

The **AI** REVOLUTION in **LIFE SCIENCES:**

Technologies, Applications,
and Future Directions



AUTHORS

Dr. Prabhavati
Dr. Soumya M. Hegde
Dr. Jagadevi Shivaputrappa
Dr. Mahesh Vilas Tanwade

The AI Revolution in Life Sciences: Technologies, Applications, and Future Directions



**India | UAE | Nigeria | Uzbekistan | Montenegro | Iraq |
Egypt | Thailand | Uganda | Philippines | Indonesia**
www.empyrealpublishinghouse.com

The AI Revolution in Life Sciences: Technologies, Applications, and Future Directions

Authored by:

Dr. Prabhavati

Assistant Professor

Department of Zoology, K.L.E. Society's S. Nijalingappa College,
Rajajinagar, Bengaluru, Karnataka, India

Dr. Soumya M. Hegde

Assistant Professor

Department of Botany, KLE Society's S. Nijalingappa College, Rajajinagar,
Bengaluru, Karnataka, India

Dr. Jagadevi Shivaputrappa

Guest Lecturer

Department of Botany, Gulbarga University, Kalaburagi, Karnataka, India

Dr. Mahesh Vilas Tanwade

Assistant Professor

Department of Postgraduate Studies and Research in Zoology, Sharnbasva
University, Kalaburagi, Karnataka, India

Copyright 2026 by Dr. Prabhavati, Dr. Soumya M. Hegde, Dr. Jagadevi Shivaputrappa and Dr. Mahesh Vilas Tanwade

First Impression: 2026

The AI Revolution in Life Sciences: Technologies, Applications, and Future Directions

ISBN: 978-93-49359-23-9

DOI:

Rs. 1000/- (\$80)

No part of the book may be printed, copied, stored, retrieved, duplicated and reproduced in any form without the written permission of the editor/publisher.

DISCLAIMER

Information contained in this book has been published by Empyreal Publishing House and has been obtained by the authors from sources believed to be reliable and correct to the best of their knowledge. The authors are solely responsible for the contents of the articles compiled in this book. Responsibility of authenticity of the work or the concepts/views presented by the author through this book shall lie with the author and the publisher has no role or claim or any responsibility in this regard. Errors, if any, are purely unintentional and readers are requested to communicate such error to the author to avoid discrepancies in future.

Published by:
Empyreal Publishing House

Preface

Artificial Intelligence (AI) has emerged as one of the most transformative technologies of the twenty-first century, and its influence on the life sciences is growing at an unprecedented pace. The convergence of AI with biology, medicine, biotechnology, genomics, pharmaceuticals, agriculture, and other life-science disciplines is creating new possibilities for research, discovery, diagnosis, prediction, and innovation. The AI Revolution in Life Sciences: Technologies, Applications, and Future Directions has been developed to provide a comprehensive overview of this rapidly evolving landscape.

The increasing availability of large and complex datasets, advances in machine learning and deep learning, improvements in computational infrastructure, and the development of generative AI have significantly changed how scientific and biological problems are approached. AI systems can identify patterns within vast datasets, support the interpretation of complex biological information, accelerate drug discovery, assist in disease prediction, enable personalised healthcare, and contribute to more efficient and sustainable biological and agricultural systems.

This book explores the fundamental technologies underlying AI and examines their applications across diverse areas of the life sciences. It highlights how machine learning, deep learning, natural language processing, computer vision, predictive analytics, and emerging AI approaches are being integrated into scientific and healthcare workflows. Particular attention is given to applications such as drug discovery and target identification, disease diagnosis and prediction, precision medicine, genomics, biotechnology, bioengineering, agriculture, and other emerging areas.

Beyond current applications, the book also considers the challenges associated with the growing use of AI in life sciences. Issues related to data quality, privacy, security, algorithmic bias, interpretability, ethical decision-making, regulatory frameworks, and responsible AI adoption are increasingly important as AI systems become part of critical scientific and healthcare environments. Understanding these challenges is essential for ensuring that technological progress remains scientifically reliable, ethically responsible, and socially beneficial.

The book also looks towards the future, considering emerging developments such as generative AI, AI-driven biological discovery, digital twins,

autonomous laboratories, advanced computational biology, and the potential convergence of AI with quantum computing. These developments may further accelerate scientific discovery and open new pathways for addressing some of the most complex challenges in human health, biotechnology, food security, and biological research.

This volume is intended for undergraduate and postgraduate students, researchers, academicians, healthcare professionals, biotechnology practitioners, and professionals seeking to understand the role of AI in modern life sciences. It may also serve as a reference for readers who wish to explore the opportunities and limitations of AI-enabled scientific innovation without requiring highly specialised knowledge of every underlying computational technique.

