs), surpassing traditional models and identifying new SNPs linked to gene expression peculiarities.

Enhancing Gene Expression Analysis through Deep Learning

Gene expression analysis, a critical aspect of understanding cellular function, has significantly benefited from deep learning techniques. High-throughput screening technologies provide valuable insights into gene expression regulation, but often struggle to explore large biological sequence regions. Deep learning models, such as Enformer, have improved the accuracy of gene expression prediction by integrating data from long-range interactions within the genome. These models enable predictions of saturation mutagenesis and natural genetic variants, thereby enhancing our understanding of gene expression.

Deep generative models (DGMs) have also been employed for gene expression analysis, identifying underlying structures from omics data and facilitating joint analyses across multiple loci to understand multigenic diseases. Deep learning approaches with DGMs on high-dimensional SNP data enable predictions of how nucleotide changes affect DNA levels, expanding our understanding beyond genetic expression datasets alone.The integration of deep learning into genomic variant detection and gene expression analysis has not just advanced, but accelerated our understanding of genetics, paving the way for practical applications in personalized medical treatments. By accurately identifying genetic variations and interpreting gene expressions, deep learning accelerates the discovery of disease-associated genes, drug targets, and potential therapies. Despite challenges such as overfitting and interpretability, the superiority of deep-learning-derived techniques in certain contexts underscores their significance in genomics research. With numerous pipelines available, the utility of deep learning in genomics continues to expand, promising a brighter future in precision medicine.

Protein Structure Prediction using Deep Learning

The field of protein structure prediction has undergone a profound transformation by integrating deep learning techniques. Deep learning methods have revolutionized our ability to predict the three-dimensional structures of proteins with precision and efficiency, thus playing a pivotal role in understanding protein function, drug discovery, and therapeutic design. Deep learning models have achieved remarkable success in predicting protein structures by effectively capturing intricate patterns and dependencies within protein sequences.

Predicting protein structures solely from their sequences presents a significant challenge, which deep learning effectively addresses. Recent applications have successfully predicted both three-state and eight-state secondary structures in proteins, bridging the gap between primary sequence and tertiary structure predictions. The prediction of eight-state secondary structures, known as the Q8 problem, offers enhanced structural information for various applications. Deep learning techniques such as the SC-GSN network, bidirectional long short-term memory (BLSTM) approach, conditional neural field with multiple layers, and DCRNN have been employed to tackle this complex task. Additionally, advanced methods like next-step conditioned convolutional neural networks (CNNs) have been utilized to identify sequence motifs associated with specific secondary structure elements, as demonstrated by AlQuraishi's "Alphafold" model.

Deep learning also plays a crucial role in forecasting protein-protein interactions and binding sites. Models like "DeepPPI", developed by Xiuquan Du, leverage sequence information to accurately predict interactions among different proteins. DeepPPI enhances our understanding of protein assembly and function by extracting critical elements from protein sequences and structures.

Deep learning has made significant strides in tertiary structure prediction. Approaches such as deep learning contact-map methods have achieved breakthroughs in accurately predicting protein folding, as seen in the CASP13 competition. These methods leverage vast protein databases to capture intricate residue-residue dependencies, aiding in the prediction of long-range contacts and facilitating precise folding predictions. Models like "AlphaFold 2" integrate recurrent neural networks (RNNs) and attention mechanisms to achieve extraordinary precision in predicting protein tertiary structures, surpassing other techniques in competitions like CASP.

Overall, deep learning applications in protein structure prediction demonstrate its vast potential and impact. Deep learning drives significant progress in understanding protein folding, function, and interactions by uncovering complex patterns and connections within protein sequences and structures. Despite future challenges and opportunities, profound learning promises to reveal fresh insights into basic life processes, personalized medicine, and drug discovery.

PREDICTION OF PROTEIN STRUCTURE FROM DEEP LEARNING USING SEQUENCE DATA

Deep learning, a subset of machine learning, harnesses neural networks with multiple layers of processing to iteratively derive insights (41). These techniques have undergone significant advancements. Initially employed by numerous researchers in computational biology, particularly for tasks like protein prediction and interaction analysis, machine learning methods paved the way for adopting deep learning algorithms. These algorithms excel in managing vast, complex datasets while learning abstract functions. In bioinformatics, deep learning has been instrumental in data augmentation efforts (42,43). The architecture of deep learning models is not only flexible but also capable of accommodating diverse data sources such as protein 3D structures, network topologies, domain components, primary sequences, and textual data for downstream analyses. This adaptability instills confidence in the potential applications of deep learning in computational biology and bioinformatics. Moreover, essential neural network modules, including fully connected, convolutional, and recurrent layers, play crucial roles in deep learning frameworks (44,45).

