Computational Molecular Biology 2024, Vol.14, No.2, 76-83
http://bioscipublisher.com/index.php/cmb

Review and Progress Open Access

The Application and Progress of Deep Learning in Bioinformatics

Haimei Wang
Hainan Institute of Biotechnology, Haikou, 570206, Hainan, China
Corresponding email: haimei wang@hitar.org

Computational Molecular Biology, 2024, Vol.14, No.2 doi: 10.5376/cmb.2024.14.0009
Received: 17 Feb., 2024
Accepted: 29 Mar., 2024
Published: 16 Apr., 2024

Preferred citation for this article:
Wang H.M., 2024, The application and progress of deep learning in bioinformatics, Computational Molecular Biology, 14(2): 76-83 (doi: 10.5376/cmb.2024.14.0009)

Abstract As biological data explosively grows and traditional computational methods struggle to keep pace, deep learning has become a powerful tool for analyzing complex biological data, significantly improving the ability to mine and interpret large-scale biological data, including images, signals, and sequences. This study reviews successful applications of deep learning in key areas such as genomics, proteomics, and drug discovery, and the results show that deep learning models outperform traditional methods in tasks such as gene expression prediction and protein structure modeling. Deep learning offers great potential for advancing bioinformatics research to analyze biological data more accurately and efficiently, but many challenges remain, and future research should focus on addressing identified challenges and exploring new applications of deep learning in bioinformatics to fully realize its potential.

Keywords Deep learning; Bioinformatics; Neural networks; Data mining; Biomedical data

1 Introduction

Deep learning, a subset of machine learning, has revolutionized various fields by enabling the automatic extraction of high-level features from raw data through multiple layers of processing units (LeCun et al., 2015; Goh et al., 2017). This approach, known as representation learning, allows models to learn intricate patterns and hierarchies within the data without explicit guidance from domain experts (Berrar and Dubitzky, 2021). The advent of deep learning has been particularly transformative in the era of big data, where the availability of vast amounts of data has facilitated the training of complex models, leading to significant advancements in predictive accuracy and efficiency (Talukder et al., 2020). Key architectures in deep learning include deep neural networks (DNNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and more recent innovations such as graph neural networks (GNNs) and generative adversarial networks (GANs) (Liu et al., 2017; Li et al., 2019; Muzio et al., 2020).

Bioinformatics, an interdisciplinary field that combines biology, computer science, and information technology, plays a crucial role in modern biology by enabling the analysis and interpretation of complex biological data (Gauthier et al., 2019). The rise of high-throughput technologies has led to an explosion of data in genomics, proteomics, and other omics fields, necessitating advanced computational tools to manage and analyze this information (Lan et al., 2018; Berrar and Dubitzky, 2021). Bioinformatics facilitates the understanding of biological processes at a molecular level, aiding in the discovery of new biomarkers, the prediction of protein structures, and the elucidation of gene regulatory networks, among other applications (Min et al., 2016; Muzio et al., 2020). The integration of deep learning into bioinformatics has further enhanced the ability to decode complex biological interactions and predict outcomes with high accuracy, thereby driving forward research in systems biology, biomedical imaging, and drug discovery (Ravì et al., 2017).

This study reviews the applications and progress of deep learning in the field of bioinformatics, including various deep learning architectures and their specific applications in different bioinformatics fields, such as sequence analysis, structure prediction, and prediction of biomolecule interactions; And analyze the challenges and limitations of using deep learning in bioinformatics, such as overfitting and interpretability issues, propose potential future research directions, in order to provide valuable research for researchers to apply deep learning techniques in bioinformatics research.

2 Fundamental Concepts of Deep Learning

2.1 Neural networks and their structures

Neural networks are the backbone of deep learning, consisting of interconnected layers of nodes or neurons. These networks are designed to simulate the way the human brain processes information. The primary types of neural networks used in deep learning include multilayer feed-forward perceptrons, convolutional neural networks (CNNs), and recurrent neural networks (RNNs) (Min et al., 2016). Each type of network has a unique structure and is suited to different types of data and tasks. For instance, CNNs are particularly effective for image data due to their ability to capture spatial hierarchies, while RNNs are well-suited for sequential data such as time series or natural language (Li et al., 2019; Berrar and Dubitzky, 2021).

