https://doi.org/10.55453/rjmm.2026.129.6.2

Review

AI in Drug Discovery and Precision Medicine

Richa Sharma[1,*], Richa Saxena[2], Rakshith Maneshwar Hoskeri[3], Ekadashi Rajni[1], Teju Lal Chaurasia[4]

[1] Department of Microbiology, Mahatma Gandhi University of Medical Sciences & Technology, Jaipur, Rajasthan, India
[2] Department of Biotechnology, Invertis University, Bareilly- 243001, U.P., India
[3] Department of Clinical Psychology, Mahatma Gandhi University of Medical Sciences & Technology, Jaipur, Rajasthan, India
[4] School of Allied Health Sciences, Jaipur National University, Jaipur, Rajasthan, India

*Correspondence: Dr. Richa Sharma, email: richa.phd.15@gmail.com

Abstract: Biological data have grown exponentially due to single-cell multi-omics, high-content imaging, and large-scale clinical records; this has transformed modern biology. The landscape has shifted to one of data abundance, and the current approaches to computational analysis are not up to the task of dealing with the scale and complexity of this change. The information in big biological data can be characterized by the volume it occupies, velocity, variety, veracity, and value, and each of these features presents its own challenges to the researcher. Artificial intelligence (AI), including machine learning (ML), deep learning (DL), and natural language processing (NLP), has progressively turned into a requirement in tackling these issues. This chapter discusses how AI can be used to curate, represent, learn, and model the analysis of big biological data. Questions include data curation via methods to compress, represent, and outline big biological data, how to learn representations, how to model with big biological data, and how to translate into clinical problems. Using the conversation of genomics, proteomics, transcriptomics, structural biology, imaging, and electronic health records, this chapter reveals how AI is transformative in bioinformatics. In conclusion, the chapter points to enduring challenges, namely interpretability, data governance, and reproducibility, and provides a step into the future of explainable, trustworthy, and reproducible AI in biology.

Keywords: artificial intelligence; machine learning; bioinformatics; clinical genomics; reproducibility; benchmarking

INTRODUCTION

Biology is a field that has been submerged in a major change in the last several decades. No longer characterized by hypothesis-based investigations that involved small-scale experiments, it is now influenced by the flood of data that high-throughput technologies and large-scale biomedical projects are generating. Experiments such as whole-genome sequencing projects of entire populations, single-cell atlases, mapping millions of cells, multi-modal imaging of tissues and organisms, and the digitization of clinical and health records mean that data measurement has to be in terms of terabytes to petabytes. This paradigm requires not only different analytical methods but a totally different way of thinking about the issues at hand (artificial intelligence (AI)) in which the means of finding meaning in complexity is created [1]. The real meaning of this challenge can be summed up in a term coined as big biological data. The amount of data being handled is enormous, and in most cases, it surpasses the capacity of customary data storage and processing environments. The frequency of data creation- whether it be a continuous drive of sequencing reads or streaming clinical notes- necessitates real-time or near-real-time processing. The diversity of biological modalities, encompassing DNA and RNA sequences, protein structure, matrices of expression, clinical time series, and even unstructured biomedical prose, is the source of diversity [2]. Making this even more difficult is veracity, since biological data is erratic and/or incomplete and inconsistent, with batch effects, missing values, and annotation errors. The utility of these data rests on the capacity to coordinate, combine, and transform this data into knowledge that may be used to move fundamental science and translational medicine. However, these challenges have a unique solution in AI, which can learn complicated patterns, generalize across domains, and adapt to a new task. The importance of AI in bioinformatics can only be appreciated in terms of the analyses that it facilitates. Traditional pipelines were

Citation: Sharma R, Saxena R, Hoskeri RM, Rajni E, Chaurasia TL. *AI in Drug Discovery and Precision Medicine.* R. J. Mil. Med. 2026, CXXIX(6): 600-609. https://doi.org/10.55453/rjmm.2026.129.6.2

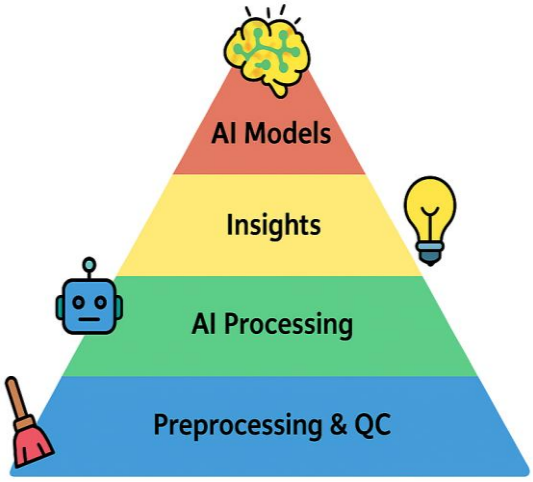
Academic Editor: Dan Spinu

Received: 06 October 2025
Revised: 22 July 2026
Accepted: 28 August 2026

also highly dependent upon manually designed heuristics, hard-coded thresholds, and statistical models of limited scope that were unable to keep up with the sophistication of contemporary data.