Protein interaction prediction using deep learning

The latest advancements in deep learning techniques for Protein-Protein Interaction (PPI) models, such as Deep Convolutional Neural Networks (CNNs), are showing great promise. Torrisi et al. (44) have demonstrated the efficacy of using structural network information alongside sequence-based features to predict protein interactions. The adaptability of image-related methodologies to protein structures is also being explored, as seen in the use of the pre-trained ResNet50 model to extract structural information from 2D volumetric protein representations (45). However, it's important to note that these techniques, while promising, do face challenges such as high computational costs and interpretability issues.

Various deep-learning methods can be applied to protein-protein interaction networks. DeepPPI, for instance, utilizes a multilayer perceptron learning structure with protein sequences as input features. It employs concatenation for combining seven sequence-based features, which are specific characteristics of the protein sequences that are known to influence protein interactions (46). DPPI, another approach, uses a convolutional neural network architecture with protein sequences as input, utilizing protein-positioning specific scoring matrices (PSSM) derived from PSI-BLAST for encoding (47). DeePFE-PPI, introduced in 2019, employs a multilayer perceptron with protein sequences as input, utilizing pre-trained model embeddings (Word2vec) for encoding (48). S-VGAE utilizes graph convolutional neural networks with protein sequences and topology information, employing a conjoint encoding method combined through concatenation (49).

Deep learning is not just a theoretical concept, but a practical tool in drug discovery. It is being used to optimize drug properties, identify new drugs, and predict drug-target interactions. It's also playing a role in predicting molecular properties such as solubility, bioactivity, and toxicity, and even generating novel molecules with desired properties. In QSAR studies, deep neural networks are being used to predict drug bioactivity and chemical structures (50).

Deep learning methods also enhance traditional in silico drug discovery efforts. For instance, AtomNet by Atomwise utilizes convolutional neural networks to predict molecular bioactivity in proteins, marking a significant application of deep learning in Drug-Target Interaction (DTI) prediction (51). In docking studies, deep learning techniques improve the accuracy of traditional docking modules, which are computational tools used to predict the binding of a small molecule to a target protein, and scoring functions, which are algorithms used to evaluate the quality of a predicted binding pose. This enhancement enhances binding mode prediction accuracy and demonstrates successful integration into rational docking processes (52).

Recent advancements in deep learning for analyzing protein function and evolution, a collaborative effort of biologists, bioinformaticians, and researchers, have showcased notable progress. One prominent application involves integrating deep learning algorithms into protein analysis methodologies, such as combining them with homology modelling. Homology modelling, a widely used method for predicting protein structures, relies on principles that leverage amino acid sequences and preserve 3D structures concerning the primary structure.

Despite its popularity, homology modelling presents challenges like dealing with weak sequence structures and modelling rigid body shifts. However, the integration of deep learning models has significantly enhanced the accuracy of protein structure prediction, inspiring a new era of research. These deep-learning techniques are crucial in refining each step of template-based protein modelling. For instance, DLPAlign, a deep learning-based approach combined with sequence alignment, enhances the accuracy of progressive multiple sequence analysis by employing convolutional neural networks.

Additionally, methods like DESTINI employ deep learning techniques for predicting protein residues and residue contacts alongside template-based structure modelling. Deep learning techniques have particularly excelled in quality assessment (QA), a pivotal stage following structure predictions aimed at quantifying deviations from native protein structures, be it in template-based or template-free techniques.

However, despite these advancements, challenges persist in utilizing deep learning for biological analyses, especially in the complex task of protein structure prediction. One significant challenge lies in the demand for substantial amounts of high-quality data, limiting the scope of biological analyses to the availability of such data. Moreover, the single-task nature of deep learning models poses limitations, as they can only address one issue at a time.

Interpretability remains another pressing challenge, as understanding how deep learning models arrive at their predictions proves challenging. However, the future trajectory of deep learning in protein analysis is promising. Researchers are actively developing new techniques to enhance interpretability. This involves creating hybrid models by integrating other machine learning techniques aimed at improving both performance and interpretability.