2.2 Key algorithms and models

Deep learning leverages a variety of algorithms and models to process and analyze data. Some of the most prominent models include deep neural networks (DNNs), CNNs, RNNs, and more advanced architectures like generative adversarial networks (GANs) and variational autoencoders (VAEs) (Li et al., 2019; Berrar and Dubitzky, 2021). These models are trained using optimization algorithms such as stochastic gradient descent (SGD) and its variants, which adjust the weights of the network to minimize the error in predictions (Shrestha and Mahmood, 2019). The choice of model and algorithm depends on the specific application and the nature of the data being analyzed (Min et al., 2016; Li et al., 2019).

2.3 Advantages of deep learning over traditional methods

Deep learning offers several advantages over traditional machine learning methods. One of the key benefits is its ability to automatically learn and extract features from raw data, eliminating the need for manual feature engineering (Shrestha and Mahmood, 2019). This capability is particularly valuable in bioinformatics, where the data is often high-dimensional and complex (Mamoshina et al., 2016). Additionally, deep learning models can capture intricate patterns and relationships within the data, leading to improved accuracy and performance in tasks such as classification, prediction, and clustering (Min et al., 2016). The scalability of deep learning models also allows them to handle large datasets effectively, making them well-suited for the big data era in bioinformatics (Mamoshina et al., 2016; Li et al., 2019; Berrar and Dubitzky, 2021). By leveraging these fundamental concepts, deep learning has the potential to revolutionize bioinformatics, providing deeper insights and more accurate predictions than traditional methods.

3 Applications of Deep Learning in Bioinformatics

3.1 Genomics and sequence analysis

3.1.1 Deep learning for genome annotation

Deep learning has significantly advanced the field of genome annotation by enabling the identification of complex patterns within large genomic datasets (Zhang et al., 2020). These methods have been particularly effective in annotating functional elements of the genome, such as regulatory regions and non-coding RNAs. For instance, deep learning models have been employed to predict the functional impact of genetic variants and to annotate sequence elements with high accuracy (Libbrecht and Noble, 2015; Zou et al., 2018; Routhier and Mozziconacci, 2022). The ability of deep learning to handle vast amounts of data and to learn intricate patterns has made it a valuable tool in genome annotation, surpassing traditional methods in both accuracy and efficiency (Koumakis, 2020; Talukder et al., 2020).

3.1.2 Variant calling and mutation detection

Variant calling and mutation detection are critical tasks in genomics, where deep learning has shown remarkable promise. By leveraging high-throughput sequencing data, deep learning models can accurately identify genetic variants and mutations, which are essential for understanding genetic diseases and developing personalized medicine approaches. These models have been integrated into pipelines for next-generation sequencing (NGS) data analysis, providing robust and scalable solutions for variant calling (Zou et al., 2018; Li et al., 2019). The application of deep learning in this area has led to improved detection rates and reduced false positives, making it a preferred choice for genomic researchers (Schmidt and Hildebrandt, 2020).

3.1.3 Integration with high-throughput sequencing data

The integration of deep learning with high-throughput sequencing data has revolutionized the analysis of genomic information. Deep learning frameworks can process and analyze large-scale sequencing data, enabling the discovery of novel genomic features and the interpretation of complex biological processes. For example, deep learning models have been used to integrate multi-omics data, such as genomics, epigenomics, and transcriptomics, to provide a comprehensive understanding of cellular mechanisms (Min et al., 2016; Talukder et al., 2020). This integration has facilitated advancements in precision medicine by allowing for the detailed analysis of individual genetic profiles and the identification of potential therapeutic targets (Cao et al., 2020; Koumakis, 2020).