The objective of this book is not merely to present Artificial Intelligence as a technological advancement, but to examine it as an evolving scientific tool that is reshaping the way knowledge is generated, interpreted, and applied in the life sciences. By bringing together technological foundations, practical applications, emerging trends, and future directions, this book aims to encourage informed discussion and further research in this interdisciplinary field.

As AI continues to evolve, its greatest potential may lie not in replacing scientific expertise, but in augmenting human capabilities and enabling researchers and professionals to address problems that were previously difficult, time-consuming, or impossible to solve. The responsible integration of human knowledge with computational intelligence will therefore remain central to the future of life-science innovation.

We hope this book provides readers with a useful foundation for understanding the present state of AI in life sciences while inspiring further exploration, research, and innovation in the years ahead.

Acknowledgement

The authors would like to express their sincere gratitude to all those who contributed, directly or indirectly, to the development of this book, *The AI Revolution in Life Sciences: Technologies, Applications, and Future Directions*.

The rapid advancement of Artificial Intelligence has created remarkable opportunities across life sciences, transforming the way biological data are analysed, diseases are diagnosed, drugs are discovered, and healthcare and biotechnology are delivered. Bringing together these developments into a coherent and accessible volume has been possible through the support, encouragement, and intellectual contributions of numerous individuals and institutions.

We are deeply thankful to our colleagues, researchers, educators, healthcare professionals, scientists, and students whose discussions, insights, and continuing work in Artificial Intelligence and life sciences have enriched our understanding of this rapidly evolving field. Their dedication to research and innovation continues to inspire new approaches to solving complex challenges in biology, medicine, biotechnology, agriculture, and related disciplines.

We also acknowledge the valuable contributions of the wider scientific and academic community whose research, publications, datasets, technological developments, and scholarly discussions have provided the foundation for many of the concepts presented in this book.

Our sincere appreciation goes to our families, friends, colleagues, and well-wishers for their patience, encouragement, and constant support throughout the preparation of this work. Their understanding and motivation played an important role in bringing this book to completion.

Finally, we extend our gratitude to the publishing and editorial team for their cooperation and assistance in shaping this work into its present form.

We hope that this book will serve as a useful resource for students, researchers, academicians, healthcare professionals, biotechnology practitioners, and all those interested in understanding the transformative potential of Artificial Intelligence in life sciences.

About the Authors



Dr. Prabhavati is an Assistant Professor in the Department of Zoology at K.L.E. Society's S. Nijalingappa College, Rajajinagar, Bengaluru, Karnataka, India. She received her Ph.D. in Zoology from Gulbarga University, Kalaburagi in 2019, with a dissertation named "Hydrobiological Studies of Kamalavati River of Sedam Taluka, Kalaburagi District," which focused on limnology, ecotoxicology and aquatic biodiversity. Prior to her current position, she worked as an Assistant Professor for eight years of teaching both undergraduate and postgraduate courses, actively contributing to academic curriculum development and guiding numerous student research projects. Her main research and teaching interests include Ecology, Biodiversity Assessment, Wildlife Management, Ethology, Genetics and Molecular Biology, Anatomy, Physiology and Reproductive Biology. Dr. Prabhavati, has published six research papers, serves as a peer reviewer for academic journals, develops undergraduate instruction and laboratory manuals and has presented at national and international conferences. She has been honoured with the Women Icon Award (2025).



Dr. Soumya M. Hegde is an Assistant Professor in the Department of Botany at KLE Society's S. Nijalingappa College, Rajajinagar, Bengaluru, Karnataka, India. She completed her Ph.D. in Botany from Sahyadri Science College, Kuvempu University, Shivamogga in 2024 with the research title "Diversity and Biology of Terrestrial Orchids in Shivamogga District, Karnataka." Her doctoral work focused on terrestrial orchid diversity, distribution, phenology, reproductive Biology, mycorrhizal associations, and conservation in the Western Ghats of Karnataka.

Her areas of teaching and research interest include Plant Taxonomy, Orchid Biology, Plant Ecology, Biodiversity Assessment, Plant Conservation, Mycorrhizal Associations, Plant Anatomy, Phytochemistry, Ethnobotany, Medicinal Plants, and Plant Secondary Metabolites. She has conducted extensive field surveys across the Western Ghats, documented orchid diversity, carried out anatomical and ecological studies, and contributed to

biodiversity conservation research.

Dr. Hegde has published 18 research papers in the areas of orchid diversity, phenology, plant anatomy, and biodiversity of the Western Ghats. She is actively involved in undergraduate teaching, laboratory instruction, curriculum development, preparation of practical manuals, question banks, and academic mentoring. She has also served as a reviewer for graduate research synopses.