AI can make discoveries about the underlying structures and relationships directly based on the data, forming representations that reflect these relationships. In genomics, AI has been used to produce precise variant calling and prediction of pathogenicity, and in transcriptomics, deep learning models have helped to reveal latent states in cells and developmental processes [3]. AlphaFold models, such as the first RCSB PDB-equipped protein of a flexible type, transformed the field of structural biology by demonstrating that AI-directed protein-folding prediction was possible with higher accuracy compared to previous methods. In clinical practice, AI is useful in patient stratification, risk modeling, and individualized treatment plans with the incorporation of genomic, phenotypic, and electronic health records. Significantly, AI is not just about prediction but also about scale and integration. Modern AI systems have a layered structure: data layers that support the findability, accessibility, interoperability and reusability (FAIR principles) of data; engineering layers that deal with data management, onto which data is ingested, how to handle data quality, reconciliation, and how to manage the data; computational layers that rely on both distributed training and high-performance computing infrastructures and modeling layers that focus on predictive, generative and multi-modal learning. The bio-AI stack: this hierarchical reasoning, commonly called the bio-AI stack, acknowledges the fact that biological insights are achieved through synergistic interactions of data engineering, complex modeling, and systems-level reasoning (Fig.1). Lastly, there are substantial ethical, legal, and social concerns regarding the incorporation of AI in bioinformatics. With human-derived data, issues of privacy, consent, algorithmic bias, or the fair distribution of any resulting benefits are of the essence. New strategies like federated learning and differential privacy are trying to develop a compromise between AI utility and protection of sensitive data [4]. To prevent the arrival of a world in which AI is used to play God, we must have governance frameworks, clear reporting practices, and explicable models that ensure that AI does not just move the field of biological discovery forward but does so responsibly and sustainably [5].

Figure 1: Workflow of AI in Bioinformatics



MATERIALS AND METHODS

This study is a narrative review that synthesizes recent advances in artificial intelligence applications in drug discovery and precision medicine. A comprehensive literature search was conducted to identify relevant studies on the application of artificial intelligence in drug discovery and precision medicine.

Databases Searched: The search was performed across four electronic databases: PubMed, Scopus, Web of Science and Google Scholar. The search was conducted across four electronic databases to ensure broad, multidisciplinary coverage. PubMed was selected as the primary source for biomedical and life sciences literature, ensuring access to foundational research in medicine and pharmacology. Scopus and Web of Science were included as large, interdisciplinary databases covering the intersection of computer science, engineering, and biomedical research, providing high-impact, peer-reviewed sources. Google Scholar was incorporated to capture cutting-edge developments in artificial intelligence and machine learning, including preprints and conference papers that may not yet be indexed in traditional databases but represent the latest methodological advancements.

Timeframe for Review Study: The timeframe was limited to articles published between January 2021 and December 2025 to capture the most recent advancements in this rapidly evolving field in the last 5 years. This five-year window was strategically chosen to focus

on the most recent advancements in this rapidly evolving field, capturing the emergence of transformative technologies such as deep learning, generative models, and large language models, while ensuring the findings remain current and relevant.

Keywords for Search: The search strategy utilized a combination of keywords, including Artificial Intelligence, Virtual Screening, Bioinformatics, Drug Discovery, Protein Structure, Precision Medicine, Personalized Medicine, and Target drug interaction. These were employed for a review of the literature at the intersection of AI methodologies and their biomedical applications.

Selection Approach: The selection process prioritized peer-reviewed original research articles, conference papers, and relevant preprints that focused on the application of AI methodologies within drug discovery and precision medicine. Peer-reviewed articles and conference papers were included to ensure a foundation of validated, high-quality research, while preprints were incorporated to reflect the latest innovations and emerging trends in AI development. Studies were included if they demonstrated a clear focus on AI applications, such as target identification, drug design, patient stratification, or biomarker discovery. Articles were excluded if they were opinion pieces, editorials, or lacked a substantive connection to the specified domains.

Evidence Synthesis: Evidence from the included studies was synthesized using a thematic narrative approach. Major themes, technological advancements, clinical applications, limitations, and future directions were systematically identified and organized.

RESULTS

Data Management and Engineering Foundations

Bioinformatics and the effective use of artificial intelligence lie in the efficient processing of big biological data prior to getting into the modeling process. Data management in this scenario cannot be merely reduced to storage: it is much more about the systematic intake, strict quality monitoring and assurance, harmonization of various modalities, and reproducibility of the work by means of standard workflows [5]. Biological data can be extremely heterogeneous, and genomics, transcriptomics, proteomics, metabolomics, imaging, and clinical information may be collected using different platforms and standards. The non-uniformity may cause biases, impair integration, and make downstream analysis more complicated. Hence, data management and data engineering are the basis of bioinformatics with AI [6]. The first challenge is data ingestion, in which raw results of high-throughput technologies are transformed into structured data. Such a process usually entails the derivation of relevant metadata, their compatibility with other existing schemas, and storing them in repositories that comply with the FAIR principles of data, which in this case are Findable, Accessible, Interoperable, and Reusable. The ingestion pipelines should also be able to cope with this velocity to support real-time sequencing results or a patient whose electronic health records are constantly being updated [7]. To meet these demands, cloud-based infrastructures and distributed file systems have been gaining popularity recently because they are scalable solutions to large volumes of data and are collaboratively accessible.