3.2 Protein structure prediction

Deep learning has also made significant strides in the prediction of protein structures, a fundamental challenge in bioinformatics. By utilizing deep neural networks, researchers have been able to predict the three-dimensional structures of proteins with unprecedented accuracy. These models can learn from vast amounts of protein sequence and structure data, enabling the prediction of protein folding patterns and interactions. The success of deep learning in this domain has been exemplified by models such as AlphaFold, which have achieved remarkable performance in protein structure prediction competitions (Libbrecht and Noble, 2015; Min et al., 2016). The ability to accurately predict protein structures has profound implications for understanding biological functions and designing new drugs.

3.3 Drug discovery and design

In the realm of drug discovery and design, deep learning has emerged as a powerful tool for identifying potential drug candidates and optimizing their properties. Deep learning models can analyze large datasets of chemical compounds and biological targets to predict the efficacy and safety of new drugs. These models have been used to screen vast libraries of compounds, identify promising drug candidates, and optimize their chemical structures for better performance (Min et al., 2016; Cao et al., 2020). The integration of deep learning with bioinformatics has accelerated the drug discovery process, reducing the time and cost associated with developing new therapeutics and enabling the discovery of novel treatments for various diseases (Li et al., 2019; Wang et al., 2023).

4 Challenges in Applying Deep Learning to Bioinformatics

4.1 Data quality and availability

4.1.1 Dealing with noisy and incomplete data

One of the primary challenges in applying deep learning to bioinformatics is the presence of noisy and incomplete data. Biological datasets often contain errors, missing values, and inconsistencies due to the limitations of experimental techniques and the complexity of biological systems. These issues can significantly affect the performance of deep learning models, leading to inaccurate predictions and unreliable results. Strategies to mitigate these problems include data cleaning, imputation techniques, and robust model training methods that can handle noise and missing data effectively (Min et al., 2016; Lan et al., 2018; Saha et al., 2023).

4.1.2 Strategies for data augmentation

Data augmentation is a crucial technique to enhance the quality and quantity of training data, especially when dealing with limited datasets. In bioinformatics, data augmentation can involve generating synthetic data, applying transformations to existing data, or using domain-specific knowledge to create new training examples. These strategies help in improving the generalization ability of deep learning models and reducing overfitting. For instance, techniques such as generative adversarial networks (GANs) and variational autoencoders (VAEs) have been employed to generate realistic biological data, thereby augmenting the training datasets (Li et al., 2019; Tang et al., 2019; Jin et al., 2020).

4.1.3 Access to large and diverse datasets

The effectiveness of deep learning models heavily relies on the availability of large and diverse datasets. However, in bioinformatics, obtaining such datasets can be challenging due to privacy concerns, data sharing restrictions, and the high cost of data generation. Collaborative efforts and open-access initiatives are essential to overcome

these barriers. Additionally, integrating data from multiple sources and different omics layers (e.g., genomics, transcriptomics, proteomics) can provide a more comprehensive understanding of biological systems and improve model performance (Li et al., 2020; Auslander et al., 2021; Saha et al., 2023).

4.2 Model interpretability and complexity

Deep learning models, particularly deep neural networks, are often criticized for their lack of interpretability. In bioinformatics, where understanding the underlying biological mechanisms is crucial, the black-box nature of these models poses a significant challenge. Researchers are actively working on developing interpretable models and techniques to explain the predictions of deep learning models. Methods such as attention mechanisms, feature importance analysis, and model-agnostic interpretation techniques are being explored to enhance the interpretability of deep learning models in bioinformatics (Li et al., 2019; Auslander et al., 2021).

4.3 Computational resources and scalability

Deep learning models require substantial computational resources for training and inference, which can be a limiting factor in bioinformatics research. The high-dimensional nature of biological data and the complexity of deep learning architectures necessitate the use of powerful hardware, such as GPUs and TPUs, and efficient algorithms to handle large-scale data. Scalability is another critical aspect, as models need to be capable of processing increasing amounts of data without a significant loss in performance. Advances in distributed computing, cloud-based platforms, and optimization techniques are being leveraged to address these challenges and make deep learning more accessible and scalable in bioinformatics (Lan et al., 2018; Cao et al., 2020; Li et al., 2020).