RECENT ADVANCES IN DEEP LEARNING FOR COMPUTATIONAL BIOLOGY

Deep learning applications in protein-protein interaction Forecasting and pharmaceutical discovery

The cutting-edge deep learning techniques utilized in protein-protein interaction (PPI) models, such as deep convolutional neural networks (CNNs), are a source of fascination and excitement. These techniques, renowned for their ability to extract features from structural data, are pushing the boundaries of our understanding. For instance, as Torrisi et al. (2019) have shown, the integration of structural network information with sequence-based features is revolutionizing the prediction of protein interactions. Moreover, the use of the pre-trained ResNet50 model to extract structural information from 2D volumetric representations of proteins is a testament to the adaptability of image-related methodologies for protein structures. However, it's important to note that these techniques, while innovative, do have their limitations, such as high computational costs and limited interpretability.

Several distinct deep learning methods are employed in the context of PPI networks, each with its unique approach and strengths. One such method is DeepPPI, a multilayer perceptron that leverages protein sequences as input features, combining seven sequence-based features through concatenation. Another method, DPPI, is a CNN that also uses protein sequences as input features, with protein-positioning specific scoring matrices (PSSM) derived from PSI-BLAST as the encoding method. DeePFE-PPI, introduced in 2019, utilizes a multilayer perceptron with protein sequences as input and pre-trained model embeddings (Word2vec) for encoding. S-VGAE, on the other hand, is an example of a graph convolutional neural network that utilizes protein sequences and the topology information of PPI networks, combining features via a conjoint method and concatenation. These methods, each with their unique approach, contribute to the diverse landscape of deep learning in PPI networks.

Deep learning is not just a theoretical concept, but a practical tool extensively applied in drug discovery. It's optimizing drug properties, identifying new drugs, and predicting drug-target interactions. Deep learning is also predicting crucial molecular properties of drugs, such as solubility, bioactivity, and toxicity, and even generating novel molecules with desired properties. In the realm of quantitative structure-activity relationship (QSAR) studies, deep neural networks are predicting the bioactivity of drugs based on their chemical structures, providing tangible and valuable insights.

Deep learning is not a replacement for traditional drug discovery methods, but a powerful enhancement. For instance, AtomNet, a groundbreaking development by Atomwise, represents the first significant application of deep learning in drug-target interaction (DTI) prediction. By using CNNs, AtomNet predicts molecular bioactivity in proteins with unprecedented accuracy. Additionally, deep learning techniques have significantly improved the accuracy of docking modules and scoring functions, thereby enhancing binding mode prediction accuracy. These advancements underscore the reliability and potential of deep learning in rational docking processes, providing reassurance about the future of drug discovery.

New advancements in the analysis of protein function and evolution using deep learning methods.

Recent developments in protein analysis have incorporated deep learning algorithms to enhance accuracy and efficiency. One notable example is the integration of deep learning with homology modelling. Homology modelling, the most widely used protein structure prediction method, generates the 3D structure of a protein based on amino acid sequences and the preservation of 3D structures in homologous proteins with similar sequences. This approach effectively builds 3D models using known structures of identical proteins. However, challenges such as weak sequence similarities and rigid body shifts remain. Incorporating deep learning models has significantly improved the accuracy of these protein models.

Deep learning-based methods enhance accuracy in various stages of template-based protein modelling. For instance, DLPAlign combines deep learning with sequence alignment. This method leverages convolutional neural networks (CNNs) to improve the accuracy of progressive multiple sequence alignment. Another recent method, DESTINI, applies deep learning for protein residue and residue contact prediction and template-based structure modelling.

In summary, deep learning techniques have significantly improved collaborative sectors like model quality assessment (QA), a crucial stage in protein structure prediction. QA follows structure predictions to quantify deviations from native protein structures using template-based and template-free techniques.

Future Outlook and Potential Impact of Deep Learning on Biological Research and Clinical Practice

Deep learning models are revolutionizing biological research and clinical practices, offering immense potential for advancement. For instance, they can be used to analyze vast and complex genomic datasets, leading to the discovery of new disease markers. In drug discovery, deep learning algorithms can predict the efficacy of potential drugs, accelerating the development process. In personalized treatment strategies, these models can analyze patient data to recommend the most effective treatment plan. Leveraging deep learning algorithms allows researchers to delve into biological data, unraveling novel patterns and molecules that shed light on intricate biological mechanisms. Furthermore, these models accelerate drug target development, refine diagnostic procedures, and enhance clinical trial designs.

Looking ahead, the prospects for deep learning in both biological research and clinical settings are promising. Continuous development in technology and algorithms holds the promise of ushering humanity into a new era of disease diagnosis, precise treatment, and preventive care. Moreover, deep learning models offer the potential to not just improve, but optimize healthcare delivery systems by streamlining tasks and expediting diagnostic processes, instilling a sense of hope and reassurance about the future of healthcare.

