

REVIEW ARTICLE

Applications of large-scale artificial intelligence models in bioinformatics

Mingjing Li | Qichen Shang | Ziyang Dong | Zhixuan You | Le Zhang |
Ming Xiao

College of Computer Science, Sichuan University, Chengdu, China

Correspondence

Ming Xiao and Le Zhang.
Email: xiaoming@scu.edu.cn and
zhangle06@scu.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 62372316; Noncommunicable Chronic Diseases-National Science and Technology Major Project, Grant/Award Number: 2024ZD0532900; Sichuan Science and Technology Program Key Project, Grant/Award Numbers: 2024YFHZ0091, 2025YFHZ0066

Abstract

Large-scale artificial intelligence (AI) models can mine potential patterns from massive amounts of data and provide more accurate analyses. This capability has enabled its gradual application in various areas of bioinformatics. However, few reviews have comprehensively summarized the applications of different types of large-scale AI models in key areas of bioinformatics. Therefore, we first introduce the concept of large-scale AI models and classify them into three types. Second, we summarize the key methods, applications, and resources of these three types of bioinformatics models. Finally, we discuss challenges and directions for future research. This review provides researchers with a comprehensive perspective to better understand the applications of large-scale AI models in bioinformatics.

KEYWORDS

artificial intelligence, bioinformatics, large-scale language models, large-scale multimodal models, large-scale vision models

1 | INTRODUCTION

High-throughput technologies generate large-scale and complex omics data that challenge classical bioinformatics methods in terms of data management and interpretation [1]. In this context, with the rapid development of artificial intelligence (AI) technologies, various classical machine learning (ML) and deep learning (DL) AI methods have been widely applied in bioinformatics, significantly improving the efficiency and accuracy of bioinformatics algorithms [2–5]. For instance, self-supervised learning techniques can enhance the accuracy of protein structure prediction [6]; DL enables effective prediction of single-cell deoxyribonucleic acid

(DNA) methylation states [7, 8]; and variational autoencoders [9] have been applied to biomolecular design.

Despite the significant achievements of AI methods in bioinformatics, they still face three major challenges. First, the performance improvement of these methods relies heavily on a large amount of high-quality annotated data [10]. For instance, deep neural networks are often ineffective when the number of annotated proteins is low. Correctly integrating various data types and annotations also poses challenges for the accurate prediction of microbial protein functions [11]. Second, classical AI methods often perform well on specific datasets but exhibit weak generalization ability when faced with unseen data or slightly different data

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Quantitative Biology* published by John Wiley & Sons Australia, Ltd on behalf of Higher Education Press.

distributions [12–14], mainly limited by data heterogeneity, noise, and data distribution sensitivity [15]. Finally, classical AI methods also exhibit limitations when analyzing large-scale datasets [16, 17].

Recently, large-scale AI models [18] have been increasingly applied in various fields. For instance, AlphaFold2 [19] revolutionized protein structure prediction (the Nobel Prize in Chemistry 2022), whereas its successor, AlphaFold3 [20], extended predictions to protein complexes, nucleic acids, and ligands. Concurrently, Evo [21] enables multi-omics sequence analysis/design, and BioBERT [22] and other pre-trained language models [23, 24] significantly enhance biomedical text mining through domain-specific pre-training. These models demonstrate strong potential for addressing the three core challenges outlined earlier by mining patterns from massive datasets to deliver precise analyses.

Several reviews have discussed the applications of large-scale AI models in bioinformatics. For instance, Wang et al. [25] summarized the progress of ChatGPT in bioinformatics and explored its potential and challenges in areas such as omics and drug discovery. Zhang et al. and Chen et al. [26, 27] have examined the applications of transformer-based language models in bioinformatics, and Shahid and Ruan et al. [28, 29] analyzed the challenges and limitations of fine-tuning large language models (LLMs) in various bioinformatics domains. However, these reviews

primarily focused on the applications of LLMs in bioinformatics and lacked a comprehensive summary of the applications of different types of large-scale AI models across the key areas of bioinformatics.

Therefore, in this review, we first introduce the concepts and classifications of large-scale AI models. Second, we summarize the practical applications of large-scale AI models in bioinformatics (Figure 1). Third, future research directions and challenges are discussed. Overall, this review provides researchers with a comprehensive perspective to better understand the applications of large-scale AI models in bioinformatics.

2 | OVERVIEW OF LARGE-SCALE AI MODELS

2.1 | What are large-scale AI models?

Large-scale AI models refer to AI models with numerous parameters and complex architectures [18, 30, 31], which are typically trained on massive datasets and can handle complex tasks, such as natural language processing (NLP) and image recognition [32–35]. Large-scale AI models generally have characteristics such as numerous parameters, complex architectures, high computational resource demands, and strong adaptability to various tasks [36].

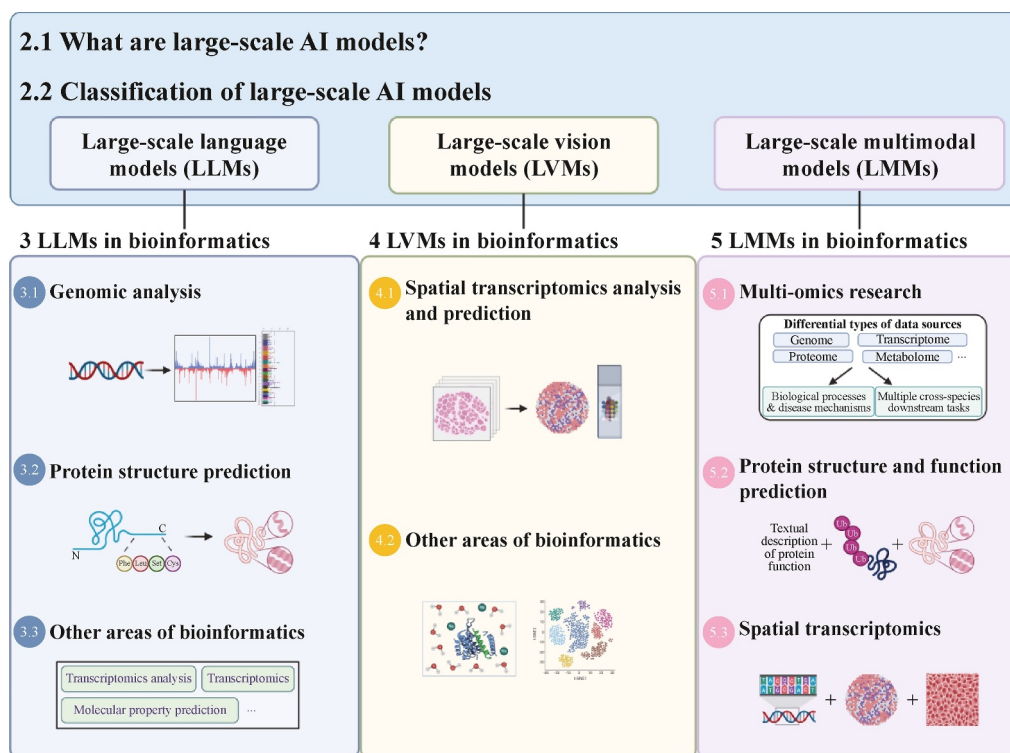


FIGURE 1 Main topics of the review.

2.2 | Classification of large-scale AI models

Based on the pretraining data modalities [37] and task requirements [18], large-scale AI models can be classified into three types: large-scale language models (LLMs), large-scale vision models (LVMs), and large-scale multimodal models (LMMs) [37]. This classification is based on consensus in the field and the latest trends. Furthermore, the reasons for our choice of a modality-based framework over other potential structures, such as organizing by biological problems or by model architecture, are as follows. On the one hand, a problem-based structure would lead to significant redundancy, as a single model type such as LLMs finds applications across numerous biological domains. On the other hand, although an architecture-based approach is technically precise, it is less intuitive for an interdisciplinary bioinformatics audience. Classification based on pretraining data modalities and task requirements not only provides a clear narrative, but also serves as a practical guide for researchers seeking to identify the most appropriate AI tools for their specific data types.

LLMs [38] are mainly used for processing and generating linguistic data and have shown strong performance and broad application potential across various NLP tasks. Recently, the rapid development of LLMs has primarily relied on the widespread application of the Transformer [39]. The Transformer comprises an encoder and decoder, both of which are composed of multiple stacked identical layers. Each layer contains two main components: self-attention [39] and a feed-forward neural network (FNN) [40]. Self-attention computes the association weights between any position in a sequence, thereby overcoming the limitations of traditional neural networks in capturing long-range dependencies. The FNN introduces nonlinear transformations, further enhancing the capacity of the model to express complex patterns. Through stacking, these components form a multilayered structure, enabling the Transformer to progressively construct and refine representations of the input sequences, thereby effectively processing various sequential tasks.

Building on this foundation, BERT [41] and GPT [42], two typical applications of the Transformer architecture, exhibit different design characteristics and applicable scenarios. BERT employs a bidirectional encoder structure, making it suitable for tasks that require a deep understanding of contextual information [43]. In contrast, GPT focuses on an autoregressive decoder structure that performs better in tasks that require generating coherent text [44].