5 Recent Advances and Breakthroughs

5.1 Innovations in model architectures

Recent advancements in deep learning have led to the development of novel model architectures that significantly enhance the performance and applicability of bioinformatics tools. For instance, the integration of deep neural networks, convolutional neural networks, and recurrent neural networks has been pivotal in transforming biomedical big data into valuable knowledge (Min et al., 2016). Additionally, emergent architectures such as graph neural networks and generative adversarial networks (GANs) have shown promise in handling complex biological data, offering new ways to model and interpret biological systems (Li et al., 2019). Autoencoders, particularly, have become a dominant approach in single-cell RNA-seq data analysis, demonstrating their utility in tackling computational challenges in this emerging area (Zheng and Wang, 2019).

5.2 Integration of deep learning with other AI techniques

The synergy between deep learning and other AI techniques has opened new avenues for bioinformatics research. Ensemble deep learning, which combines the strengths of ensemble methods and deep learning models, has led to improvements in model accuracy, stability, and reproducibility across various bioinformatics applications (Cao et al., 2020). This integration has been particularly beneficial in areas such as sequence analysis and systems biology, where traditional methods fall short. Moreover, the incorporation of machine learning techniques within established bioinformatics frameworks has enhanced the efficiency of studying complex biological systems by enabling automatic feature extraction, selection, and predictive model generation (Auslander et al., 2021).

5.3 Application to emerging areas in bioinformatics

Deep learning has been increasingly applied to emerging areas in bioinformatics, demonstrating its versatility and potential. In genomics, for example, deep learning models have been used to annotate genomes, identify sequence determinants of genome functions, and even design synthetic genomic sequences (Routhier and Mozziconacci, 2022). The application of deep learning in genomics has shown higher accuracies in specific tasks compared to traditional methodologies, highlighting its potential in precision medicine (Koumakis, 2020). Furthermore, the use of deep learning in analyzing omics data and biomedical imaging has provided new insights and facilitated the discovery of novel biological patterns and relationships (Tang et al., 2019; Li et al., 2020). By leveraging these recent advances and breakthroughs, researchers are better equipped to address the complex challenges in bioinformatics, paving the way for more accurate and efficient biological data analysis and interpretation.

6 Case Studies

6.1 Successful applications in genomics

Deep learning has significantly impacted genomics, providing higher accuracies in specific tasks compared to traditional methodologies. For instance, deep learning models have been successfully applied to regulatory genomics, variant calling, and pathogenicity scoring, demonstrating their ability to identify complex patterns in large genomic datasets (Zou et al., 2018). Additionally, the integration of deep learning with high-throughput sequencing technologies has enabled researchers to analyze vast amounts of genomics data, facilitating precision medicine through multi-scale and multimodal data analysis (Figure 1) (Koumakis, 2020). These advancements underscore the potential of deep learning to accelerate changes in genomics research, particularly in the analysis and interpretation of big data (Min et al., 2016).