Following ingestion, intense quality control (QC) is required. Biological measurements may be burdened with batch effects, sequencing errors, incomplete values and annotation inconsistencies [8]. Unless provided with systematic QC, AI models will suffer spurious learning or deliver unreliable results. Outlier detection, statistical normalization and missing values imputation are regularly deployed, whereas mounting AI techniques like autoencoders and generative adversarial networks can be successfully utilized to resolve noise and repair data. Vetting truth at this level has a direct bearing on the accuracy of subsequent modeling. Another basis of data management is harmonization and integration. Multi-omics studies such as those that combine genomic, transcriptomic and proteomic data into a single analysis require the combination of disparate data sets of different modalities into a common data environment [9]. This integration is hampered by the fact that measurement scales, noise profiles, and metadata conventions differ. Ontologies and controlled vocabularies - e.g., the Gene Ontology (GO), or Human Phenotype Ontology (HPO) - are crucial to the alignment of heterogeneous data. More imminently, graph-based models and knowledge graphs have been used to model biological entities and relations, and provide more sound data integration. Reproducibility is also of essence. With the advent of big data, workflows may include tens of computational steps, and without thorough documentation, websites creating results may be very difficult to recreate [10]. This is one of the reasons why workflow management systems like Nextflow, Snakemake and Galaxy have become popular; such systems can automate pipelines, ensure version control, and are portable across different computational environments. Such reproducibility is further ensured through the use of containerization tools such as Docker and Singularity, since they can be used to confine software dependencies and help to standardize analyses to make them easy to reproduce across systems.

Machine Learning and Deep Learning in Bioinformatics Applications

ML and DL have become essential in the new bioinformatics, providing highly sophisticated computational options to handle, process and understand the increasing and multi-layered complexity of biological data. The classical statistical tools have the power to analyze big data; however, they are generally unable to cope with vast dimensionality, heterogeneity, and non-linear shapes of omics data, imaging information, and molecular structures [11]. DL and ML can help dissect these difficulties through learning patterns directly from data, which would allow the implementation of predictive modeling, feature extraction, and novelty discovery of unknown relationships between biological entities. Random forests, support vector machines, and gradient boosting ML algorithms find wide use in genomics, where they can be used to predict the expression of genes, call genetic variants, and identify disease-related

mutations. The convolutional neural networks (CNNs) have been shown to be particularly effective in the analysis of next-generation sequencing (NGS) data and in genome-wide association studies (GWAS) [12]. As one example, CNNs are applied to the prediction of transcription factor binding sites, enhancers, and epigenetic modifications on raw DNA sequences. The potential associated with such capabilities is to transform gene regulation and functional genomics studies. In proteomics and structural biology, AlphaFold has become an important tool in protein structure prediction and has reached accuracy resolution near that of experimental results. This paradigm shift not only helps understand the protein folding mechanism better, but also helps expedite drug discovery and protein engineering [13]. Learned models such as recurrent neural networks (RNNs) and transformers have also demonstrated improvements in sequence-based prediction of protein-protein interactions, post-translational modifications, and enzyme-substrate specificities, the basis of systems biology and drug design.

The second broad use case is in biomedical imaging: ML and DL models can be used to analyze large numbers of microscopy images, histopathological slides, and medical scans. Computer-assisted image recognition techniques are fast approaching the level of accuracy in classifying cell types, measuring morphological characteristics of tissues, and even in the diagnosis of diseases like cancer as accurate as a human pathologist [14]. A combination of imaging data and the omics profile adds more dimensions to personalized medicine by detecting the disease mechanism more comprehensively. However, with these successes, there are impediments to the use of ML and DL in bioinformatics. Biological datasets are usually noisy, imbalanced, and even small compared to typical computer vision or natural language processing datasets. This poses a question of overfitting, reproducibility, and generalizability of the models. Moreover, many DL algorithms have a black box, which makes biological interpretation difficult, and solutions are needed to create explainable AI (XAI) methods. To overcome these drawbacks, hybrid models that allow domain knowledge to marry with data-driven learning are being actively explored. To conclude, ML and DL make bioinformatics transformational as they allow finding predictive, scalable, and interpretable models of large bioinformatics data. Through these techniques, we get to know much more about basic biological processes, and they also lead to new innovations in the field of diagnosing illness, pharmaceuticals, and precision medicine [14]. With coming computational sophistication as well as hardware and software capabilities, machine learning and deep learning are poised to make even greater inroads into the world of bioinformatics, bringing the data of biology closer to productive biomedical uses.

AI in Omics Data Analysis

Omics, including genomics, transcriptomics, proteomics, metabolomics, and epigenomics, produce huge and complex datasets that are at center stage in modern biology and medicine. Artificial intelligence (AI) has become an effective method of analyzing this high-dimensional data as it provides methods that extend beyond traditional statistics. With its capacity to integrate different data types and learn patterns in them, AI allows novel biomarker discovery, disease signatures and therapeutic targets [7]. In the genomic world, the use of AI-related applications has incredibly enhanced the processing of next-generation sequencing (NGS) information. Machine learning and deep learning algorithms can be used in alignment of sequencing reads, calling of variants, and predicting functional effect of mutations using their high speed. As an example, AI models are also employed to differentiate pathogenic variants and non-pathogenic variants that can help to diagnose hereditary disorders. What is more, the AI is utilized in genome-wide association studies (GWAS) to compute genetic variants that are related to complex traits and can provide new insights into susceptibility to diseases. AI applications also have immense benefits in transcriptomics. The large amounts of data generated using RNA sequencing reflect dynamic patterns of gene expression, frequently at time-point snapshots, in tissues or disease conditions [15]. I models can cluster expression profiles efficiently, predict gene regulatory networks, and identify differentially expressed genes that can be linked to disease progression. NLP-based models are also being used to learn rules that define the likelihood of RNA splicing events, microRNA targets, and novel non-coding RNAs with functional roles in disease, just as they have been used with sequences such as text (Table 1).