LVMs [45] are primarily used to process image data, such as in image recognition and classification. LVMs combine the advantages of convolutional neural networks [46] and Transformers [47–49]. CNNs compute

dot products in local regions using small-sized convolutional kernels, focusing on extracting local features from images; however, they have limitations when processing long-distance dependencies. As a complement, the self-attention mechanism in Transformers dynamically computes correlations between different image segments, thereby capturing long-range dependencies in images to understand the global connections between different regions. For instance, vision Transformers (ViTs) [50] convert various visual data into “visual sentences” [51] and perform next token prediction to achieve automated image understanding. After we contrast LVMs with LLMs, we found that LLMs typically rely solely on Transformer-based architectures to process sequential data, whereas LVMs integrate CNNs’ local feature extraction with Transformers’ ability to capture long-range dependencies to achieve comprehensive image understanding. This fundamental difference explains why LVMs are uniquely suited to tasks involving biological imaging, such as spatial transcriptomics (ST).

LMMs [52] enable unified understanding and generation of data by integrating information from different modalities. This process typically involves key components such as the feature extractor, fusion layer, and generator. First, the feature extractor extracts the corresponding feature representations from the input data of each modality. Next, the fusion layer performs cross-modal interaction and integration of the extracted features. Finally, the generator generates the required output for the task based on the fused multimodal representations. For instance, the modality module in Macaw-LLM [53] can simultaneously handle data from multiple modalities, such as text, images, and audio, thereby supporting multimodal understanding and interaction tasks.

3 | LLMs IN BIOINFORMATICS

It is one of the basic studies in bioinformatics to employ various sequencing data (such as DNA, RNA, or protein sequences) to predict biomolecular structures and functions [54–60]. LLMs, particularly those based on the Transformer architecture, were originally designed to capture complex, non-local contextual dependencies in natural language. Given the analogy between sequencing data and linguistic word sequences [61], LLMs can learn the underlying grammatical and semantic rules of biological sequences. Consequently, they have become increasingly prevalent in bioinformatics applications.

By leveraging architectures, such as Transformer to process large-scale biological data, these LLM models play a pivotal role in critical bioinformatics tasks, including protein structure prediction and genome analysis. Because amino acid or nucleotide sequences are typically represented by single letters, LLMs cannot

interpret these letters directly. Therefore, one of the key steps in applying protein sequences to LLMs for structural prediction is the transformation of sequencing sequences into numerical vectors.

Figure 2 uses DNA sequences as an example to illustrate the key steps in converting sequences into numerical vectors. First, tokenization is performed individually for each sequence, where the (CLS) and (SEP) tokens are added to mark the start and end of the sequence. Second, non-overlapping k -mer segmentation [62] (assuming $k = 6$) is applied to the sequence. For bases at the end of a sequence that form an

incomplete k -mer, the model typically treats it as a single token. These tokens are the basic units used by LLMs for understanding and prediction [63], and can be used to construct biologically meaningful vocabularies.

Next, embedding is used to map the padded and masked sequences into a vector space, similar to word embeddings [64] in NLP. To meet the requirements for uniform length when LLMs process sequences in batches, we not only extend shorter sequences to a uniform length by appending padding tokens (PAD), but also apply masking to distinguish between valid sequence elements and padding tokens, which can guarantee that the model focuses exclusively on valid data during prediction [65]. In addition, to capture the relative positions of nucleotides or amino acids within a sequence, positional encoding is typically added to the embedding vectors [66].

Finally, the numerical vectors are passed to the decoder for analysis and prediction. The FNN and attention mechanisms in the decoder are common to LLMs. However, each model selects more appropriate attention mechanisms based on practical applications, such as multi-head attention mechanisms and multi-head masked attention mechanisms. Among them, the multi-head attention mechanism [67] divides input embeddings into multiple subspaces, with each self-attention head independently focusing on different features. By computing multiple self-attention subspaces in parallel, it captures various types of relationships. For instance, the decoder employs a multi-head attention mechanism to capture diverse dependencies within the input data and employs a masking mechanism [68] to reduce the weights of future positions, ensuring that the model does not use future information during prediction. The synergistic interaction of each module ensures that the decoder generates accurate predictive results.

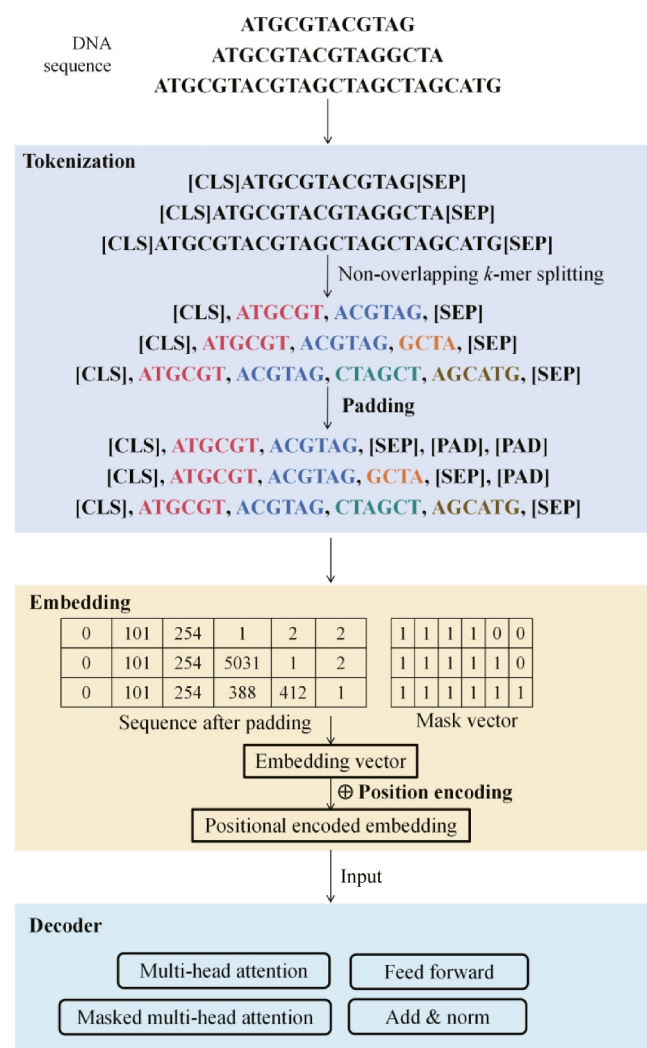


FIGURE 2 Main process of converting sequences into numerical vectors for analysis and prediction in LLMs (using DNA gene sequences as an example). Through tokenization, padding, and masking, the data undergo preprocessing while ensuring the model focuses exclusively on valid data. Subsequently, embedding operations map these sequences into vector spaces, combined with positional encoding to capture the relative positional information of nucleotides. The resulting numerical vectors are then input into the decoder for deep extraction of complex sequence features. DNA, deoxyribonucleic acid; LLMs, large language models.

3.1 | Genomic analysis

With the development of next-generation sequencing technologies, the volume of genomic data has grown exponentially [69]. However, classical analytical methods exhibit an insufficient capability to extract features based on contextual information when dealing with complex genomic data [70, 71].

Consequently, LLMs, which leverage their deep understanding of complex sequences and automatic feature extraction capabilities [72], have been gradually applied to key tasks in genomics, such as gene expression and sequence function prediction [72]. Notably, genomes contain numerous non-coding regions and repetitive sequences, and regulatory regions may be distant from the target genes. The diversity and noise in these data require stronger generalization capabilities, which increase the difficulty of sequence analysis [73].

Generally, the main steps for using LLMs in genomic analysis (Figure 2) are as follows. First, tokenization is performed individually on each sequence by adding tokens at the start and end positions and segmenting the sequence into multiple tokens to construct a vocabulary of gene sequences. The padded and masked sequences are then mapped to vector spaces through embedding. Subsequently, these numerical vectors are fed into the decoder, which employs self-attention mechanisms to effectively capture contextual relationships between positions within the sequence. During the pretraining phase, LLMs typically adopt bidirectional context pretraining [41, 74], which considers the contextual information both preceding and following a given position in the sequence. This enables the model to identify and predict key regions within gene sequences, thereby enhancing the accuracy of subsequent specific tasks such as gene expression prediction and sequence function analysis.

Currently, bidirectionally context-pretrained LLMs have been widely applied to gene expression and sequence function prediction [62, 75–79]. For instance, Enformer [75] employs an optimized self-attention mechanism to integrate interaction information spanning up to 200,000 base pairs, thereby increasing the accuracy of gene expression prediction by 5% and 22.6% compared with the ML-based ExPecto [80] in cross-gene and cross-tissue gene expression evaluations, respectively. DNABERT-2 [62], by introducing byte pair encoding and linear bias attention (ALiBi), optimizes the handling of long sequences and outperforms HyenaDNA [77] across multiple tasks. The nucleotide Transformer (NT) [76] learns the context-dependent features of nucleotides from unlabeled DNA sequences and applies them to various genomic prediction tasks through fine-tuning. In promoter prediction tasks, NT achieves a correlation coefficient that is 10% higher than that of HyenaDNA, and in gene splicing prediction tasks, NT attained a correlation coefficient of 0.96, surpassing that of Enformer, which is 0.74.