She has presented research papers at several national and international conferences and participated in numerous workshops, seminars, and training programmes related to biodiversity, orchid conservation, plant taxonomy, ecology, and research methodology.

Dr. Hegde is a life member of The Orchid Society of India and a member of the International Society for Horticultural Science (ISHS). She is committed to promoting scientific research, biodiversity conservation, and quality higher education through field-based learning, innovative teaching, and interdisciplinary research



Dr. Jagadevi Shivaputrappa is a Guest Lecturer in the department of Botany, Gulbarga University, Kalaburagi, Karnataka, India. Her areas of interest in teaching and research include Medicinal plants, Agricultural biotechnology, Plant breeding, molecular biology, microbiology and genetic engineering. She published 5 research articles in reputed national and international journals and a book chapter in Cambridge scholars publishing house, UK and presented her research work in several national and international conferences.

She is actively involved in curricular and co-curricular enrichment activities. She is an active member of MycoAsia and institutional botanical society. She aspires to make a significant contribution to research and education.



Dr. Mahesh Vilas Tanwade is an Assistant Professor in the Department of Postgraduate Studies and Research in Zoology at Sharnbasva University, Kalaburagi, Karnataka, India. With over four years of academic and research experience, he actively mentors postgraduate and doctoral scholars, currently supervising three research students. He holds a Ph.D. and an M.Sc. in Zoology from Sharnbasva University and obtained his undergraduate studies from Sharnbasveshwar College of Science, Kalaburagi affiliated to Gulbarga University, Kalaburagi.

Dr. Mahesh Vilas Tanwade's core research expertise lies at the intersection of endocrinology, reproductive biology, and ecotoxicology. Utilizing rodent animal models, his work explores male fertility, the neuroendocrine regulation of copulatory behaviors—including mounting, intromission, and ejaculation—and the adverse physiological impacts of endocrine-disrupting chemicals (EDCs). His research integrates standard toxicological and endocrine assays, such as the Hershberger, uterotrophic, and repeated-dose toxicity tests, alongside comprehensive semen analysis, to advance strategies in reproductive healthcare and environmental risk assessment.

Dr. Mahesh Vilas Tanwade is an active contributor to the scientific community and has published several research papers in peer-reviewed International Journals, including Comparative Clinical Pathology, Biological Trace Element Research, and the Journal of Molecular Structure. He has presented his scholarly findings at leading institutions, such as Dr. B. Lal Institute of Biotechnology, Jaipur, and Banaras Hindu University (BHU), Varanasi. Dr. Mahesh Vilas Tanwade holds life memberships in several professional organizations, including the Indian Society for the Study of Reproduction and Fertility (ISSRF), the Laboratory Animal Scientists Association (LASA), India, and the Society of Clinical Research and Medical Professionals (SCRMP), USA.

Table of Contents

Title of Chapters	Page No.
CHAPTER 1	1 – 19
<i>Foundations of Artificial Intelligence in Life Sciences</i>	
CHAPTER 2	20 – 38
<i>AI in Genomics and Bioinformatics</i>	
CHAPTER 3	39 – 55
<i>Artificial Intelligence in Drug Discovery and Development</i>	
CHAPTER 4	56 – 75
<i>AI Applications in Healthcare and Medical Diagnostics</i>	
CHAPTER 5	76 – 94
<i>Artificial Intelligence in Agriculture and Environmental Life Sciences</i>	
CHAPTER 6	95 - 113
<i>AI in Biotechnology and Industrial Life Sciences</i>	
CHAPTER 7	114 – 131
<i>Ethical, Legal, and Social Implications of AI in Life Sciences</i>	
CHAPTER 8	132 – 149
<i>Future Trends and Emerging Innovations in AI and Life Sciences</i>	
REFERENCES	150 – 165

CHAPTER 1

Foundations of Artificial Intelligence in Life Sciences

1.1 INTRODUCTION TO ARTIFICIAL INTELLIGENCE IN LIFE SCIENCES

Artificial Intelligence (AI) has emerged as one of the most influential technological developments of the twenty-first century, transforming the way data are generated, analyzed, interpreted, and applied across scientific and healthcare domains. In the life sciences, AI refers broadly to computational methods that enable machines to learn from biological, medical, chemical, and clinical data and use that knowledge to perform tasks such as prediction, classification, pattern recognition, decision support, and knowledge discovery. The growing availability of high-dimensional biological datasets, improvements in computational power, and advances in machine learning and deep learning have created new opportunities for applying AI to problems that were previously difficult or time-consuming to solve (Topol, 2019). (Figure 1.1)