Their ability to integrate and analyze diverse data types—from genomics to clinical records enables deep learning models to unveil hidden patterns and relationships, providing a wealth of comprehensive insights into diseases and guiding translational research. This comprehensive understanding leaves the audience enlightened and informed. Additionally, deep learning's versatility extends to addressing various challenges across domains such as image classification, object detection, speech recognition, and machine translation.

CONCLUSION

In conclusion, recent advancements in deep learning within computational biology not only highlight its transformative potential but also ignite a sense of optimism and excitement about the future of research and clinical practices. Table 2 provides a summary of these significant strides.

Table 2 Recent developments in applying deep learning to computational biology.

Deep Learning Algorithm	**Application in Computational Biology**	**References**
Convolutional Neural Networks (CNN)	Gene expression analysis	79
Recurrent Neural Networks (RNN)	DNA sequence analysis	80
Generative Adversarial Networks (GAN)	Synthetic biology and protein design	81
Deep Belief Networks (DBN)	Protein structure prediction	82
Reinforcement Learning (RL)	Drug discovery and optimization	83
Transformer Networks	RNA structure prediction	84
Autoencoders	Disease diagnosis and prognosis	85
Graph Neural Networks (GNN)	Protein-protein interaction prediction	86
Variational Autoencoders (VAE)	Single-cell genomics analysis	87
Variational Autoencoders (VAE)	Single-cell genomics analysis	87
Deep Reinforcement Learning	Drug target identification	88
Capsule Networks	Protein structure classification	89
Adversarial Autoencoders	Gene expression imputation	90
Deep Boltzmann Machines (DBM)	Epigenetic data analysis	91
Deep Learning Algorithm	Application in Computational Biology	92
Attention Mechanism	Single-cell RNA sequencing analysis	93
Deep Q-Networks (DQN)	Drug toxicity prediction	94
Capsule Networks	Protein-protein interaction prediction	95
Deep Generative Models	DNA sequence generation	96
Graph Convolutional Networks (GCN)	Drug-target interaction prediction	97
Deep Survival Analysis	Cancer survival prediction	98
Transformer Networks	Transcriptomics analysis	99
Graph Neural Networks (GNN)	Drug repurposing	100
Adversarial Networks	Image-based phenotypic screening	101
Deep Transfer Learning	Drug response prediction	102
Generative Adversarial Networks (GAN)	Synthetic data generation	103

Deep Reinforcement Learning	Protein folding	104
Variational Graph Autoencoders (VGAE)	Disease-gene prioritization	105
Deep Neural Networks (DNN)	Metagenomic analysis	106
Convolutional Recurrent Neural Networks (CRNN)	Chromatin state prediction	107
Deep Clustering	Cell type identification	108
Deep Reinforcement Learning	Protein-ligand binding affinity prediction	109
Graph Convolutional Networks (GCN)	Drug response prediction	110
Long Short-Term Memory (LSTM)	RNA splicing prediction	111
Deep Reinforcement Learning	Antibiotic resistance prediction	112
Capsule Networks	Protein function prediction	113
Autoencoders	Single-cell epigenomics analysis	114
Deep Belief Networks (DBN)	Genetic variant classification	115
Transformer Networks	Protein-protein interaction network analysis	116
Graph Convolutional Networks (GCN)	Drug-target interaction network analysis	117
Recurrent Neural Networks (RNN)	Protein secondary structure prediction	118
Deep Reinforcement Learning	Gene regulatory network inference	119
Variational Autoencoders (VAE)	Metabolomics data analysis	120
Deep Belief Networks (DBN)	Drug side effect prediction	121
Capsule Networks	Cancer subtype classification	122
Convolutional Neural Networks (CNN)	Histopathology image analysis	123
Generative Adversarial Networks (GAN)	Synthetic biology and gene synthesis	124
Transformer Networks	Protein contact prediction	125
Deep Reinforcement Learning	Genome sequence assembly	126
Graph Neural Networks (GNN)	Cell type classification in single-cell transcriptomics	127
Autoencoders	DNA motif discovery	128