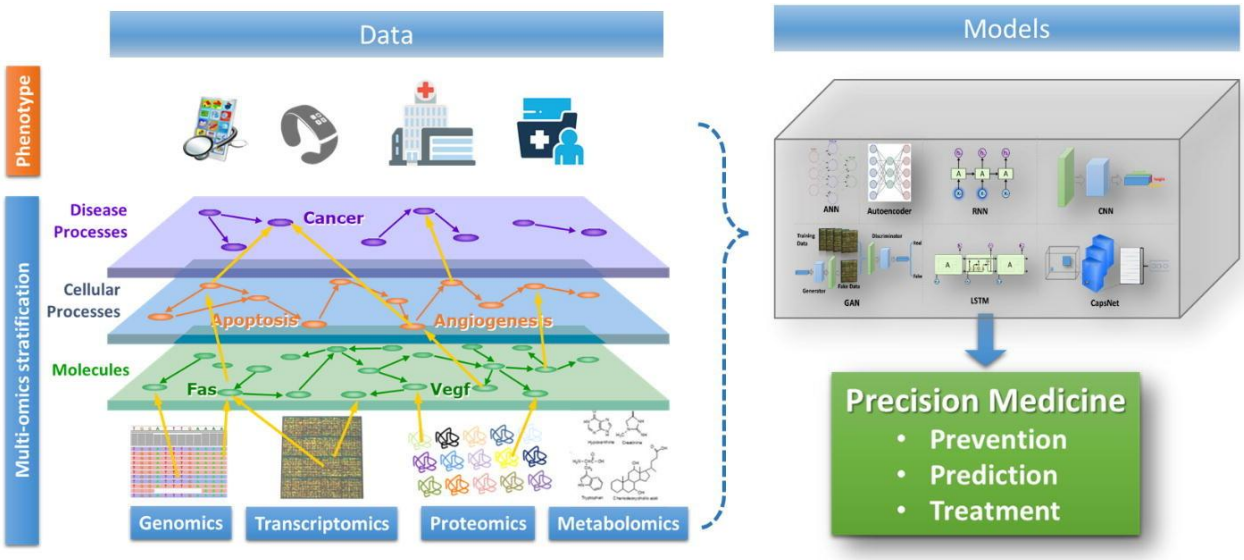

Figure 1 Multi level and multi scale -omics models (Adopted from Koumakis, 2020)

Image caption: Deep learning has the ability to deal with multimodal data effectively and genomics offers extremely heterogeneous data. The notion of precision medicine is based on the multimodal data analysis (Adopted from Koumakis, 2020)

6.2 Deep learning in drug development

The application of deep learning in drug development has shown promising results, particularly in in silico drug discovery and development. Graph neural networks (GNNs), a type of deep learning model, have been employed to predict protein-protein interactions and protein functions, which are crucial for identifying potential drug targets (Muzio et al., 2020). Furthermore, deep learning models can predict how small molecules will modulate the activity of therapeutically relevant proteins, aiding in the identification of new drug candidates (Wainberg et al., 2018). These models have also been integrated with established bioinformatics frameworks to enhance the study of complex biological systems, thereby improving the efficiency and accuracy of drug development processes (Figure 2) (Auslander et al., 2021).

6.3 Predictive modeling in disease research

Deep learning has also made significant strides in disease research through predictive modeling. For example, deep learning algorithms have been used to predict disease progression and treatment outcomes by analyzing biomedical big data (Min et al., 2016). In the context of disease diagnosis, deep learning models have been applied to gene regulatory networks to predict gene interactions and automatically diagnose diseases from data (Muzio et al., 2020). Additionally, these models have been utilized to analyze medical images and physiological data, such as electroencephalography, to understand human health and disease better (Cao et al., 2018). The ability of deep learning to integrate vast datasets and learn complex relationships has made it a valuable tool in predictive modeling for disease research (Wainberg et al., 2018).