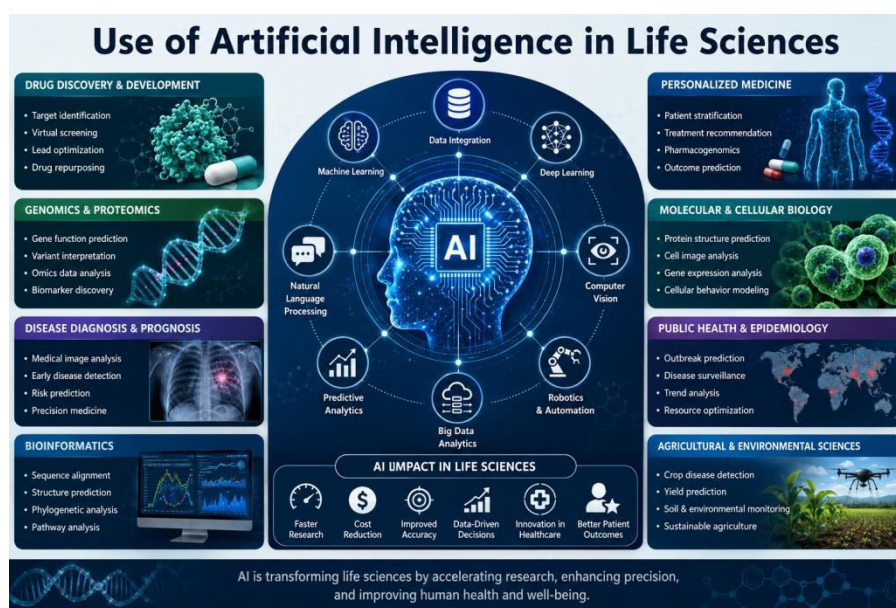


Figure 1.1: Use of Artificial Intelligence in Life Sciences

Life sciences encompass a wide range of disciplines, including biology, biotechnology, medicine, genomics, bioinformatics, pharmacology, neuroscience, epidemiology, agriculture, and environmental biology. These fields increasingly depend on large and complex datasets. Examples include

DNA and RNA sequences, protein structures, medical images, electronic health records, laboratory results, clinical-trial data, molecular structures, and information obtained from wearable and monitoring devices. Traditional analytical approaches can struggle to identify meaningful relationships within such large datasets. AI provides computational techniques capable of processing large volumes of information and identifying complex patterns that may not be immediately apparent to researchers or healthcare professionals.

Machine learning (ML), a major branch of AI, enables computational systems to learn patterns from data rather than relying exclusively on explicitly programmed rules. Supervised learning can be used for classification and prediction when labelled datasets are available, while unsupervised learning can identify structures or clusters within datasets without predefined labels. Deep learning, which uses multilayer neural networks, has been particularly important for analyzing complex data such as images, genomic sequences, and other high-dimensional biological information. The success of deep learning has been supported by advances in computing infrastructure, large datasets, and improved algorithms (Topol, 2019).

One of the most visible applications of AI in life sciences is medical diagnosis. AI-based systems can analyze medical images such as X-rays, computed tomography (CT) scans, magnetic resonance imaging (MRI), pathology slides, and dermatological photographs. For example, Esteva et al. (2017) demonstrated that a deep convolutional neural network trained using a large collection of clinical skin images could classify certain skin cancers at a level comparable to dermatologists in the experimental setting. Such research demonstrated the potential of deep learning for assisting clinicians with image-based diagnosis and screening.

AI is also contributing to genomics and precision medicine. Modern sequencing technologies generate enormous quantities of genetic information, creating a need for computational approaches capable of interpreting complex biological relationships. AI can assist researchers in identifying genetic variants, predicting disease susceptibility, analyzing gene expression, and understanding relationships between genetic characteristics and clinical outcomes. When combined with clinical and environmental information, these

approaches can support more individualized approaches to disease prevention and treatment.

However, the effectiveness of such systems depends heavily on the quality, representativeness, and appropriate interpretation of the underlying data.

Another major area of AI application is drug discovery and development. Conventional drug discovery can involve extensive laboratory experimentation, molecular screening, and repeated optimization. AI can support this process by predicting molecular properties, identifying potential drug targets, screening candidate compounds, and assisting in the design of new molecules. Machine-learning approaches have increasingly been investigated for virtual screening, molecular property prediction, and molecular generation (Deng et al., 2021). AI can therefore help researchers prioritize promising candidates before committing substantial laboratory resources to experimental testing.