REFERENCES

1. Way, G. P., Greene, C. S., Carninci, P., Carvalho, B. S., de Hoon, M., Finley, S. D., Gosline, S. J. C., Lê Cao, K.-A., Lee, J. S. H., Marchionni, L., Robine, N., Sindi, S. S., Theis, F. J., Yang, J. Y. H., Carpenter, A. E., & Fertig, E. J. (2021). A field guide to cultivating computational biology. PLoS Biology, 19(10), e3001419. https://doi.org/10.1371/journal.pbio.3001419
2. Faustine A, Pereira L, Bousbiat H, Kulkarni S. UNet-NILM: A Deep Neural Network for Multi-tasks Appliances State Detection and Power Estimation in NILM; UNet-NILM: A Deep Neural Network for Multi-tasks Appliances State Detection and Power Estimation in NILM. 2020 [cited 2023 May 9]; Available from: https://doi.org/10.1145/3427771.3427859
3. Kindel WF, Christensen ED, Zylberberg J. Using deep learning to reveal the neural code for images in primary visual cortex.

4. Machine Learning Industry Trends Report Data Book, 2022-2030 [Internet]. [cited 2023 May 9]. Available from: https://www.grandviewresearch.com/sector-report/machine-learning-industry-data-book
   Wang, H., & Raj, B. (2017, March 2). On the Origin of Deep Learning. ArXiv.org. https://doi.org/10.48550/arXiv.1702.07800
5. Wang H, Raj B. On the origin of deep learning. arXiv preprint arXiv:1702.07800. 2017 Feb 24.
6. Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2018). A primer on deep learning in genomics. Nature Genetics, 51(1), 12–18. https://doi.org/10.1038/s41588-018-0295-5
7. Liu, J., Li, J., Wang, H., & Yan, J. (2020). Application of deep learning in genomics. Science China. Life Sciences, 63(12), 1860–1878. https://doi.org/10.1007/s11427-020-1804-5
8. Jo, T., Nho, K., Bice, P., & Saykin, A. J. (2022). Deep learning-based identification of genetic variants: application to Alzheimer's disease classification. Briefings in Bioinformatics, 23(2). https://doi.org/10.1093/bib/bbac022
9. Liu, Y., Qu, H.-Q., Mentch, F. D., Qu, J., Chang, X., Nguyen, K., Tian, L., Glessner, J., Sleiman, P. M. A., & Hakonarson, H. (2022). Application of deep learning algorithm on whole genome sequencing data uncovers structural variants associated with multiple mental disorders in African American patients. Molecular Psychiatry. https://doi.org/10.1038/s41380-021-01418-1
10. Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W. R., Fajardo-Flores, S. B., Gaytan-Lugo, L. S., Santana-Mancilla, P. C., & Crossa, J. (2021). A review of deep learning applications for genomic selection. BMC Genomics, 22(1). https://doi.org/10.1186/s12864-020-07319-x
11. Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., Assael, Y., Jumper, J., Kohli, P., & Kelley, D. R. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. Nature Methods, 18(10), 1196–1203. https://doi.org/10.1038/s41592-021-01252-x
12. Treppner, M., Binder, H., & Hess, M. (2022). Interpretable generative deep learning: an illustration with single cell gene expression data. Human Genetics. https://doi.org/10.1007/s00439-021-02417-6
13. van Dijk, S. J., Tellam, R. L., Morrison, J. L., Muhlhausler, B. S., & Molloy, P. L. (2015). Recent developments on the role of epigenetics in obesity and metabolic disease. Clinical Epigenetics, 7(1). https://doi.org/10.1186/s13148-015-0101-5
14. Angermueller, C., Lee, H. J., Reik, W., & Stegle, O. (2017). DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. Genome Biology, 18(1). https://doi.org/10.1186/s13059-017-1189-z
15. Yin, Q., Wu, M., Liu, Q., Lv, H., & Jiang, R. (2019). DeepHistone: a deep learning approach to predicting histone modifications. BMC Genomics, 20(S2). https://doi.org/10.1186/s12864-019-5489-4
16. Galkin, F., Mamoshina, P., Kochetov, K., Sidorenko, D., & Zhavoronkov, A. (2021). DeepMAge: A Methylation Aging Clock Developed with Deep Learning. Aging and Disease, 12(5), 1252. https://doi.org/10.14336/ad.2020.1202
17. Zhang, B., Li, J., & Lü, Q. (2018). Prediction of 8-state protein secondary structures by a novel deep learning architecture. BMC Bioinformatics, 19(1). https://doi.org/10.1186/s12859-018-2280-5
18. AlQuraishi, M. (2019). AlphaFold at CASP13. Bioinformatics, 35(22), 4862–4865. https://doi.org/10.1093/bioinformatics/btz422