In proteomics, the incorporation of AI has given a new form of studying protein expression and interactions. Much higher sensitivity and specificity can be achieved as a result of analyzing mass spectrometry data using technology based on AI [16-17]. Transformers Deep learning techniques are granting predictions for protein structure, post-translational modifications, and the structure of protein-protein interactions. These observations are vital insights in cell signaling networks, pathogen interactions and disease mechanisms. Metabolomics is difficult because metabolites are diverse and chemically complex [18]. The most frequently applied AI models, particularly those based on neural networks, are currently used to categorize metabolite spectra, identify unknown compounds, and integrate metabolomic data with genomic and proteomic data. The integrated analysis has also contributed to a better detection of metabolic biomarkers of diseases like cancer/diabetes as well as cardiovascular disorders. One of the main advantages of AI in omics is that it allows combining multi-omics data and provides a systems-level analysis of biological processes [19]. Conventional methods tended to study each of these omics layers independently, whereas AI enables the combination of them into a unified framework. Such an integrative analysis can illustrate cross-talk among biological layers, e.g., how genetic variants generate changes to the transcriptome, which produces protein expression and metabolite changes. Such insights are vital in conceiving precision medicine and in the field of systems biology. Despite these, there are still some challenges in using AI in omics. Analysis can be problematic due to data heterogeneity, batch effects, and missing data. The interpretability of the AI models is an issue because in biology, one needs to be able to justify biological conclusions with transparent reasoning that can be experimentally tested [20]. Further, both the noise and scale, as well as the different data structures used by different omics, are vital components of multi-omics integration.

Nonetheless, explainable AI research, powerful data harmonization techniques, and high-volume and collaborative initiatives are most likely to resolve these limitations.

Table 1: Applications of AI in Omics Data Analysis

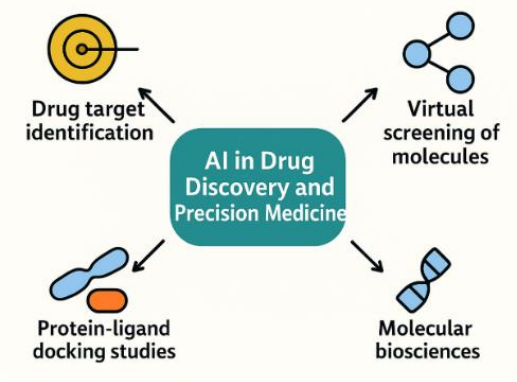
Omics Field	AI Application	Example Tools/Models
Genomics	Variant calling, mutation prediction, genome annotation	DeepVariant, AlphaFold
Transcriptomics	Gene expression profiling, splicing event prediction	Random Forest, CNN models
Proteomics	Protein structure prediction, biomarker discovery	AlphaFold, DeepCNF
Metabolomics	Pattern recognition, disease classification	SVM, Neural Networks
Epigenomics	Methylation pattern analysis, chromatin accessibility prediction	Gradient Boosting, Deep Learning

AI in Drug Discovery and Precision Medicine

Artificial intelligence (AI) has demonstrated a transformative potential in two areas: drug discovery and precision medicine. The conventional drug discovery method is lengthy, wastes many resources, and has poor rates due to only a small percentage of potential drugs making it to clinical application. With the emerging AI-based methodologies, this process has already started to be revolutionized, shortening drug target identification, predicting drug and compound interactions, and leveraging lead molecules with better efficacy and safety profiles [21]. The prediction of drug-target interactions and protein structures is one of the great contributions of AI to drug discovery. AI has enabled deep learning models to make significant advances, such as AlphaFold, which predicts protein 3D conformations with extraordinary accuracy [22]. Such structural predictions can give important hints to the possible binding pockets, allowing rational design of small molecules or biologics. Also, AI algorithms are capable of filtering immense chemical databases to find the ones that match the desired chemical structure, which is quite faster and cheaper than carrying out high-throughput screening in wet labs. The sorting of candidate molecules expected to interact best with target proteins using AI-powered screening leads to reduced drug development pipeline efficiency [23]. Other than identifying targets and predicting whether they will bind, AI has also been successful in lead optimization. With the modeling of the pharmacokinetic and pharmacodynamic (PK/PD) attributes, AI algorithms can predetermine the absorption, distribution, metabolism, excretion, and toxicity (ADMET) of compounds. This is useful in the elimination of molecules that possess undesired properties at initial stages, minimizing late-stage failures. With the development of generative models, including generative adversarial networks (GANs) and reinforcement learning systems, novel chemical entities are being generated with a view to not only meeting efficacy requirements but also safety and manufacturability constraints.