The application of AI to protein science represents another significant development. Proteins perform essential biological functions, and their three-dimensional structures are closely related to their functions and interactions. Deep-learning systems such as AlphaFold demonstrated that AI could make highly accurate predictions of protein structures from amino-acid sequences, addressing a longstanding computational challenge in structural biology (Jumper et al., 2021). Such developments have important implications for structural biology, disease research, protein engineering, and drug discovery because improved structural information can assist scientists in understanding biological mechanisms and identifying potential therapeutic targets.

AI is additionally being applied to clinical research and healthcare management. Algorithms can assist with patient risk prediction, disease monitoring, clinical decision support, patient stratification, medical documentation, and analysis of clinical-trial data. AI may also support public-health surveillance by identifying patterns in population-level health information. According to the World Health Organization (WHO), AI has considerable potential to improve diagnosis, treatment, health research, drug development, and public-health functions, including surveillance and outbreak response (WHO, 2021).

Despite these opportunities, AI in life sciences must not be viewed simply as a replacement for human expertise.

Biological systems are highly complex, context-dependent, and variable. A model that performs well on one dataset may produce less reliable results when applied to a different population, laboratory environment, healthcare system, or biological condition. Issues such as data quality, dataset bias, privacy, interpretability, reproducibility, cybersecurity, and regulatory compliance can significantly affect the reliability of AI systems. Topol (2019) emphasizes that bias, privacy, security, and lack of transparency remain important limitations in medical AI.

Ethical governance is therefore an essential component of AI adoption in life sciences. Biological and medical data can contain highly sensitive information, including genetic and health-related information. The WHO recommends that AI for health be developed and deployed with attention to human autonomy, safety, transparency, accountability, inclusiveness, and equity (WHO, 2021). These principles are particularly important because AI-generated recommendations can influence clinical decisions, research priorities, and access to healthcare.

Overall, AI represents a powerful methodological framework for modern life sciences. Its importance lies not only in automating existing tasks but also in enabling researchers to explore biological questions at a scale and complexity that traditional methods may struggle to address. From medical imaging and genomics to protein structure prediction and drug discovery, AI is increasingly becoming an integral component of interdisciplinary scientific research. Nevertheless, meaningful progress requires collaboration among computer scientists, biologists, clinicians, pharmacists, statisticians, ethicists, and policymakers. The future of AI in life sciences will therefore depend on combining computational innovation with biological understanding, rigorous scientific validation, responsible data governance, and human expertise.

1.2 MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Machine learning (ML) and deep learning (DL) represent two of the most important computational approaches within artificial intelligence (AI), enabling systems to learn patterns from data and generate predictions,

classifications, or decisions without requiring every rule to be explicitly programmed. Their importance in life sciences has increased substantially because modern biological and medical research generates large, heterogeneous, and high-dimensional datasets, including genomic sequences, transcriptomic profiles, proteomic measurements, electronic health records, medical images, physiological signals, and clinical text. Machine learning provides methods for extracting meaningful relationships from these datasets, while deep learning extends this capability by automatically learning increasingly complex representations of raw or minimally processed data (LeCun et al., 2015; Zitnik et al., 2018). (Figure 1.2)

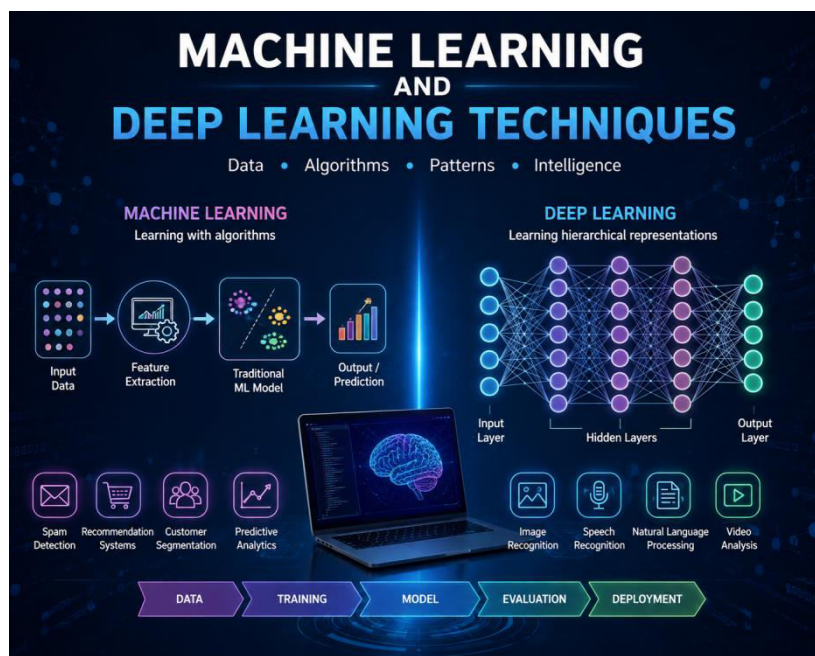


Figure 1.2: Machine Learning and Deep Learning Techniques

1.2.1 Machine Learning in Life Sciences

Machine learning refers to computational methods that learn relationships between input variables and desired outcomes from examples. Unlike conventional rule-based systems, ML models can improve their performance as they are exposed to additional data. In life sciences, machine learning is commonly divided into supervised, unsupervised, and reinforcement learning.