3.2 | Protein structure prediction

Protein structure prediction is one of the most important tasks in bioinformatics. Classical experimental techniques, such as X-ray crystallography [81] and cryo-electron microscopy [82], can provide highly precise structural information, but are usually costly and time-consuming [83]. Recently, the application of large-scale language models to protein structure prediction has attracted widespread attention. This section focuses on the application of sequence-based LLMs in protein structure prediction, and Section 5.2 will discuss LLMs approaches that integrate multiple data sources.

LLM applications for protein structure prediction, such as for genomic sequences, are also based on

sequence inputs. However, LLMs in the protein domains focus more on mapping sequences to structural spaces for protein structure prediction [84]; whereas in genomics, LLMs emphasize sequence pattern recognition and functional associations using their sequence modeling capabilities to process ultra-long DNA sequences (Figure 2), as well as to capture and predict regulatory patterns [21].

The application of LLMs in protein structure prediction can be primarily categorized into multiple sequence alignment (MSA)-based [85] and MSA-free models [86], both of which are based on the Transformer architecture. All MSA-based models require MSAs as input, whereas MSA-free models typically operate on single sequences or their embeddings. Furthermore, MSA-based models employ Transformer layers with row and column attention to extract co-evolutionary signals from MSAs, enabling them to achieve high predictive accuracy for proteins with numerous homologs [85]. In contrast, MSA-free models utilize multi-head attention to generate contextual embeddings that capture structural and functional signals. However, the performance of MSA-based models degrades significantly in the absence of sufficiently homologous sequences [19]. To date, MSA-free models such as HelixFold-Single [87] and ESM-2 [88] have achieved competitive performance.

The main steps in using MSA-free LLMs to predict protein structures include the transformation of protein amino acid sequences into numerical vectors through embedding. Subsequently, various computational models [89–93] facilitate feature fusion [94] by integrating embedding vectors with the physicochemical properties of amino acids (such as hydrophobicity and charge) to generate richer and more comprehensive feature representations. Following this integration, the decoder analyzes the integrated protein vector and predicts the protein structure. The predictive accuracy is ultimately enhanced through large-scale pretraining [95] and fine-tuning [96] methodologies applied to specific protein structure prediction tasks.

Currently, numerous LLMs have been successfully applied for protein structure prediction. For instance, ESM-2 [88] focuses on sequence feature learning based on the aforementioned approach and is 60 times faster than AlphaFold in predicting 3D protein structures. xTrimoABFold [45] is specifically designed for predicting antibody structures and protein-antibody interactions and outperforms AlphaFold2 [19] in both accuracy and computational efficiency for antibody structure prediction tasks.

3.3 | Other areas of bioinformatics

In addition, LLMs have found applications in other important areas of bioinformatics, including transcriptomics [97], single-cell sequencing analysis [98],

and molecular property prediction. For instance, transcriptomics is used to study an organism's transcriptome and the sum of all RNA transcripts. Transcriptomic analysis has enabled the study of changes in gene expression in different organisms and has been instrumental in understanding human disease [99, 100]. In transcriptomics, ERNIE-RNA [101] introduces pairing constraints into the self-attention mechanism, thereby enabling the capture of RNA structural features without MSA [102]. In single-cell sequencing analysis, scFoundation [103] is pretrained on more than 50 million single-cell transcriptomic datasets, making it widely applicable to tasks such as deep cell sequencing and cell drug response prediction. In molecular property prediction, SMILES-BERT [104] learns deep representations of molecular structures through a masked SMILES recovery task, outperforming traditional methods based on CNNs (e.g., XGBoost [105]) and GCNs (such as HRGCN+ [106]) across multiple molecular property prediction tasks. MolecularGPT [107], based on LLaMA-2 [108], is designed for few-shot molecular property prediction through instruction fine-tuning, and can handle more than 1000 tasks. In molecular zero-shot learning, MolecularGPT improves accuracy by 15.7% compared with GIMLET [109] in classification tasks and reduces the root mean square error by 17.9% in regression tasks, demonstrating its strong generalization capabilities. In addition, in the field of scientific discovery, LLM4SD [110] can extract known information from scientific literature and uncover patterns from molecular data to predict molecular properties.

4 | LVMs IN BIOINFORMATICS

With the development of bioinformatics, various experimental biological imaging datasets have been widely used. These images contain crucial biological information, such as cell morphology, tissue structure, protein localization, and molecular interactions, which are characterized by high dimensionality and complex correlations [111]. Local features and global patterns in biological images jointly determine biological significance, increasing the complexity of the analysis. Owing to their feature representation capabilities and ability to capture multi-scale dependencies, LVMs can simultaneously model details at both the global tissue and local cellular levels. LVMs are increasingly applied in biological image analysis to improve the accuracy of prediction and classification in this field.

LVMs encompass several critical phases. Because the image data in bioinformatics typically have different resolutions and qualities, it is necessary to standardize the image size by resizing and normalizing the pixel

values. The standardized images are segmented into fixed-size patches, followed by dimensional transformation through embedding mechanisms, which collectively constitute the patch-embedding process [109]. Positional encoding is subsequently incorporated into the embedding sequence before transmission to the decoder component. Within the decoder architecture, self-attention mechanisms facilitate the computation of inter-patch similarities, thereby capturing spatial relationships across distinct image regions. The implementation of multiple cascading encoder layers enables progressive feature extraction and representation optimization. The resultant feature representations extracted through LVMs after pretraining and fine-tuning protocols can be applied across various downstream analytical tasks.

4.1 | ST analysis and prediction

In ST experiments, RNA sequencing combined with microscopy or RNA imaging techniques [112] generates high-resolution images. Classical ML methods [113–116] require dimensionality reduction or feature selection when processing high-dimensional image data, which is time-consuming and prone to information loss [117]. LVMs can efficiently and accurately process high-dimensional data using patching and self-attention mechanisms. This model can generate global representations of cell populations or tissues with multiple layers of self-attention mechanisms, thereby providing critical contextual information for the inference of gene expression levels and downstream tasks, such as transcriptomic data prediction.

For instance, Hist2ST [118] combines Transformer and graph neural network to analyze spatial dependencies and image features in histological images, achieving RNA ST data prediction from histological images. Its computational efficiency significantly exceeds that of the ML-based ST-Net [119] on the breast cancer dataset. THItGene [120] combines ViT [50] with the Efficient-CapsNet [121] neural network to capture long-range dependencies and spatial hierarchies in images, which can predict gene expression data from histological images, and has achieved a performance score of 0.327 on the HER2+ dataset, surpassing that of Hist2ST, which is 0.250.

4.2 | Other areas of bioinformatics

In addition, LVMs are gradually being applied in fields, such as protein function prediction and genomics. For instance, in protein function prediction, ViTScore [122] voxelizes the protein-ligand interaction pocket into a 3D

grid, where each voxel represents a spatial region of the protein structure and labels different types of atoms or amino acids. This not only preserves complete spatial information but also highlights the details of protein-ligand interactions, which helps us accurately predict protein-ligand docking poses. In genomics, GeneViT [123] converts gene sequences into numerical vectors and then maps these vectors into a two-dimensional space using *t*-SNE [124] method to generate images. ViT [50] is subsequently applied to extract image features for the classification of gene expression data. Its accuracy on benchmark datasets is 98.5%, surpassing the 96% accuracy of the ML algorithm MFOA [125].

5 | LMMs IN BIOINFORMATICS

The field of bioinformatics involves multiple modal data types, such as genomic sequence data in text modality, transcriptomic data represented in matrix modality, and biological experimental images belonging to image modality, which are usually of high dimensionality, heterogeneity, and complex internal correlations

[126]. Classical single-modal models face inefficiency and serious information loss when handling multimodal data [127]. Because of their powerful data fusion and pattern recognition capabilities, LMMs provide a new technological solution for bioinformatics. The core of LMMs lies in fusing complementary information from different data sources to enhance the predictive capability [128]. Figure 3 illustrates the main steps of LMMs for different modal datasets. First, through their respective modal encoders, the data from the different modalities are preprocessed, and then converted into a unified representation after feature extraction for subsequent fusion and analysis. Second, a suitable fusion strategy (early fusion, mid fusion, or late fusion) [129] is used to integrate the data representations of different modalities whereas cross attention [130, 131] maps the features of each modality into three structures: query, key, and value, and each attention head is responsible for computing its own attentional scores. These scores reflect the relationship between themselves and different modalities. By dynamically weighting and aggregating the features of different modalities, it dynamically focuses on important information between different modalities, further improving