In precision medicine, AI is enabling an increasingly personalized approach in which therapies are individually selected on the basis of their genetic, molecular, and clinical characteristics [24]. With the multi-omics data combined with clinical histories and real-life evidence, AI can divide patients into subpopulations with differing reactions to treatments. AI models will enable investigators to determine which immunotherapies or targeted therapies are likely to work for cancer patients, thereby limiting trial-and-error approaches to treatment selection [23]. This stratification not only enhances patient outcomes but also reduces healthcare costs from unnecessary side effects. It is also being used to determine biomarker discovery, which is key to precision medicine. By analyzing large genomic and proteomic data sets, AI systems are able to discover biomarkers predictive of drug response, disease progression, or risk of relapse. These biomarkers are essential to making clinical judgments and to the creation of companion diagnostics that match treatment to patient-specific traits. Although it has a positive future, the integration of AI into drug discovery and precision medicine is not without difficulties. Annotated data is the key to training successful models, but it is usually hindered by privacy implications, proprietary policies, and differences in data standards [25]. Additionally, the explanation of AI models is a topical concern, where clinicians and regulators demand clear and understandable decision-making to keep patients safe (Fig.2). Ethical concerns, including data privacy, algorithmic unfairness, and fair access to AI-enabled treatment, also have to be accommodated. AI is transforming pharmaceutical research and precision medicine to make development in therapeutics faster, more affordable, and more personalized. Ranging between structural biology and virtual screening, all the way to new biomarker discovery and patient stratification, AI-driven advancements are opening the doors toward the next age of healthcare.

Figure 2: AI in Drug Discovery and Precision Medicine



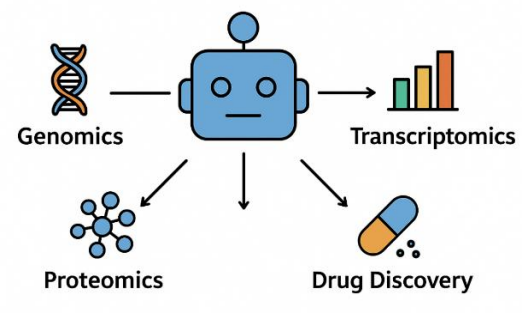
AI in Medical Imaging and Diagnostics

The diagnostic healthcare industry is one of the most pronounced and influential areas of artificial intelligence (AI) in the medical field today. Radiology, pathology, and clinical diagnostics collect image data, patient data, and other data every day in vast amounts, and AI-driven technologies have shown great promise in interpreting this data quickly and efficiently. In contrast to traditional manual data analysis, which is time-consuming and prone to human error, AI models based on deep learning are much faster, providing the ability to analyze thousands of images and identify trends that could not be noticed even by a skilled clinician [26]. In radiology, the use of AI algorithms is at its peak, where they are used to identify and categorize anomalies in X-rays, computed tomography (CT), magnetic resonance imaging (MRI), and ultrasonic images. As an example, CNNs have been trained to detect early symptoms of cardiovascular abnormalities, brain tumors, and lung cancer with accuracy levels superior to or equal to those of professional radiologists. In addition to marking suspicious areas, such systems also give quantitative measures that help clinicians make more assured decisions more swiftly. In pursuit of the same goals, AI is also being used in mammography to screen for breast cancer; this is particularly important because early and proper screening results in a high chance of patient survival [27]. There has also been a paradigm shift in pathology with the emergence of AI. Histological samples on slides can be scanned in high resolution using whole-slide imaging/digital pathology systems, generating terabytes of data that can be amenable to computational analysis. Pathway prediction: Neuro AI-powered pathology applications can categorize tissue type, identify malignant cells, and measure biomarkers within the tumor microenvironment at a very high degree of precision [28]. These capabilities are essential in the field of oncology, where precise staging and grading have a direct impact on treatment design and prognosis. By eliminating time-consuming steps, AI liberates pathologists to work on challenging diagnostic cases and decreases inter-observer variance.

In addition to imaging applications, AI is now transforming laboratory testing through interpretation of the rich data produced in genomics, proteomics, and clinical laboratory assays (Fig.3). Machine learning algorithms can be used to capture a multitude of patient information—such as lab results, medical history, and lifestyle data—and piece it together in order to make a complete diagnostic picture [29]. One particularly useful application is the fact that AI tools are now being used to help predict the risk of sepsis presenting in inpatient services by constantly monitoring electronic health record (EHR)-available data and intervening at the appropriate time, potentially saving lives. Some of the additional contributions that AI makes to precision diagnostics are that AI provides like-and-like comparisons. Diagnostic algorithms can also stratify patients by minute phenotypic dissimilarity, and this can facilitate targeted therapies. As another example, neurodegeneration, including Alzheimer's, comes at a time when MRI and PET scans using AI models can reveal early biomarkers before clinical symptoms come through.

The possibility of chain reaction prevention exists with early detection that will lead to the prevention of intervention and better management of the patient. Even though AI is successful in medical imaging and diagnostics, there are challenges related to its adoption. Data confidentiality, normalization of imaging models and incorporation of AI systems into actual clinical processes are among the topical issues [30]. Further, surveillance of diagnostic AI tools by regulatory bodies needs to have stringent validation as a means of ensuring accuracy, reproducibility, and fairness. Black-box models are vulnerable to scrutiny of interpretation, which makes clinicians reluctant to accept the outputs of the algorithms without any intervention by a human. Ethical aspects such as potential bias in algorithms that can result in healthcare inequality, based on demographic groups, must also be considered. In summary, AI has become a game-changer in the field of medical imaging and diagnostics to advance accuracy, efficiency, and personalization of health [31]. Applications include radiology and pathology, laboratory diagnostics, and predictive models. Clinical applications: In radiology and pathology, AI is used to aid the detection of diseases earlier and in monitoring progression more accurately, with AI-driven systems also being used in laboratory diagnostics and predictive modeling to deliver specifically targeted interventions.