Supervised learning uses labelled datasets in which the desired output is already known. Classification and regression are the two major forms of supervised learning. Classification algorithms can be used to distinguish between disease and healthy states, classify cancer subtypes, identify pathogenic microorganisms, or predict whether a patient is at high or low risk of developing a particular condition. Regression models, in contrast, predict continuous outcomes such as drug response, disease progression, biomarker levels, or physiological measurements. Common algorithms include logistic regression, linear regression, decision trees, random forests, support vector machines (SVMs), and gradient-boosting methods.

Unsupervised learning operates on datasets without predefined labels. Its objective is generally to discover hidden structures or relationships within the data. Clustering techniques such as k-means, hierarchical clustering, and density-based methods can group patients according to molecular or clinical characteristics. In genomics and transcriptomics, unsupervised approaches can identify gene-expression patterns and potentially reveal previously unrecognized disease subtypes. Dimensionality-reduction techniques, including principal component analysis (PCA), can also simplify complex biological datasets while retaining important patterns.

Reinforcement learning (RL) represents another learning paradigm in which an algorithm learns through interaction with an environment and receives rewards or penalties for its actions. Although less widespread than supervised and unsupervised learning in routine biomedical analysis, reinforcement learning has potential applications in treatment optimization, clinical decision support, adaptive experimentation, and drug-development strategies.

An important strength of machine learning in life sciences is its ability to integrate heterogeneous biological information. Biological systems cannot generally be understood through a single data type; genomic, epigenomic, transcriptomic, proteomic, phenotypic, and clinical information may provide complementary perspectives. Machine learning can therefore be used to combine multiple data sources and identify relationships that may not be apparent through traditional statistical analysis alone (Zitnik et al., 2018).

1.2.2 Deep Learning Techniques

Deep learning is a specialized branch of machine learning based on artificial neural networks containing multiple processing layers. These layers enable models to learn hierarchical representations, moving from relatively simple features toward increasingly complex patterns. Unlike many traditional ML approaches that depend heavily on manually engineered features, deep learning can learn useful representations directly from large datasets (LeCun et al., 2015; Goodfellow et al., 2016).

Convolutional neural networks (CNNs) are particularly effective for image-based life-science applications. CNNs automatically learn spatial features from images and have been widely investigated for medical image analysis, including the interpretation of radiological images, histopathological slides, retinal photographs, and other biomedical images. Deep learning-enabled computer vision has consequently become an important area of biomedical AI research (Esteva et al., 2021).

Recurrent neural networks (RNNs) and related architectures such as long short-term memory (LSTM) networks are designed to process sequential information. They can be applied to biological sequences, physiological time-series data, clinical records, and biomedical language. More recently, transformer-based architectures have become increasingly important because their attention mechanisms can model relationships across long sequences and can be adapted to genomic, proteomic, molecular, and clinical datasets.

Autoencoders are another important deep-learning technique. They learn compact representations of input data and can be used for dimensionality reduction, noise removal, anomaly detection, and feature extraction. In biomedical research, these capabilities can be useful when dealing with high-dimensional molecular datasets in which the number of measured variables greatly exceeds the number of available observations.

Deep learning has also produced major advances in structural biology. A prominent example is AlphaFold, a deep-learning system developed for protein-structure prediction. Its ability to predict highly accurate protein structures demonstrated how neural networks could address fundamental

biological problems that had traditionally required extensive experimental and computational effort (Jumper et al., 2021).

1.2.3 Applications across Life Sciences

The combination of ML and DL has created applications across multiple areas of life sciences. In drug discovery, models can predict molecular properties, identify potential drug–target interactions, estimate toxicity, and assist in the prioritization of candidate compounds. In genomics, ML techniques can identify disease-associated variants, classify genomic patterns, and support prediction of molecular phenotypes. In precision medicine, models can combine clinical, molecular, imaging, and lifestyle information to estimate individual disease risks and potential treatment responses.

In healthcare, AI systems have demonstrated applications in medical imaging, clinical decision support, patient monitoring, and risk prediction. The availability of electronic health records, medical imaging, wearable sensors, genomic sequencing, and other data sources has also encouraged the development of multimodal AI, in which different types of biomedical information are processed together (Rajpurkar et al., 2022; Acosta et al., 2022).