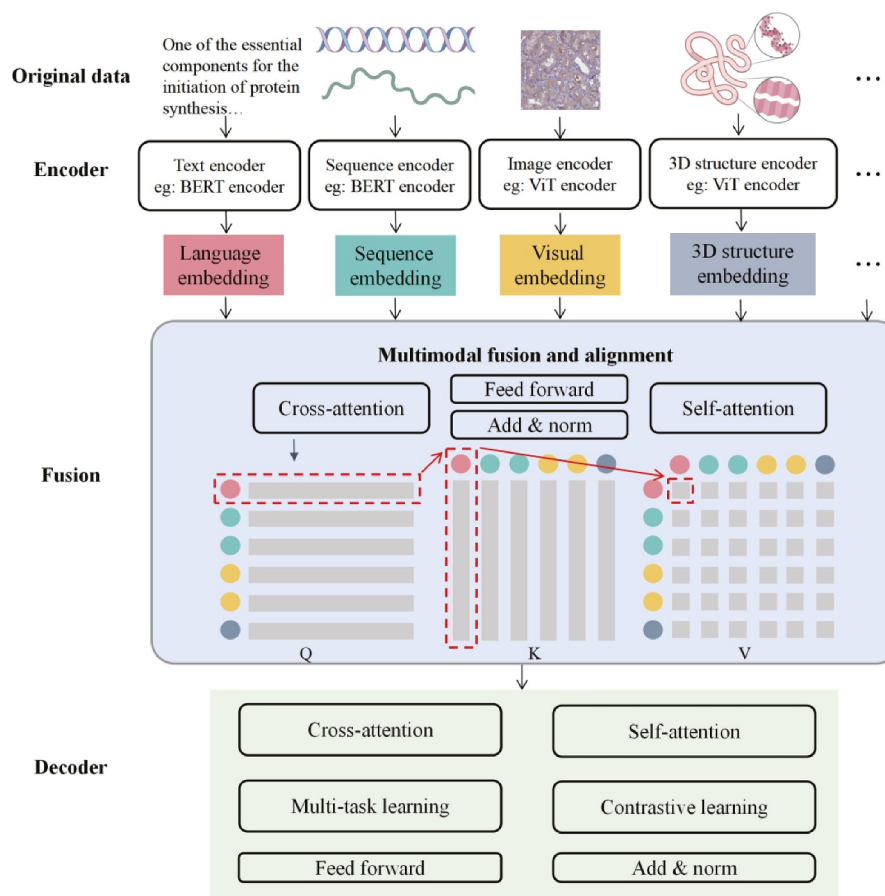


FIGURE 3 Main steps of multimodal data fusion for large-scale multimodal models. Different modal data are transformed into unified representations through preprocessing and feature extraction. Appropriate fusion strategies can be adopted to integrate data, and decoders can dynamically weight and aggregate features from different modalities to increase fusion effects.

the fusion effect and flexibility. Finally, LMMs are generally pretrained on large-scale and diverse datasets to learn generalized feature representations. Strategies such as multi-task learning [132] and contrastive learning [133] are employed during the training process to further increase the performance and generalization ability of the model. Subsequently, it is adapted to specific bioinformatics tasks through transfer learning or fine-tuning.

5.1 | Multi-omics research

Multi-omics research involves data from different omics to comprehensively understand biological processes and disease mechanisms [47, 70, 113, 116, 134]. LMMs can be used to discover deeper biological phenomena by integrating the information from different omics data. For instance, DeepOmix [46] provides new perspectives for understanding complex biological processes by integrating protein, genomic, and transcriptomic data, which demonstrates significantly better survival prediction performance than DL-based survival models such as DeepSur [135] and DeepHit [136]. Evo [21] is able to parse and design DNA, RNA, and protein sequences, which performs well in predicting the effects of mutations on bacterial proteins and RNAs, as well as in modeling gene regulatory mechanisms. In the protein expression prediction task, the correlation coefficient of Evo reaches 0.60, far exceeding the 0.11 of GenSLMs [137].

Meanwhile, LMMs perform well in multiple cross-species downstream tasks, providing new perspectives for understanding the relationships between gene expression and diseases among different species. For instance, the GeneCompass [138] model integrates 126 million single-cell transcriptome data points from humans and mice, which not only predicts changes in cellular states but also reveals gene expression and regulation. The SATURN [139] model realizes the integration of non-homologous genes across species by integrating scRNA-seq data, cellular annotations, and protein expression from multiple species.

5.2 | Protein structure and function prediction

In addition to unimodal, sequence-based LLMs discussed in Section 3.2, multimodal data-driven LMMs can also be used for protein structure prediction research. In protein research, common data modalities include amino acid sequences, 3D structures, experimental data, and textual descriptions. By joint learning from different modality data, we can better capture the complex properties and interaction mechanisms of

proteins, thereby improving the comprehensiveness and accuracy of predictions [140, 141].

For instance, ESM3 [142] integrates the sequences, structures, and functions of proteins into embedding vectors in a unified way, effectively capturing both local and global structural and functional information. Helix-ProtX [120] integrates data from textual descriptions, amino acid sequences, and 3D structures, enabling the generation of any modality from any other. It outperforms LLMs such as Chroma [143] and HelixFold-Single [87] in structure prediction. Released in 2024, AlphaFold3 [20] significantly outperforms AlphaFold2 [19] in predicting protein structures and protein-protein interaction structures by reducing reliance on MSA and introducing technical improvements such as diffusion modules. Chai-1 [144] can be used to predict the molecular structures of proteins, small molecules, DNA, and other molecules without MSA. In protein antibody prediction, the accuracy without using MSA is 0.479, which is comparable to that when using MSA, and surpasses the accuracy of AlphaFold 2.3, which is 0.380. However, for protein monomer predictions, the accuracy of Chai-1 without MSA is 0.852, and this is lower than that of AlphaFold 2.3, which was 0.903. Furthermore, by processing protein sequences, ligand information, and experimental constraint data, Chai-1 achieved a success rate of 77% in the PoseBusters benchmark test [145], which is comparable to the success rate of 76% for AlphaFold 3.

5.3 | Spatial transcriptomics

ST data involve various modalities, such as spatial information, gene expression data, histomorphological images, and even epigenetics data. Gene expression data not only depend on gene expression levels, but also closely relate to their spatial locations within tissues and specific cell types [146]. LMMs can effectively integrate information from different data sources and predict transcriptomic data by learning the relationships among different data modalities.

For instance, mclSTExp [147] uses hematoxylin and eosin stained images and ST data with spatial locations for multimodal comparative learning. It exhibited excellent performance in predicting spatial gene expression by combining intrinsic features with spatial contexts using self-attention mechanisms. On the HER2+ dataset, mclSTExp outperforms BLEEP [148] by at least 23.01%. GHTNet [149] integrates different types of biological information, such as chromatin features, epigenomics data, and protein-DNA interaction data. This integration enables better prediction of the binding specificity between transcription factors and DNA. Table 1 summarizes the representative large-scale AI models used in bioinformatics.

TABLE 1 Representative large-scale AI models applied in bioinformatics.

Model name	Year	Type	Main tasks	References
ESM-2	2022	LLM	Protein 3D structure and function prediction	[88]
xTrimoABFold	2022	LLM	Antibody structure prediction	[45]
Enformer	2021	LLM	Protein 3D structure prediction	[75]
DNABERT-2	2023	LLM	Genomic sequence classification and prediction	[62]
Nucleotide transformer	2024	LLM	DNA molecular phenotype prediction	[76]
ERNIE-RNA	2024	LLM	RNA structural and functional predictions	[101]
scFoundation	2024	LLM	Drug response prediction and single-cell perturbation prediction	[103]
Hist2ST	2022	LVM	Spatial transcriptomics data prediction	[118]
THIttoGene	2024	LVM	Spatial gene expression prediction	[120]
Cellvit	2024	LVM	Segmentation of nucleus	[150]
SCTS	2023	LVM	Segmentation of nucleus	[151]
CellFormer	2023	LVM	Identify nucleus	[152]
ViTScore	2023	LVM	Protein-ligand docking prediction	[122]
GeneViT	2023	LVM	Classifying cancerous gene expression	[123]
Evo	2024	LMM	Predict gene essentiality	[21]
GeneCompass	2024	LMM	Cross-species biological investigations	[138]
SATURN	2024	LMM	Encoding genes' biological properties	[139]
ESM3	2024	LMM	Protein 3D structure prediction	[142]
AlphaFold3	2024	LMM	Predict the structure and interactions of proteins, DNA, RNA, etc.	[20]
HelixProtX	2024	LMM	Protein sequence, structure and function prediction of proteins	[120]
Chai-1	2024	LMM	Molecular structure prediction	[144]
mclSTExp	2024	LMM	Spatial transcriptomics expression prediction	[147]
GHTNet	2022	LMM	Predicting TF-DNA binding specificity	[149]

Abbreviations: AI, artificial intelligence; LLM, large language model; LMM, large-scale multimodal model; LVM, large-scale vision model; TF-DNA, transcription factor-DNA.

6 | CHALLENGES AND FUTURE PERSPECTIVE

In general, although large-scale AI models have been increasingly used in various fields of bioinformatics, there are still many shortcomings and challenges that we need to address.

In terms of LLMs, LLMs have achieved great success in bioinformatics such as protein structure prediction. However, current protein structure predictions are mostly static [19, 20], and fail to reflect the highly dynamic nature and multiple conformations of proteins in organisms. Several studies [19] have attempted to employ LLMs to analyze protein dynamics; however, they performed poorly in identifying conformational fuzziness owing to insufficient dynamic information in the training data.

Therefore, one of the most important future research directions is to increase the predictive accuracy of protein dynamic conformations by increasing

the ability of the model to learn the dynamic changes of protein conformations across states. Furthermore, how to balance and improve the accuracy and speed of large-scale AI models in tasks, such as protein structure prediction and genetic data analysis, becomes a multi-dimensional challenge. Relevant studies [153–155] have addressed this issue.