Figure 3: AI in Omics and Bioinformatics



Ethical Imperatives and Emerging Directions for AI in Bioinformatics

On the one hand, the emergence of artificial intelligence in bioinformatics and healthcare systems has provided new and outstanding prospects in the regulation and analysis of biological big data (Table 2). On the other hand, however, there is indeed a complicated array of challenges in such assimilation. Such obstacles cut across the technical, ethical, and socio-regulatory dimensions and require comprehensive solutions until AI can reach mass and reliable use. An awareness of such concerns is essential to determining the course to be taken by AI in life sciences. Among the biggest obstacles are data quality and heterogeneity. Biological data are produced by a variety of sources, which include genomic sequencing, transcriptomics, proteomics, electronic health record systems, and imaging platforms [32]. These data are usually not homogeneous, do not scale up well, and have inconsistent reliability. Missing values, noise, and biases that occur during sample collection or sequencing methods can adversely affect the performance of the model. The need to provide common data exchange formats and standardized exchange across platforms is a critical means of ensuring that data-driven analysis using AI can be maximized. The other obstacle is algorithm transparency and interpretability. Some of the most effective AI models, especially deep learning neural networks, are so powerful and opaque that little is understood about their decision-making mechanisms. In healthcare, where the actions of clinicians may have a direct effect on patient well-being, this unpredictability in the absence of explainability erodes trust [33]. There is a need to develop interpretable AI models and integrate explainable AI (XAI) frameworks, whereby clinicians and researchers know not only that a particular decision is predicted, but why it is predicted. The issues of moral inclinations also make the landscape more tricky. The privacy and security of data are paramount, especially because genomic and medical information is sensitive in nature.

Ensuring that the global regulations, including GDPR (General Data Protection Regulation) or HIPAA (Health Insurance Portability and Accountability Act), are adhered to is not a trivial task in the case of large-scale datasets that are being processed across borders [34]. In addition, an algorithm can be biased, i.e., it can reinforce unfair treatment in healthcare due to the underrepresentation of certain ethnic, gender, or socioeconomic groups in the training set. As an example, an AI may be developed based on mostly European genomic data and perform poorly when used in underrepresented groups, resulting in errors in diagnosis. There is still a major hurdle in terms of regulatory implications involved in the validation and approval of AI tools to be used in the clinical setting. Regulatory agencies require evidence of reproducibility, generalizability, and safety at a high level before algorithms can be included in the flow of diagnostics or treatment [35]. Such a high degree of validation would demand large and disparate datasets, multinational cooperation between industry and science, and universities, and hospitals. Looking into the future, it is anticipated that AI in bioinformatics is going to shift from predictive analytics to prescriptive and generative AI. The rise of new paradigms like federated learning could help alleviate privacy issues since it means that AI models can be trained using decentralized data without data being shared between organizations. In the same way, the combination of quantum computing with artificial intelligence has the potential to make the analysis of ultra-large biological data, e.g., generated by single-cell sequencing, faster. A second emerging opportunity is the bringing of AI to systems biology, in which global models of biological networks can promise a richer understanding of disease processes and drug targets [36]. What is also significant is the creation of ethical frameworks and governance models that will bring balance between innovation and responsibility. Cross-sector partnerships (with biologists, clinicians, data scientists, ethicists, and policymakers) will be paramount to ensure that the merits of AI-based bioinformatics are equally distributed among all world populations. Training of researchers and clinicians in the use of AI tools will also be instrumental in how they will be effectively and usefully adopted (Table 3).

Table 2: AI Techniques and Their Role in Bioinformatics

AI Technique	Role in Bioinformatics
Machine Learning (ML)	Classification of gene expression data, biomarker discovery
Deep Learning (DL)	Protein structure prediction, drug-target interaction
Natural Language Processing (NLP)	Mining biomedical literature, gene–disease association prediction
Reinforcement Learning (RL)	Optimization in drug design and molecular docking simulations
Graph Neural Networks (GNN)	Modeling biological networks and pathway interactions

Table 3: Comparison of Traditional and AI-Driven Approaches in Bioinformatics

Aspect	Traditional Methods	AI-Driven Approaches
Data Processing	Manual, rule-based	Automated, data-driven
Pattern Detection	Limited statistical models	Deep learning & ML
Scalability	Low	High
Multi-omics Integration	Difficult	Efficient
Drug Discovery Speed	Slow	Accelerated
Predictive Accuracy	Moderate	High

CONCLUSIONS

Artificial intelligence incorporated into bioinformatics sets a paradigm shift in the manner in which large and intricate biological information is handled, inspected, and interpreted. Predictive modeling and multi-omics integration powered by high-throughput data mining have allowed the deployment of AI to transform traditional biological studies into a data-driven scientific approach that can solve previously untenable problems. The applications include genome annotation, protein structure prediction, personalized medicine, and drug discovery, forming just a few features of the central role of AI in fast-tracking scientific discovery and clinical translation.

The potential of AI may not be realized without addressing the issues of information bias, computer program intelligibility, moral standards, and control. Genomic and medical data require care regarding privacy protection and the mitigation of bias. Similarly, transparent interpretability, support for reproducible solutions, and equitable access to solutions across populations are also critical to the successful development of trust in bioinformatics applications powered by AI systems.

There is enormous promise in more disruptive potential, like federated learning, explainable AI, and quantum-enhanced computing, to accelerate and enhance the potential of biological data analysis. To make this vision a reality, however, multidisciplinary cooperation is needed, combining know-how from computational science, biology, medicine, and ethics. Alongside this is a need to create resilient governance and global networks that can coordinate technological advances in accordance with the needs of societies. Despite rapid progress, several research gaps remain. The need for interpretable AI models, standardized benchmarking datasets, cross-population validation, and integration of multi-modal biological data continues to challenge large-scale clinical translation. Future research should prioritize trustworthy AI frameworks, equitable data representation, and regulatory-ready validation pipelines.