1.2.4 Challenges and Considerations

Despite their potential, ML and DL techniques face important challenges in life-science applications. Biomedical datasets may contain missing values, measurement errors, class imbalance, limited sample sizes, and inconsistent labels. Models trained on one population may perform poorly when applied to another because of differences in demographics, clinical practices, equipment, or data collection procedures. Consequently, external validation and careful evaluation are essential.

Another major concern is interpretability. Highly complex deep-learning models may produce accurate predictions without providing sufficiently understandable explanations for researchers or clinicians. Bias, privacy, data security, and transparency are additional concerns, particularly when models use sensitive patient information (Topol, 2019; Rajpurkar et al., 2022).

Overall, machine learning and deep learning provide a powerful computational foundation for modern life-science research. Traditional ML remains valuable

for structured datasets and interpretable predictive modelling, whereas DL is particularly advantageous for large-scale, complex, and unstructured data such as images, sequences, and multimodal biomedical information. Their future impact will depend not only on increasingly sophisticated algorithms but also on high-quality datasets, rigorous validation, interdisciplinary collaboration, ethical governance, and meaningful integration with biological and clinical expertise.

1.3 ROLE OF BIG DATA AND COMPUTATIONAL BIOLOGY

The rapid development of high-throughput technologies has fundamentally changed the way biological and medical research is conducted. Modern life sciences generate enormous quantities of heterogeneous data from genome sequencing, transcriptomics, proteomics, metabolomics, medical imaging, electronic health records, clinical trials, and environmental observations. This rapid expansion has created what is commonly described as the big data era in biology. Computational biology provides the analytical foundation for managing and interpreting these datasets, while artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), enables researchers to identify complex patterns that may not be readily detectable through conventional statistical approaches (Libbrecht & Noble, 2015; Min et al., 2017). (Figure 1.3)

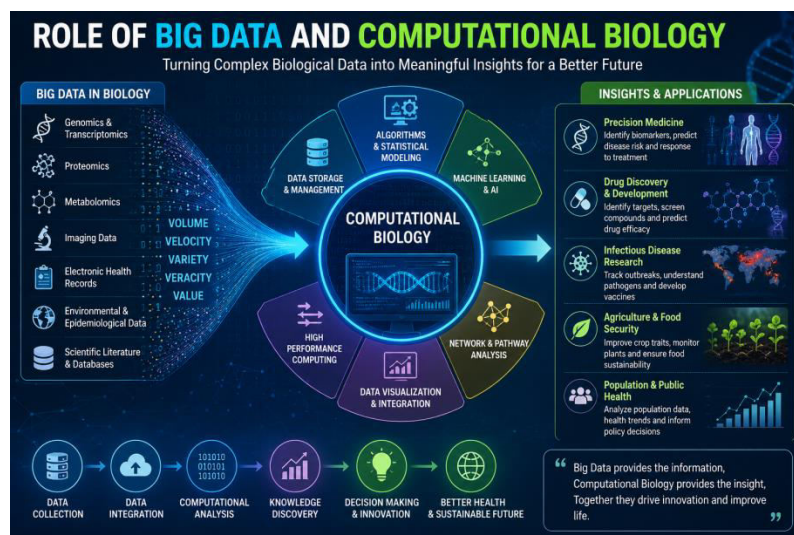


Figure 1.3: Role of Big Data and Computational Biology

1.3.1 Big Data in Life Sciences

Big data in life sciences is characterized not simply by its volume but also by its variety, velocity, complexity, and heterogeneity. For example, genomic studies can produce millions or billions of DNA sequence measurements, while transcriptomic and proteomic technologies generate information about gene expression and protein abundance across thousands of biological samples. Single-cell technologies add another dimension by measuring molecular characteristics at the level of individual cells. Consequently, biological datasets frequently contain millions of variables but comparatively fewer observations, creating substantial computational and statistical challenges (Hasin et al., 2017).

The value of biological big data depends heavily on the ability to organize, integrate, and interpret it. Computational infrastructure such as high-performance computing, cloud platforms, distributed storage, and specialized databases has therefore become an essential component of modern life-science research. In parallel, principles such as FAIR-Findable, Accessible, Interoperable, and Reusable-data management are increasingly important for ensuring that biological datasets can be effectively shared and reused across research groups and institutions (Mahmud & Banerjee, 2026).

1.3.2 Computational Biology as an Analytical Framework

Computational biology applies mathematical models, algorithms, statistics, and computer science to biological problems. It serves as an important bridge between experimentally generated data and biological interpretation. Traditional computational biology has been used for sequence alignment, genome assembly, phylogenetic analysis, molecular modelling, protein structure analysis, and biological network reconstruction. With the emergence of large-scale datasets, however, computational biology has increasingly incorporated machine learning and AI-based approaches.