For example, parameter-efficient fine-tuning (PEFT, such as adapter and LoRA) can significantly reduce computational requirements during the fine-tuning process. In protein function prediction tasks, PEFT can reduce the training time while maintaining prediction accuracy [153]. In genomic data analysis, adopting dynamic batching strategies can adaptively adjust batch sizes based on sample complexity, fully utilizing computational resources and improving analysis speed without sacrificing accuracy [154]. Additionally, specific optimizations for different hardware platforms (such as CPU, GPU, and TPU [155]) can further enhance the performance.

LVMs have been gradually applied in ST, particularly in the extraction and analysis of spatial structural information. However, annotation of ST data is often difficult and costly [156]. Therefore, one of the future research challenges is how to train efficient models using limited annotated data. Additionally, LVMs can be applied to analyze image data in genomics, such as chromosomal banding patterns or fluorescence images of chromosomes, to study chromosomal features or abnormal information [157], which may contribute to research on major diseases such as cancer.

In terms of LMMs, due to their ability in cross-modal learning and feature fusion, LMMs are an important technological direction for parsing the interrelationships of multi-omics data and thus elucidating the mechanisms of complex biological processes, owing to their ability in cross-modal learning and feature fusion. However, the integration of multimodal information requires appropriate fusion techniques and timing.

In practical bioinformatics applications, achieving perfect matching of cross-modal data is often challenging. How to process missing information, complete incomplete data, and resolve conflicting signals between modalities remain critical issues requiring urgent resolution. Recent studies have developed various strategies to address these challenges. For instance, feature fusion techniques (such as the multimodal feature fusion module [158]) can integrate complementary information between modalities to improve data completeness. For cross-modal conflict issues, researchers have proposed commonality- and discrepancy-sensitive encoders (CDS-Encoder) and dynamic feature unification modules, which mitigate conflicts by capturing modal invariance and specificity, and dynamically integrating multimodal features, respectively [159, 160]. Furthermore, latent feature learning methods align the consistency and specificity between modalities through contrastive learning, effectively completing missing data and ensuring the spatial consistency of the cross-modal data [161, 162].

Although these methods have demonstrated advantages in their respective fields, cross-modal bioinformatics integration still lacks a systematic framework that balances modal generalization and specificity while improving the computational efficiency in practical applications.

It should be noted that while large-scale AI models have provided powerful analytical and predictive capabilities for bioinformatics, challenges regarding model accessibility and reproducibility remain unresolved. These models typically demand substantial computational resources (GPU memory and training time), complex environmental configurations, and reliance on massive pretraining datasets. Concurrently, the increase in model parameters has led to an exponential increase in computational costs and energy consumption [163]. Addressing these issues requires

multi-dimensional coordination for model optimization. For example, the computational load can be reduced by optimizing the model architecture (e.g., model distillation [159] and compression [164]). Distributed computing (e.g., model slicing [165] and joint learning [166]) can be used to increase the throughput. With the maturation of model optimization technologies and new computing architectures [167], future research should focus on developing low-resource, high-performance bioinformatics models.

7 | CONCLUSION

In summary, large-scale AI models have exhibited great potential in bioinformatics. We systematically explored the burgeoning applications and transformative potential of LLMs, LVMs, and LMMs within a diverse bioinformatic landscape. However, challenges such as high computational resource demands, low generalization and interpretability, difficulties in cross-modal data integration, and a lack of corresponding web services still remain.

Therefore, addressing how to improve the interpretability and generalization ability of large-scale AI models through model optimization, develop corresponding algorithmic web services in combination with high-performance computing technologies, or even provide low-cost solutions, will be the main direction for future research on large-scale AI models in bioinformatics.

AUTHOR CONTRIBUTIONS

Mingjing Li: Conceptualization; investigation; visualization; writing—original draft. **Qichen Shang:** Investigation; validation; visualization. **Ziyang Dong:** Investigation; validation. **Zhixuan You:** Investigation; validation. **Le Zhang:** Funding acquisition; resources; writing—review and editing. **Ming Xiao:** Methodology; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

This work was supported by grants from Noncommunicable Chronic Diseases-National Science and Technology Major Project (No. 2024ZD0532900); National Natural Science Foundation of China (No. 62372316); Sichuan Science and Technology Program Key Project (No. 2024YFHZ0091 and No. 2025YFHZ0066).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

This review article does not generate any new data.

ETHICS STATEMENT

This review article is based on a comprehensive analysis of existing literature and does not involve human or

animal subjects, experimental research, or personal data collection.

REFERENCES

- [1] Chen M, Harrison A, Shanahan H, Orlov Y. Biological big bytes: integrative analysis of large biological datasets. *J Integr Bioinform.* 2017;14(3):20170052.
- [2] Lai K, Twine N, O'Brien A, Guo Y, Bauer D. Artificial intelligence and machine learning in bioinformatics. *Life Sci.* 2019; 1:272–86.
- [3] Theodoris CV. Perspectives on benchmarking foundation models for network biology. *Quant Biol.* 2024;12(4):335–8.
- [4] Zhang MQ. Dialog between artificial intelligence & natural intelligence. *Quant Biol.* 2023;11(4):359–62.
- [5] Li B, Tan K, Lao AR, Wang H, Zheng H, Zhang L. A comprehensive review of artificial intelligence for pharmacology research. *Front Genet.* 2024;15:1450529.
- [6] Jisna VA, Jayaraj PB. Protein structure prediction: conventional and deep learning perspectives. *Protein J.* 2021;40(4): 522–44.
- [7] Angermueller C, Lee HJ, Reik W, Stegle O. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. *Genome Biol.* 2017;18(1):67.
- [8] Wei R, Zhang L, Zheng H, Xiao M. A systematic review of the application of machine learning in CpG island (CGI) detection and methylation prediction. *Curr Bioinform.* 2024;19(3): 235–49.
- [9] Kingma DP, Welling M. An introduction to variational autoencoders. *Found Trends Mach Learn.* 2019;12(4):307–92.
- [10] Greenwald NF, Miller G, Moen E, Kong A, Kagel A, Dougherty T, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat Biotechnol.* 2022;40(4):555–65.
- [11] Arden Z, Chakraborty S, Lenk F, Kaster A-K. Elucidating the functional roles of prokaryotic proteins using big data and artificial intelligence. *FEMS Microbiol Rev.* 2023;47(1):fuad003.
- [12] Bryant P, Pozzati G, Zhu W, Shenoy A, Kundrotas P, Elofsson A. Predicting the structure of large protein complexes using AlphaFold and Monte Carlo tree search. *Nat Commun.* 2022; 13(1):6028.
- [13] Wang Q, Wang B, Xu Z, Wu J, Zhao P, Li Z, et al. PSSM-distil: protein secondary structure prediction (PSSP) on low-quality PSSM by knowledge distillation with contrastive learning. *Proc AAAI Conf Artif Intell.* 2021;35(1):617–25.
- [14] Ghobashi A, Ma Q. Perspectives on integrating artificial intelligence and single-cell omics for cellular plasticity research. *Quant Biol.* 2025;13(4):e70004.
- [15] Gavrikov P, Keuper J. Can biases in ImageNet models explain generalization? In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 2024. p. 22184–94.
- [16] Li Y, Huang C, Ding L, Li Z, Pan Y, Gao X. Deep learning in bioinformatics: introduction, application, and perspective in the big data era. *Methods.* 2019;166:4–21.
- [17] Ezugwu AE, Ikotun AM, Oyelade OO, Abualigah L, Agushaka JO, Eke CI, et al. A comprehensive survey of clustering algorithms: state-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Eng Appl Artif Intell.* 2022;110(C):104743.
- [18] Wang DQ, Feng LY, Ye JG, Zou JG, Zheng YF. Accelerating the integration of ChatGPT and other large-scale AI models into biomedical research and healthcare. *MedComm Futur Med.* 2023;2(2):e43.
- [19] Yang Z, Zeng X, Zhao Y, Chen R. AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduct Targeted Ther.* 2023;8(1):115.
- [20] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature.* 2024;630(8016):493–500.
- [21] Nguyen E, Poli M, Durrant MG, Kang B, Katrekar D, Li DB, et al. Sequence modeling and design from molecular to genome scale with Evo. *Science.* 2024;386(6723):ead09336.
- [22] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics.* 2020;36(4): 1234–40.
- [23] Shue E, Liu L, Li B, Feng Z, Li X, Hu G. Empowering beginners in bioinformatics with ChatGPT. *Quant Biol.* 2023;11(2): 105–8.
- [24] Xu D. ChatGPT opens a new door for bioinformatics. *Quant Biol.* 2023;11(2):204–6.
- [25] Wang J, Cheng Z, Yao Q, Liu L, Xu D, Hu G. Bioinformatics and biomedical informatics with ChatGPT: year one review. *Quant Biol.* 2024;12(4):345–59.
- [26] Zhang S, Fan R, Liu Y, Chen S, Liu Q, Zeng W. Applications of transformer-based language models in bioinformatics: a survey. *Bioinform Adv.* 2023;3(1):vbad001.
- [27] Chen Z, Wei L, Gao G. Foundation models for bioinformatics. *Quant Biol.* 2024;12(4):339–44.
- [28] Shahid U. Leveraging fine-tuned large language models in bioinformatics: a research perspective. *Qeios.* 2023.
- [29] Ruan W, Lyu Y, Zhang J, Cai J, Shu P, Ge Y, et al. Large language models for bioinformatics. 2025. Preprint at arXiv: 2501.06271.
- [30] Feng J. Simulating the whole brain as an alternative way to achieve AGI. *Quant Biol.* 2023;11(4):471–3.
- [31] Jiang Z, Dai W, Wei Q, Qin Z, Wei R, Li M, et al. Diffusion model-based multi-channel EEG representation and forecasting for early epileptic seizure warning. *Interdiscipl Sci Comput Life Sci.* 2025:1–12.
- [32] Song H, Chen L, Cui Y, Li Q, Wang Q, Fan J, et al. Denoising of MR and CT images using cascaded multi-supervision convolutional neural networks with progressive training. *Neurocomputing.* 2022;469(C):354–65.
- [33] Gao J, Lao Q, Liu P, Yi H, Kang Q, Jiang Z, et al. Anatomically guided cross-domain repair and screening for ultrasound fetal biometry. *IEEE J Biomed Health Inform.* 2023;27(10): 4914–25.
- [34] Gao J, Lao Q, Kang Q, Liu P, Du C, Li K, et al. Boosting your context by dual similarity checkup for in-context learning medical image segmentation. *IEEE Trans Med Imag.* 2025; 44(1):310–9.
- [35] Jiang Z, Cheng D, Qin Z, Gao J, Lao Q, Ismoilovich AB, et al. TV-SAM: increasing zero-shot segmentation performance on multimodal medical images using GPT-4 generated descriptive prompts without human annotation. *Big Data Min Anal.* 2024;7(4):1199–211.
- [36] Lin H-Y. Large-scale artificial intelligence models. *Computer.* 2022;55(5):76–80.
- [37] Qiu J, Li L, Sun J, Peng J, Shi P, Zhang R, et al. Large AI models in health informatics: applications, challenges, and the future. *IEEE J Biomed Health Inform.* 2023;27(12):6074–87.
- [38] Chang Y, Wang X, Wang J, Wu Y, Zhu K, Chen H, et al. A survey on evaluation of large language models. *ACM Trans Intell Syst Technol.* 2024;15(3):1–45.
- [39] Han K, Xiao A, Wu E, Guo J, Xu C, Wang Y. Transformer in transformer. *Adv Neural Inf Process Syst.* 2021;34:15908–19.
- [40] Moller M. Efficient training of feed-forward neural networks. In: Browne A, editor. *Neural network analysis, architectures and applications.* Boca Raton: CRC Press; 1997. p. 136–73.
- [41] Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 conference of the North*