AI in bioinformatics is not the simple processing of big data, in other words, but rather an enabler, rewriting the threshold of biological discovery and healthcare innovation. By taking responsibility for exploiting its opportunities and dealing proactively with its constraints, the scientific community can use AI to unravel the complexity of life, propel sustainable medical breakthroughs, and eventually support the health of people on a global scale.

Conflicts of interest and sources of funding

The authors declare no conflict of interest.

This research received no external funding.

Acknowledgments

The authors used AI-assisted tools for language editing and grammatical refinement. The authors independently developed all scientific content, conceptualization, analysis, and interpretations.

Authors' contribution

Conceptualization, R.S. and R.S.; methodology, R.S.; validation, E.R.; resources, R.S.; data curation, R.S.; writing—original draft preparation, R.S.; writing—review and editing, H.M.R.; visualization, T.L.C.; supervision, R.S. All authors have read and agreed to the published version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

References

1. Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, Zhang J. et al. Artificial intelligence: A powerful paradigm for scientific research. The Innovation. 2021; 2(4): 100179. doi:10.1016/j.xinn.2021.100179

2. Lilhore UK, Kumar A, Vyas N, Simaiya S, Dutt V. (Eds.). (2025). Multimodal Data Fusion for Bioinformatics Artificial Intelligence. John Wiley & Sons. 2025; ISBN: 978-1-394-26993-8

3. Chandrashekar K, Niranjan V, Vishal A, Setlur AS. Integration of artificial intelligence, machine learning and deep learning techniques in genomics: review on computational perspectives for NGS analysis of DNA and RNA seq data. Current Bioinformatics.2024; 19(9): 825-844. doi: https://doi.org/10.2174/0115748936284044240108074937

4. Nguyen D, Le T, Le B. Explainability graph neural networks with nearest neighbor estimate interpretations. Neural Computing and Applications. 2025; 37(20): 15891-15922.doi:10.1007/s00521-025-11339-7

5. Cooper DR, Grabowski M, Zimmerman MD, Porebski PJ, Shabalin IG, Woinska M, Minor W. State-of-the-Art data management: improving the reproducibility, consistency, and traceability of structural biology and in vitro biochemical experiments. Structural Genomics: General Applications. 2021; 2199: 209-236. doi: 10.1007/978-1-0716-0892-0_13

6. Bayer JM, Thompson PM, Ching CR, Liu M, Chen A, Panzenhagen AC, Sämann PG. Site effects how-to and when: An overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. Frontiers in Neurology. 2022; 13: 923988. doi: https://doi.org/10.3389/fneur.2022.923988

7. Marques L, Costa B, Pereira M, Silva A, Santos J, Saldanha L, Vale N. Advancing precision medicine: A review of innovative in silico approaches for drug development, clinical pharmacology and personalized healthcare. Pharmaceutics. 2024; 16(3): 332. doi:10.3390/pharmaceutics16030332

8. Yu Y, Mai Y, Zheng Y, Shi L. Assessing and mitigating batch effects in large-scale omics studies. Genome Biology. 2024; 25(1): 254. doi:10.1186/s13059-024-03401-9

9. Hayes CN, Nakahara H, Ono A, Tsuge M, Oka S. (2024). From omics to multi-omics: a review of advantages and tradeoffs. Genes. 2024; 15(12): 1551. doi:10.3390/genes15121551

10. van Kampen AH, Mahamune U, Jongejan A, van Schaik BD, Balashova D, Lashgari D, Moerland PD. ENCORE: a practical implementation to improve reproducibility and transparency of computational research. Nature Communications. 2024; 15(1): 8117. doi: 10.1038/s41467-024-52446-8

11. Mukherjee A, Abraham S, Singh A, Balaji S, Mukunthan KS. From data to cure: A comprehensive exploration of multi-omics data analysis for targeted therapies. Molecular biotechnology. 2025; 67(4): 1269-1289. doi: 10.1007/s12033-024-01133-6

12. Pimpalkar A, Gandhewar N, Shelke N, Patil S, Chhabria S. An efficient deep convolutional neural networks model for genomic sequence classification. Genomics at the Nexus of AI, Computer Vision, and Machine Learning. 2024; 345-375. doi:10.1002/9781394268832.ch16

13. Gupta U, Pranav A, Kohli A, Ghosh S, Singh D. The contribution of artificial intelligence to drug discovery: Current progress and prospects for the future. Microbial Data Intelligence and Computational Techniques for Sustainable Computing. 2024; 1-23. doi:10.1007/978-981-99-9621-6_1

14. Mezei T, Kolcsár M, Joó A, Gurzu S. Image analysis in histopathology and cytopathology: from early days to current perspectives. Journal of Imaging. 2024; 10(10): 252. doi:10.3390/jimaging10100252

15. Kumar A. AI-driven precision oncology: predictive biomarker discovery and personalized treatment optimization using genomic data. Int J Adv Res Publ Rev. 2024; 1(3): 21-38.

16. Zhang Z, Sun Y, Peng Q, Li T, Zhou P. Integrating dynamical systems modeling with spatiotemporal scRNA-Seq Data Analysis. Entropy. 2025; 27(5): 453. https://doi.org/10.3390/e27050453.