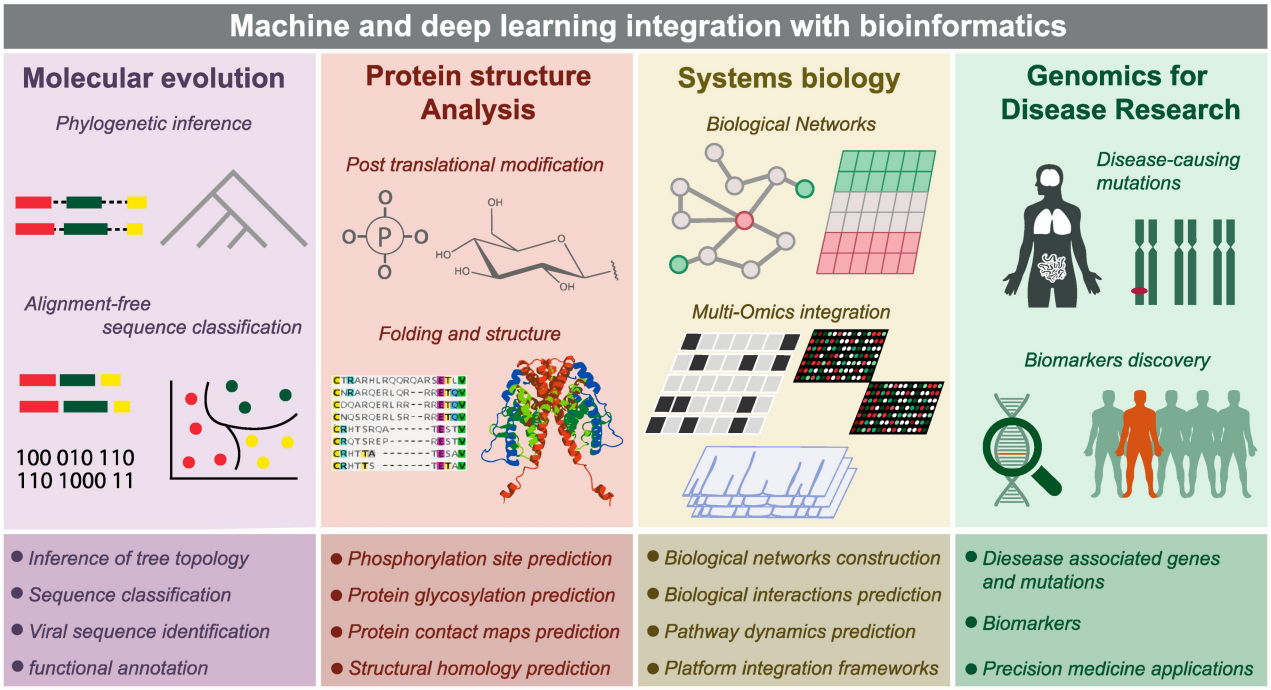

Figure 2 Applications of integrated machine learning techniques with bioinformatics in molecular evolution, protein structure analysis, systems biology, and disease genomics (Adopted from Auslander et al., 2021)

7 Future Directions in Deep Learning for Bioinformatics

7.1 Ethical and societal considerations

As deep learning continues to revolutionize bioinformatics, it is crucial to address the ethical and societal implications of these advancements. The integration of deep learning in bioinformatics raises several ethical concerns, including data privacy, consent, and the potential for bias in predictive models. For instance, the use of large-scale genomic data necessitates stringent measures to protect individual privacy and ensure that data usage complies with ethical standards (Koumakis, 2020). Additionally, the development of deep learning models must consider the potential for algorithmic bias, which can lead to disparities in healthcare outcomes if not properly managed (Berrar and Dubitzky, 2021). Addressing these ethical and societal considerations is essential to ensure that the benefits of deep learning in bioinformatics are equitably distributed and that the technology is used responsibly.

7.2 Developing interdisciplinary approaches

The future of deep learning in bioinformatics lies in fostering interdisciplinary collaborations that bring together expertise from various fields such as computer science, biology, and medicine. The complexity of biological data and the challenges associated with its analysis require a multidisciplinary approach to develop robust and effective deep learning models (Min et al., 2016; Li et al., 2020). For example, combining knowledge from genomics, computational biology, and machine learning can lead to the development of more accurate predictive models and novel applications in precision medicine (Koumakis, 2020; Routhier and Mozziconacci, 2022). Interdisciplinary approaches can also facilitate the integration of diverse data types, such as omics data, biomedical imaging, and clinical records, to provide a more comprehensive understanding of biological processes and disease mechanisms (Li et al., 2019; Tang et al., 2019).