- American chapter of the association for computational linguistics: human language technologies. 1. 2019. p. 4171–86.
- [42] Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, et al. GPT-4 technical report. 2023. Preprint at arXiv: 2303.08774.
 - [43] Yin X, Huang Y, Zhou B, Li A, Lan L, Jia Y. Deep entity linking via eliminating semantic ambiguity with BERT. *IEEE Access*. 2019;7:169434–45.
 - [44] Yenduri G, Ramalingam M, Selvi GC, Supriya Y, Srivastava G, Maddikunta PKR, et al. GPT (generative pre-trained transformer)—a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*. 2024;12:54608–49.
 - [45] Wang J, Liu Z, Zhao L, Wu Z, Ma C, Yu S, et al. Review of large vision models and visual prompt engineering. *Meta Radiol*. 2023;1(3):100047.
 - [46] Zhao L, Dong Q, Luo C, Wu Y, Bu D, Qi X, et al. DeepOmix: a multi-omics scalable and interpretable deep learning framework and application in cancer survival analysis. *Comput Struct Biotechnol J*. 2021;19:2719–25.
 - [47] Lai X, Zhou J, Wessely A, Heppt M, Maier A, Berking C, et al. A disease network-based deep learning approach for characterizing melanoma. *Int J Cancer*. 2022;150(6):1029–44.
 - [48] You Y, Lai X, Pan Y, Zheng H, Vera J, Liu S, et al. Artificial intelligence in cancer target identification and drug discovery. *Signal Transduct Targeted Ther*. 2022;7(1):156.
 - [49] You Y, Zhang L, Tao P, Liu S, Chen L. Spatiotemporal transformer neural network for time-series forecasting. *Entropy*. 2022;24(11):1651.
 - [50] Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: a survey. *ACM Comput Surv*. 2022;54(10):1–41.
 - [51] Bai Y, Geng X, Mangalam K, Bar A, Yuille AL, Darrell T, et al. Sequential modeling enables scalable learning for large vision models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2024. p. 22861–72.
 - [52] Zhang D, Yu Y, Dong J, Li C, Su D, Chu C, et al. MM-LLMs: recent advances in multimodal large language models. In: *Findings of the Association for Computational Linguistics ACL 2024*. 2024. p. 12401–30.
 - [53] Lyu C, Wu M, Wang L, Huang X, Liu B, Du Z, et al. Macaw-LLM: multi-modal language modeling with image, audio, video, and text integration. 2023. Preprint at arXiv: 2306.09093.
 - [54] Xiao M, Yang X, Yu J, Zhang L. CGIDLA: developing the web server for CpG island related density and LAUPs (lineage-associated underrepresented permutations) study. *IEEE/ACM Trans Comput Biol Bioinf*. 2020;17(6):2148–54.
 - [55] Zhang L, Bai W, Yuan N, Du Z. Comprehensively benchmarking applications for detecting copy number variation. *PLoS Comput Biol*. 2019;15(5):e1007069.
 - [56] Zhang L, Dai Z, Yu J, Xiao M. CpG-island-based annotation and analysis of human housekeeping genes. *Briefings Bioinf*. 2021;22(1):515–25.
 - [57] You Y, Tan K, Jiang Z, Zhang L. Developing a predictive platform for salmonella antimicrobial resistance based on a large language model and quantum computing. *Engineering*. 2025;48:174–84.
 - [58] Li B, Xiao X, Zhang C, Xiao M, Zhang L. DGHNN: a deep graph and hypergraph neural network for pan-cancer related gene prediction. *Bioinformatics*. 2025;41(7):btaf379.
 - [59] Xiao M, Ma F, Yu J, Xie J, Zhang Q, Liu P, et al. A computer simulation of SARS-CoV-2 mutation spectra for empirical data characterization and analysis. *Biomolecules*. 2022;13(1):63.
 - [60] Xiao M, Xiao Y, Yu J, Zhang L. PCGIMA: developing the web server for human position-defined CpG islands methylation analysis. *Front Genet*. 2024;15:1367731.
 - [61] Childers MC, Towse C-L, Daggett V. The effect of chirality and steric hindrance on intrinsic backbone conformational propensities: tools for protein design. *Protein Eng Des Sel*. 2016;29(7):271–80.
 - [62] Zhou Z, Ji Y, Li W, Dutta P, Davuluri RV, Liu H. DNABERT-2: efficient foundation model and benchmark for multi-species genomes. In: *Proceedings of the 12th International Conference on Learning Representations (ICLR 2024)*, Vienna, Austria. 2024. p. 23768–91.
 - [63] Grzankowski A, Downes SM, Forber P. LLMs are not just next token predictors. *Inquiry*. 2025:1–11.
 - [64] Verma VK, Pandey M, Jain T, Tiwari PK. Dissecting word embeddings and language models in natural language processing. *J Discrete Math Sci Cryptogr*. 2021;24(5):1509–15.
 - [65] Min B, Ross H, Sulem E, Veyseh APB, Nguyen TH, Sainz O, et al. Recent advances in natural language processing via large pre-trained language models: a survey. *ACM Comput Surv*. 2023;56(2):1–40.
 - [66] Ding R-X, Palomares I, Wang X, Yang G-R, Liu B, Dong Y, et al. Large-scale decision-making: characterization, taxonomy, challenges and future directions from an artificial intelligence and applications perspective. *Inf Fusion*. 2020;59:84–102.
 - [67] Voita E, Talbot D, Moiseev F, Titov I, Sennrich R. Analyzing multi-head self-attention: specialized heads do the heavy lifting, the rest can be pruned. In: *ACL 2019-57th annual meeting of the association for computational linguistics, proceedings of the conference*. 2020. p. 5797–808.
 - [68] Nicolson A, Paliwal KK. Masked multi-head self-attention for causal speech enhancement. *Speech Commun*. 2020;125:80–96.
 - [69] Tripathi R, Sharma P, Chakraborty P, Varadwaj PK. Next-generation sequencing revolution through big data analytics. *Front Life Sci*. 2016;9(2):119–49.
 - [70] Zhang L, Zhao J, Bi H, Yang X, Zhang Z, Su Y, et al. Bio-informatic analysis of chromatin organization and biased expression of duplicated genes between two poplars with a common whole-genome duplication. *Hortic Res*. 2021;8(1):62.
 - [71] Xiao M, Wei R, Yu J, Gao C, Yang F, Zhang L. CpG island definition and methylation mapping of the T2T-YAO genome. *Genom Proteom Bioinform*. 2024;22(2):qzae009.
 - [72] Raiaan MAK, Mukta MSH, Fatema K, Fahad NM, Sakib S, Mim MMJ, et al. A review on large language models: architectures, applications, taxonomies, open issues and challenges. *IEEE Access*. 2024;12:26839–74.
 - [73] Dozmorov MG. Epigenomic annotation-based interpretation of genomic data: from enrichment analysis to machine learning. *Bioinformatics*. 2017;33(20):3323–30.
 - [74] Artetxe M, Du J, Goyal N, Zettlemoyer L, Stoyanov V. On the role of bidirectionality in language model pre-training. In: *Findings of the association for computational linguistics: EMNLP 2022*. 2022. p. 3973–85.
 - [75] Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nat Methods*. 2021;18(10):1196–203.
 - [76] Dalla-Torre H, Gonzalez L, Mendoza-Revilla J, Lopez Carranza N, Grzywaczewski AH, Oteri F, et al. Nucleotide transformer: building and evaluating robust foundation models for human genomics. *Nat Methods*. 2025;22(2):287–97.
 - [77] Nguyen E, Poli M, Faizi M, Thomas A, Birch-Sykes C, Wornow M, et al. HyenaDNA: long-range genomic sequence modeling at single nucleotide resolution. *Adv Neural Inf Process Syst*. 2023;43:177–201.
 - [78] Wang X, Gu R, Chen Z, Li Y, Ji X, Ke G, et al. UNI-RNA: universal pre-trained models revolutionize RNA research. 2023. Preprint at bioRxiv: 2023.07.11.548588.
 - [79] Ji Y, Zhou Z, Liu H, Davuluri RV. DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. *Bioinformatics*. 2021;37(15):2112–20.