17. Son A, Kim W, Park J, Park Y, Lee W, Lee S, Kim H. Mass spectrometry advancements and applications for biomarker discovery, diagnostic innovations, and personalized medicine. International Journal of Molecular Sciences. 2024; 25(18): 9880. doi:10.3390/ijms25189880

18. Prashar DB, Sharma M, Sharma D, Mir Z, Sood H, Sharma M. Exploring the complexity of plant stress response: role of metabolomics and signalling pathways in metabolic diversification. Journal of Plant Biochemistry and Biotechnology. 2025; 1-14. doi:10.1007/s13562-025-00986-5

19. Savitha S, Keerthana R, Logeswaran K, Keerthika P, Sharmila V, Sangeetha M. Integration of Multi-Omics Data: Genomics, Proteomics,

Metabolomics. Harnessing AI and Machine Learning for Precision Wellness. IGI Global Scientific Publishing. 2025; 149-184. doi:10.4018/979-8-3693-9521-9.ch006.

20. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, Hussain A. Interpreting black-box models: a review on explainable artificial intelligence. Cognitive Computation. 2024; 16(1): 45-74. doi:10.1007/s12559-023-10179-8

21. Serrano DR, Luciano FC, Anaya BJ, Ongoren B, Kara A, Molina G, Lalatsa A. Artificial intelligence (AI) applications in drug discovery and drug delivery: Revolutionizing personalized medicine. Pharmaceutics. 2024; 16(10): 1328. doi:10.3390/pharmaceutics16101328

22. Chen L, Li Q, Nasif KFA, Xie Y, Deng B, Niu S, Xie CY. AI-driven deep learning techniques in protein structure prediction. International journal of molecular sciences. 2024; 25(15): 8426. doi:10.3390/ijms25158426

23. Zhang K, Yang X, Wang Y, Yu Y, Huang N, Li G, Yang S. Artificial intelligence in drug development. Nature medicine. 2025; 31(1): 45-59. doi: 10.1038/s41591-024-03434-4

24. Ahmad S, Shah SNA, Parveen R, Raza K. Machine Learning for Genomic Profiling and Drug Discovery in Personalised Lung Cancer Therapeutics. Journal of Drug Targeting. 2025; 1-20. doi: 10.1080/1061186X.2025.2530656

25. Karim MM, Khan S, Van DH, Liu X, Wang C, Qu Q. Transforming Data Annotation with AI Agents: A Review of Architectures, Reasoning, Applications, and Impact. Future Internet. 2025; 17(8): 353. https://doi.org/10.3390/fi17080353

26. Hanna MG, Pantanowitz L, Dash R, Harrison JH, Deebajah M, Pantanowitz J, Rashidi HH. Future of artificial intelligence (AI)-machine learning (ML) trends in pathology and medicine. Modern Pathology. 2025; 36(4): 100705. https://doi.org/10.1016/j.modpat.2025.100705

27. Lauritzen AD, Lillholm M, Lynge E, Nielsen M, Karssemeijer N, Vejborg I. Early indicators of the impact of using AI in mammography screening for breast cancer. Radiology. 2024; 311(3): e232479. https://doi.org/10.1148/radiol.232479

28. Khalighi S, Reddy K, Midya A, Pandav KB, Madabhushi A, Abedalthagafi M. Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment. NPJ Precision Oncology. 2024; 8(1): 80. doi: 10.1038/s41698-024-00575-0

29. Sathyavathi S, Kavitha RK, Mary GPA, Baskaran KR. Machine Learning-Based Techniques for Predictive Diagnostics in Healthcare. In Cybersecurity and Data Science Innovations for Sustainable Development of HEICC. 2025; 47-61. doi:10.1201/9781032711300-4

30. Obuchowicz R, Strzelecki M, Piórkowski A. Clinical applications of Artificial Intelligence in Medical Imaging and Image Processing—A Review. Cancers. 2024; 16(10): 1870. https://doi.org/10.3390/cancers16101870

31. Zhou J, Park S, Dong S, Tang X, Wei X. Artificial intelligence-driven transformative applications in disease diagnosis technology. Medical Review. 2025; https://doi.org/10.1515/mr-2024-0097

32. Chaudhari JK, Pant S, Jha R, Pathak RK, Singh DB. Biological big-data sources, problems of storage, computational issues, and applications: a comprehensive review. Knowledge and Information Systems. 2024; 66(6): 3159-3209. doi:10.1007/s10115-023-02049-4

33. Rosenbacke R, Melhus Å, McKee M, Stuckler D. How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Health Care: Systematic Review. JMIR AI 3. 2024; e53207. doi:10.2196/53207

34. Ike JE, Kessie JD, Popoola R, Azeez MA, Onibokun T. A Novel Approach to Cloud Data Encryption using Homomorphic Encryption. 2024; 5(6): 19-27. doi: https://doi.org/10.54660/.IJFMR.2024.5.6.19-27

35. Caminero A, Tropini C, Valles-Colomer M, Shung DL, Gibbons SM, Surette MG, Verdu EF. Credible inferences in microbiome research: ensuring rigour, reproducibility and relevance in the era of AI. Nature Reviews Gastroenterology & Hepatology. 2025; doi: 10.1038/s41575-025-01100-9

36. Saraf S, De A, Tripathy BK. Effective use of computational biology and artificial intelligence in the domain of medical oncology. In Computational Intelligence for Oncology and Neurological Disorders. 2024; 228-252. doi:10.1201/9781003450153-17