7.3 Expanding applications beyond current scope

While deep learning has already demonstrated significant potential in various bioinformatics applications, there is still much room for expansion beyond the current scope. Future research should explore the application of deep learning to new and emerging areas within bioinformatics, such as synthetic biology and personalized medicine. For instance, deep learning can be used to design synthetic genomic sequences with desired properties, opening up new possibilities in genetic engineering and biotechnology (Routhier and Mozziconacci, 2022). Additionally, the

integration of deep learning with other advanced technologies, such as CRISPR and single-cell sequencing, can lead to breakthroughs in understanding complex biological systems and developing targeted therapies (Cao et al., 2020; Karim et al., 2020). By expanding the applications of deep learning, researchers can unlock new insights and drive innovation in bioinformatics and related fields.

Acknowledgments

I would like to express my special thanks to Professor Shakti from Hainan University for providing in-depth guidance during the research process.

Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Auslander N., Gussow A., and Koonin E., 2021, Incorporating machine learning into established bioinformatics frameworks, International Journal of Molecular Sciences, 22(6): 2903.
https://doi.org/10.3390/ijms22062903

Berrar D., and Dubitzky W., 2021, Deep learning in bioinformatics and biomedicine, Briefings in Bioinformatics, 22(2): 1513-1514.
https://doi.org/10.1093/bib/bbab087

Cao C.S., Liu F., Tan H., Song D.S., Shu W.J., Li W.Z., Zhou Y.M., Bo X.C., and Xie Z., 2018, Deep learning and its applications in biomedicine, Genomics, Proteomics and Bioinformatics, 16(1): 17-32.
https://doi.org/10.1016/j.gpb.2017.07.003

Cao Y., Geddes T.A., Yang J.Y., and Yang P.Y., 2020, Ensemble deep learning in bioinformatics, Nature Machine Intelligence, 2(9): 500-508.
https://doi.org/10.1038/s42256-020-0217-y

Gauthier J., Vincent A.T., Charette S.J., and Derome N., 2019, A brief history of bioinformatics, Briefings in Bioinformatics, 20(6): 1981-1996.
https://doi.org/10.1093/bib/bby063

Goh G.B., Hodas N.O., and Vishnu A., 2017, Deep learning for computational chemistry, Journal of Computational Chemistry, 38(16): 1291-1307.
https://doi.org/10.1002/jcc.24764

Jin S., Zeng X., Xia F., Huang W., and Liu X., 2020, Application of deep learning methods in biological networks, Briefings in Bioinformatics, 22(2): 1902-1917.
https://doi.org/10.1093/bib/bbaa043

Karim M., Beyan O., Zappa A., Costa I., Rebholz-Schuhmann D., Cochez M., and Decker S., 2020, Deep learning-based clustering approaches for bioinformatics, Briefings in Bioinformatics, 22(1): 393-415.
https://doi.org/10.1093/bib/bbz170

Koumakis L., 2020, Deep learning models in genomics; are we there yet, Computational and Structural Biotechnology Journal, 18: 1466-1473.
https://doi.org/10.1016/j.csbj.2020.06.017

Lan K., Wang D.T., Fong S., Liu L.S., Wong K.K.L., and Dey N.J., 2018, A survey of data mining and deep learning in bioinformatics, Journal of Medical Systems, 42: 1-20.
https://doi.org/10.1007/s10916-018-1003-9

LeCun Y., Bengio Y., and Hinton G., 2015, Deep learning, Nature, 521: 436-444.
https://doi.org/10.1038/nature14539

Li H.Y., Tian S.Y., Li Y., Fang Q.M., Tan R.B., Pan Y.J., Huang C., Xu Y., and Gao X., 2020, Modern deep learning in bioinformatics, Journal of Molecular Cell Biology, 12(11): 823-827.
https://doi.org/10.1093/jmcb/mjaa030

Li Y., Huang C., Ding L., Li Z., Pan Y., and Gao X., 2019, Deep learning in bioinformatics: introduction, application, and perspective in the big data era, Methods, 166: 4-21.
https://doi.org/10.1016/J.YMETH.2019.04.008