- [80] Lan AY, Corces MR. Deep learning approaches for noncoding variant prioritization in neurodegenerative diseases. *Front Aging Neurosci.* 2022;14:1027224.
- [81] Maveyraud L, Mourey L. Protein x-ray crystallography and drug discovery. *Molecules.* 2020;25(5):1030.
- [82] Bai X-C, McMullan G, Scheres SH. How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci.* 2015;40(1):49–57.
- [83] Sanjeevi M, Hebbar PN, Aiswarya N, Rashmi S, Rahul CN, Mohan A, et al. Methods and applications of machine learning in structure-based drug discovery. In: Tripathi T, Dubey VK, editors. *Advances in protein molecular and structural biology methods.* New York: Academic Press; 2022. p. 405–37.
- [84] Ali S, Chourasia P, Patterson M. When protein structure embedding meets large language models. *Genes.* 2023;15(1):25.
- [85] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–9.
- [86] Peng C-X, Liang F, Xia Y-H, Zhao K-L, Hou M-H, Zhang G-J. Recent advances and challenges in protein structure prediction. *J Chem Inf Model.* 2023;64(1):76–95.
- [87] Fang X, Wang F, Liu L, He J, Lin D, Xiang Y, et al. A method for multiple-sequence-alignment-free protein structure prediction using a protein language model. *Nat Mach Intell.* 2023;5(10):1087–96.
- [88] Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science.* 2023;379(6637):1123–30.
- [89] Elnaggar A, Heininger M, Dallago C, Rehawi G, Wang Y, Jones L, et al. ProtTrans: toward understanding the language of life through self-supervised learning. *IEEE Trans Pattern Anal Mach Intell.* 2021;44(10):7112–27.
- [90] Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, et al. De novo design of protein structure and function with RFdiffusion. *Nature.* 2023;620(7976):1089–100.
- [91] Alam R, Mahub S, Bayzid MS. Pair-EGRET: enhancing the prediction of protein–protein interaction sites through graph attention networks and protein language models. *Bioinformatics.* 2024;40(10):btac588.
- [92] Wu J, Liu B, Zhang J, Wang Z, Li J. DL-PPI: a method on prediction of sequenced protein–protein interaction based on deep learning. *BMC Bioinf.* 2023;24(1):473.
- [93] You Y, Zhou F, Yue Y, Qiu Y, Wang X, Yu Y, et al. The classical iterative HHL-based hemodynamic simulation quantum linear equation algorithm for abdominal aortic aneurysm. *Eur Phys J Spec Top.* 2024:1–11.
- [94] Dai Y, Gieseke F, Oehmcke S, Wu Y, Barnard K. Attentional feature fusion. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*; 2021. p. 3560–9.
- [95] Hendrycks D, Lee K, Mazeika M. Using pre-training can improve model robustness and uncertainty. In: *International conference on machine learning*; 2019. p. 2712–21.
- [96] Church KW, Chen Z, Ma Y. Emerging trends: a gentle introduction to fine-tuning. *Nat Lang Eng.* 2021;27(6):763–78.
- [97] Zhang L, Badai J, Wang G, Ru X, Song W, You Y, et al. Discovering hematoma-stimulated circuits for secondary brain injury after intraventricular hemorrhage by spatial transcriptome analysis. *Front Immunol.* 2023;14:1123652.
- [98] Huang H, Yang Y, Zhang Q, Yang Y, Xiong Z, Mao S, et al. S100a4+ alveolar macrophages accelerate the progression of precancerous atypical adenomatous hyperplasia by promoting fatty acid metabolism. *eLife.* 2025;13:RP101731.
- [99] Liu T, Hao M, Liu X, Zhao H. A perspective on developing foundation models for analyzing spatial transcriptomic data. *Quant Biol.* 2025;13(4):e70010.
- [100] Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol.* 2017;13(5):e1005457.
- [101] Yin W, Zhang Z, He L, Jiang R, Zhang S, Liu G, et al. ERNIE-RNA: an RNA language model with structure-enhanced representations. 2024. Preprint at bioRxiv: 2024.2003.2017.585376.
- [102] Ranwez V, Chantret N. Strengths and limits of multiple sequence alignment and filtering methods. In: *Phylogenetics in the genomic era.* 2020. p. 2.2:1–2.2:36.
- [103] Hao M, Gong J, Zeng X, Liu C, Guo Y, Cheng X, et al. Large-scale foundation model on single-cell transcriptomics. *Nat Methods.* 2024;21(8):1481–91.
- [104] Wang S, Guo Y, Wang Y, Sun H, Huang J. SMILES-BER: large scale unsupervised pre-training for molecular property prediction. In: *Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics*; 2019. p. 429–36.
- [105] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.* 2016. p. 785–94.
- [106] Wu Z, Jiang D, Hsieh CY, Chen G, Liao B, Cao D, et al. Hyperbolic relational graph convolution networks plus: a simple but highly efficient QSAR-modeling method. *Briefings Bioinf.* 2021;22(5):bbab112.
- [107] Liu Y, Ding S, Zhou S, Fan W, Tan Q. MolecularGPT: open large language model (LLM) for few-shot molecular property prediction. 2024. Preprint at arXiv: 2406.12950.
- [108] Touvron H, Martin L, Stone KR, Albert P, Almahairi A, Babaei Y, et al. Llama 2: open foundation and fine-tuned chat models. 2023. Preprint at arXiv: 2307.09288.
- [109] Zhao H, Liu S, Ma C, Xu H, Fu J, Deng Z, et al. GIMLET: a unified graph-text model for instruction-based molecule zero-shot learning. *Adv Neural Inf Process Syst.* 2023;36:5850–87.
- [110] Zheng Y, Koh HY, Ju J, Nguyen ATN, May LT, Webb GI, et al. Large language models for scientific discovery in molecular property prediction. *Nat Mach Intell.* 2025;7(3):437–47.
- [111] Li Y, Wu FX, Ngom A. A review on machine learning principles for multi-view biological data integration. *Briefings Bioinf.* 2018;19(2):325–40.
- [112] Zhang L, Xiong Z, Xiao M. A review of the application of spatial transcriptomics in neuroscience. *Interdiscipl Sci Comput Life Sci.* 2024;16(2):243–60.
- [113] Liu S, You Y, Tong Z, Zhang L. Developing an embedding, Koopman and autoencoder technologies-based multi-omics time series predictive model (EKATP) for systems biology research. *Front Genet.* 2021;12:761629.
- [114] Xiao M, Liu G, Xie J, Dai Z, Wei Z, Ren Z, et al. 2019nCoVAS: developing the web service for epidemic transmission prediction, genome analysis, and psychological stress assessment for 2019-nCoV. *IEEE/ACM Trans Comput Biol Bioinf.* 2021;18(4):1250–61.
- [115] Zhang L, Zhang L, Guo Y, Xiao M, Feng L, Yang C, et al. MCDB: a comprehensive curated mitotic catastrophe database for retrieval, protein sequence alignment, and target prediction. *Acta Pharm Sin B.* 2021;11(10):3092–104.
- [116] Zhang L, Fan S, Vera J, Lai X. A network medicine approach for identifying diagnostic and prognostic biomarkers and exploring drug repurposing in human cancer. *Comput Struct Biotechnol J.* 2022;21:34–45.
- [117] Kan A. Machine learning applications in cell image analysis. *Immunol Cell Biol.* 2017;95(6):525–30.
- [118] Zeng Y, Wei Z, Yu W, Yin R, Yuan Y, Li B, et al. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings Bioinf.* 2022;23(5):bbac297.

- [119] He B, Bergenstr hle L, Stenbeck L, Abid A, Andersson A, Borg  , et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng.* 2020; 4(8):827–34.
- [120] Jia Y, Liu J, Chen L, Zhao T, Wang Y. THltoGene: a deep learning method for predicting spatial transcriptomics from histological images. *Briefings Bioinf.* 2023;25(1):bbad464.
- [121] Mazzia V, Salvetti F, Chiaberge M. Efficient-CapsNet: capsule network with self-attention routing. *Sci Rep.* 2021;11(1):14634.
- [122] Guo L, Qiu T, Wang J. ViTScore: a novel three-dimensional vision transformer method for accurate prediction of protein–ligand docking poses. *IEEE Trans NanoBioscience.* 2023; 22(4):734–43.
- [123] Gokhale M, Mohanty SK, Ojha A. GeneViT: gene vision transformer with improved DeepInsight for cancer classification. *Comput Biol Med.* 2023;155:106643.
- [124] Kobak D, Berens P. The art of using *t*-SNE for single-cell transcriptomics. *Nat Commun.* 2019;10(1):5416.
- [125] Dabba A, Tari A, Meftali S, Mokhtari R. Gene selection and classification of microarray data method based on mutual information and moth flame algorithm. *Expert Syst Appl.* 2021; 166:114012.
- [126] Chen T, Abadi AJ, L  Cao K-A, Tyagi S. Multiomics: a user-friendly multi-omics data harmonisation R pipeline. *F1000Research.* 2023;10:538.
- [127] Lahat D, Adali T, Jutten C. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proc IEEE.* 2015;103(9):1449–77.
- [128] Wang X, Chen G, Qian G, Gao P, Wei X-Y, Wang Y, et al. Large-scale multi-modal pre-trained models: a comprehensive survey. *Mach Intell Res.* 2023;20(4):447–82.
- [129] Gaw N, Yousefi S, Gahrooei MR. Multimodal data fusion for systems improvement: a review. *IJSE Trans.* 2021;54(11): 1098–116.
- [130] Chen C-FR, Fan Q, Panda R. CrossViT: cross-attention multi-scale vision transformer for image classification. In: *Proceedings of the IEEE/CVF international conference on computer vision.* 2021. p. 347–56.
- [131] Wei X, Zhang T, Li Y, Zhang Y, Wu F. Multi-modality cross attention network for image and sentence matching. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 2020. p. 10938–47.
- [132] Zhang Y, Yang Q. A survey on multi-task learning. *IEEE Trans Knowl Data Eng.* 2021;34(20):5586–609.
- [133] Meyer J, Eitel A, Brox T, Burgard W. Improving unimodal object recognition with multimodal contrastive learning. In: *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS).* 2020. p. 5656–63.
- [134] Li B, Xiao M, Zeng R, Zhang L. Developing a multiomics data-based mathematical model to predict colorectal cancer recurrence and metastasis. *BMC Med Inf Decis Making.* 2025; 25(Suppl 2):188.
- [135] Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Med Res Methodol.* 2018;18(1):24.
- [136] Lee C, Zame W, Yoon J, Van Der Schaar M. DeepHit: a deep learning approach to survival analysis with competing risks. In: *Proceedings of the AAAI conference on artificial intelligence,* vol. 32(1). 2018. p. 2314–21.
- [137] Zvyagin M, Brace A, Hippe K, Deng Y, Zhang B, Bohorquez CO, et al. GenSLMs: genome-scale language models reveal SARS-CoV-2 evolutionary dynamics. *Int J High Perform Comput Appl.* 2022;37(6):683–705.
- [138] Yang X, Liu G, Feng G, Bu D, Wang P, Jiang J, et al. GeneCompass: deciphering universal gene regulatory mechanisms with knowledge-informed cross-species foundation model. *Cell Res.* 2024;34(12):830–45.
- [139] Rosen Y, Brbi  M, Roohani Y, Swanson K, Li Z, Leskovec J. Toward universal cell embeddings: integrating single-cell RNA-seq datasets across species with SATURN. *Nat Methods.* 2024;21(8):1492–500.
- [140] Jha K, Saha S, Karmakar S. Prediction of protein-protein interactions using vision transformer and language model. *IEEE/ACM Trans Comput Biol Bioinf.* 2023;20(5):3215–25.
- [141] Jha K, Saha S. Analyzing effect of multi-modality in predicting protein-protein interactions. *IEEE/ACM Trans Comput Biol Bioinf.* 2022;20(1):162–73.
- [142] Hayes T, Rao R, Akin H, Sofroniew NJ, Oktay D, Lin Z, et al. Simulating 500 million years of evolution with a language model. *Science.* 2025;387(6736):850–8.
- [143] Ingraham JB, Baranov M, Costello Z, Barber KW, Wang W, Ismail A, et al. Illuminating protein space with a programmable generative model. *Nature.* 2023;623(7989):1070–8.
- [144] Boitreaud J, Dent J, McPartlon M, Meier J, Reis V, Rogozhnikov A, et al. Chai-1: decoding the molecular interactions of life. 2024. Preprint at bioRxiv: 2024.2010.2010.615955.
- [145] Buttenschoen M, Morris GM, Deane CM. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem Sci.* 2023; 15(9):3130–9.
- [146] Chen X, Teichmann SA, Meyer KB. From tissues to cell types and back: single-cell gene expression analysis of tissue architecture. *Annu Rev Biomed Data Sci.* 2018;1:29–51.
- [147] Min W, Shi Z, Zhang J, Wan J, Wang C. Multimodal contrastive learning for spatial gene expression prediction using histology images. *Briefings Bioinf.* 2024;25(6):bbae551.
- [148] Xie R, Pang K, Chung S, Perciani C, MacParland S, Wang B, et al. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. *Adv Neural Inf Process Syst.* 2023;36:70626–37.
- [149] Zhang Y, Liu Y, Wang Z, Wang M, Xiong S, Huang G, et al. Uncovering the relationship between tissue-specific TF-DNA binding and chromatin features through a transformer-based model. *Genes.* 2022;13(11):1952.
- [150] H rst F, Rempe M, Heine L, Seibold C, Keyl J, Baldini G, et al. CellViT: vision transformers for precise cell segmentation and classification. *Med Image Anal.* 2024;94:103143.
- [151] Zhou Y, Li W, Yang G. SCTS: instance segmentation of single cells using a transformer-based semantic-aware model and space-filling augmentation. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision.* 2023. p. 5944–53.
- [152] Pham V-T, Vu N-T, Pham V-T, Tran T-T. CellFormer: a DaViT transformer-based method for nuclei segmentation. In: *2023 RIVF international conference on computing and communication technologies (RIVF).* 2023. p. 165–70.
- [153] Zeng S, Wang D, Jiang L, Xu D. Parameter-efficient fine-tuning on large protein language models improves signal peptide prediction. *Genome Res.* 2024;34(9):1445–54.
- [154] Alser M, Lindegger J, Firtina C, Almadhoun N, Mao H, Singh G, et al. From molecules to genomic variations: accelerating genome analysis via intelligent algorithms and architectures. *Comput Struct Biotechnol J.* 2022;20:4579–99.
- [155] Nandamuri S. AI at scale: the infrastructure revolution enabling GPT-class large language models. *J Comput Sci Technol Stud.* 2025;7(4):321–8.
- [156] Liu B, Li Y, Zhang L. Analysis and visualization of spatial transcriptomic data. *Front Genet.* 2022;12:785290.
- [157] Xu H, Xu Q, Cong F, Kang J, Han C, Liu Z, et al. Vision transformers for computational histopathology. *IEEE Rev Biomed Eng.* 2024;17:63–79.
- [158] Zhou T. Feature fusion and latent feature learning guided brain tumor segmentation and missing modality recovery network. *Pattern Recogn.* 2023;141:109665.

- [159] Wang H, Luo S, Hu G, Zhang J. Gradient-guided modality decoupling for missing-modality robustness. *Proc AAAI Conf Artif Intell.* 2024;38(14):15483–91.
- [160] Zhang Y, Peng C, Wang Q, Song D, Li K, Kevin Zhou S. Unified multi-modal image synthesis for missing modality imputation. *IEEE Trans Med Imag.* 2025;44(1):4–18.
- [161] Zhou T, Liu M, Thung KH, Shen D. Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. *IEEE Trans Med Imag.* 2019;38(10):2411–22.
- [162] Wang Z, Luo P, Xiao M, Wang B, Liu T, Sun X. Recover then aggregate: unified cross-modal deep clustering with global structural information for single-cell data. *Briefings Bioinf.* 2024;25(6):bbae485.
- [163] Chien AA. GenAI: Giga\$\$\$, TeraWatt-Hours, and GigaTons of CO₂. *Commun ACM.* 2023;66(8):5.
- [164] Rohanian O, Nouriborji M, Kouchaki S, Clifton DA. On the effectiveness of compact biomedical transformers. *Bioinformatics.* 2023;39(3):btad103.
- [165] Fang J, Zhu Z, Li S, Su H, Yu Y, Zhou J, et al. Parallel training of pre-trained models via chunk-based dynamic memory management. *IEEE Trans Parallel Distr Syst.* 2023;34(1):304–15.
- [166] Nguyen TV, Dakka MA, Diakiw SM, VerMilyea MD, Perugini M, Hall JMM, et al. A novel decentralized federated learning approach to train on globally distributed, poor quality, and protected private medical data. *Sci Rep.* 2022;12(1):8888.
- [167] Kramer D. Will AI's growth create an explosion of energy consumption? *J Physics Today.* 2024;77(4):28–9.

How to cite this article: Li M, Shang Q, Dong Z, You Z, Zhang L, Xiao M. Applications of large-scale artificial intelligence models in bioinformatics. *Quantitative Biology.* 2026; e70026. <https://doi.org/10.1002/qub2.70026>