Libbrecht M.W., and Noble W.S,, 2015, Machine learning applications in genetics and genomics, Nature Reviews Genetics, 16(6): 321-332.
https://doi.org/10.1038/nrg3920

Liu W.B., Wang Z.D., Liu X.H., Zeng N.Y., Liu Y.R., and Alsaadi F.E., 2017, A survey of deep neural network architectures and their applications, Neurocomputing, 234: 11-26.
https://doi.org/10.1016/j.neucom.2016.12.038

Mamoshina P., Vieira A., Putin E., and Zhavoronkov A., 2016, Applications of deep learning in biomedicine, Molecular Pharmaceutics, 13(5): 1445-1454.
https://doi.org/10.1021/acs.molpharmaceut.5b00982

Min S., Lee B., and Yoon S., 2016, Deep learning in bioinformatics, Briefings in Bioinformatics, 18(5): 851-869.
https://doi.org/10.1093/bib/bbw068

Muzio G., O'Bray L., and Borgwardt K., 2020, Biological network analysis with deep learning, Briefings in Bioinformatics, 22(2): 1515-1530.
https://doi.org/10.1093/bib/bbaa257

Ravì D., Wong C., Deligianni F., Berthelot M., Andreu-Perez J., Lo B., and Yang G., 2017, Deep learning for health informatics, IEEE Journal of Biomedical and Health Informatics, 21(1): 4-21.
https://doi.org/10.1109/JBHI.2016.2636665

Routhier E., and Mozziconacci J., 2022, Genomics enters the deep learning era, PeerJ, 10: e13613.
https://doi.org/10.7717/peerj.13613

Saha G.C., Babur M., Khan M.J., Saha H., Kumar D., and Avinash, 2023, Integrative analysis of multi-omics data with deep learning: challenges and opportunities in bioinformatics, Tuijin Jishu/Journal of Propulsion Technology, 44(3): 1384-1392.
https://doi.org/10.52783/tjjpt.v44.i3.488

Schmidt B., and Hildebrandt A., 2020, Deep learning in next-generation sequencing, Drug Discovery Today, 26(1): 173-180.
https://doi.org/10.1016/j.drudis.2020.10.002

Shrestha A., and Mahmood A., 2019, Review of deep learning algorithms and architectures, IEEE Access, 7: 53040-53065.
https://doi.org/10.1109/ACCESS.2019.2912200

Talukder A., Barham C., Li X., and Hu H., 2020, Interpretation of deep learning in genomics and epigenomics, Briefings in Bioinformatics, 22(3): bbaa177.
https://doi.org/10.1093/bib/bbaa177

Tang B.H., Pan Z.X., Yin K., and Khateeb A., 2019, Recent advances of deep learning in bioinformatics and computational biology, Frontiers in Genetics, 10: 214.
https://doi.org/10.3389/fgene.2019.00214

Wainberg M., Merico D., Delong A., and Frey B., 2018, Deep learning in biomedicine, Nature Biotechnology, 36: 829-838.
https://doi.org/10.1038/nbt.4233

Wang R., Jiang Y., Jin J., Yin C., Yu H., Wang F., Feng J., Su R., Nakai K., Zou Q., and Wei L., 2023, DeepBIO: an automated and interpretable deep-learning platform for high-throughput biological sequence prediction, functional annotation and visualization analysis, Nucleic Acids Research, 51(7): 3017-3029.
https://doi.org/10.1093/nar/gkad055

Zhang Y., Yan J., Chen S., Gong M., Gao D., Zhu M., and Gan W., 2020, Review of the applications of deep learning in bioinformatics, Current Bioinformatics, 15(8): 898-911.
https://doi.org/10.2174/1574893615999200711165743

Zheng J., and Wang K., 2019, Emerging deep learning methods for single-cell RNA-seq data analysis, Quantitative Biology, 7(4): 247-254.
https://doi.org/10.1007/s40484-019-0189-2

Zou J., Huss M., Abid A., Mohammadi P., Torkamani A., and 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