Machine learning is particularly valuable because biological datasets often contain complex and nonlinear relationships. Supervised learning can be used when datasets contain labelled examples, such as disease versus healthy samples, whereas unsupervised learning can identify previously unknown patterns or groups within unlabelled datasets. Semi-supervised approaches can combine both types of information (Libbrecht & Noble, 2015).

Deep learning has further expanded these capabilities. Neural networks can automatically learn representations from high-dimensional biological data, reducing the need for manually engineered features. Applications include DNA and RNA sequence analysis, protein structure and function prediction, biomedical image analysis, disease classification, and drug-response prediction. Reviews of deep learning in bioinformatics demonstrate its growing application across omics, biomedical imaging, and biomedical signal-processing domains (Min et al., 2017).

1.3.3 Big Data and Multi-Omics Integration

One of the most significant contributions of computational biology is the integration of different biological data layers. Genomics describes DNA variation, transcriptomics examines RNA expression, proteomics investigates proteins, and metabolomics characterizes small molecules and metabolic processes. Studying these datasets independently can provide valuable information, but integrating them can provide a more comprehensive representation of biological systems.

Multi-omics analysis attempts to connect these different molecular layers to understand how genetic variation influences gene expression, protein activity, metabolism, and ultimately disease phenotypes. Computational methods and AI models can identify relationships across these datasets and help researchers discover biomarkers, disease mechanisms, and potential therapeutic targets. Hasin et al. (2017) emphasized that integrating multiple omics layers can provide insights into the flow of biological information underlying disease that may not be apparent from a single data type.

1.3.4 AI-Driven Analysis of Biological Big Data

AI provides computational biology with increasingly powerful methods for transforming raw data into biological knowledge. Machine learning algorithms can identify disease-associated genomic variants, predict gene expression patterns, classify cells, and estimate patient-specific disease risks. More recent AI systems also support protein structure prediction, molecular design, biological sequence modelling, and multi-omics prediction (Mahmud & Banerjee, 2026).

The importance of big data is particularly evident in precision medicine. Patient-level information-including genomic profiles, clinical histories, medical images, laboratory measurements, and treatment responses-can be integrated to develop more individualized predictions and treatment strategies. AI can process these heterogeneous datasets at a scale that would be difficult to manage manually. Topol (2019) noted that the combination of labelled big data, improved computational power, and cloud storage has been an important foundation for the growing use of AI in medicine.

1.3.4 Challenges and Future Directions

Despite its potential, the combination of big data, computational biology, and AI presents several challenges. Biological datasets may contain missing values, technical noise, inconsistent measurements, population biases, and batch effects. Large datasets do not automatically guarantee reliable conclusions; data quality and appropriate experimental design remain essential. AI models can also suffer from overfitting, limited generalizability, and lack of interpretability. In clinical applications, privacy, security, ethical governance, and transparency are particularly important considerations (Topol, 2019).

Another challenge is the integration of increasingly diverse datasets. Future computational biology systems will need to combine genomic, transcriptomic, proteomic, imaging, clinical, and environmental information. Emerging multimodal and foundation-model approaches are being developed to address this challenge, with the goal of creating models capable of learning general biological representations across multiple data types (Mahmud & Banerjee, 2026).

Overall, big data provides the raw information, computational biology provides the analytical framework, and AI provides increasingly sophisticated methods for pattern recognition, prediction, and biological discovery. Their convergence is transforming life-science research from predominantly hypothesis-driven analysis toward a complementary model in which large-scale data can generate new hypotheses, identify previously hidden biological relationships, and support more precise biomedical decision-making. The future of AI in life sciences will therefore depend not only on increasingly powerful algorithms but also on high-quality datasets, robust computational

infrastructure, interdisciplinary expertise, transparent validation, and responsible data governance.

1.4 ETHICAL AND REGULATORY CHALLENGES IN AI

Artificial Intelligence (AI) is increasingly transforming the life sciences by supporting medical diagnosis, drug discovery, genomics, biomedical research, clinical decision-making, personalized medicine, and public-health surveillance. The ability of AI systems to analyze large and complex datasets can improve efficiency and generate insights that may be difficult to obtain through conventional approaches. However, the increasing dependence on AI also introduces significant ethical, legal, and regulatory challenges. Unlike conventional software, AI systems may learn from data, generate probabilistic outputs, and change their performance when exposed to new populations or datasets. Consequently, responsible deployment requires attention not only to technical accuracy but also to privacy, fairness, transparency, accountability, safety, and human oversight (World Health Organization [WHO], 2021). (Figure 1.4)

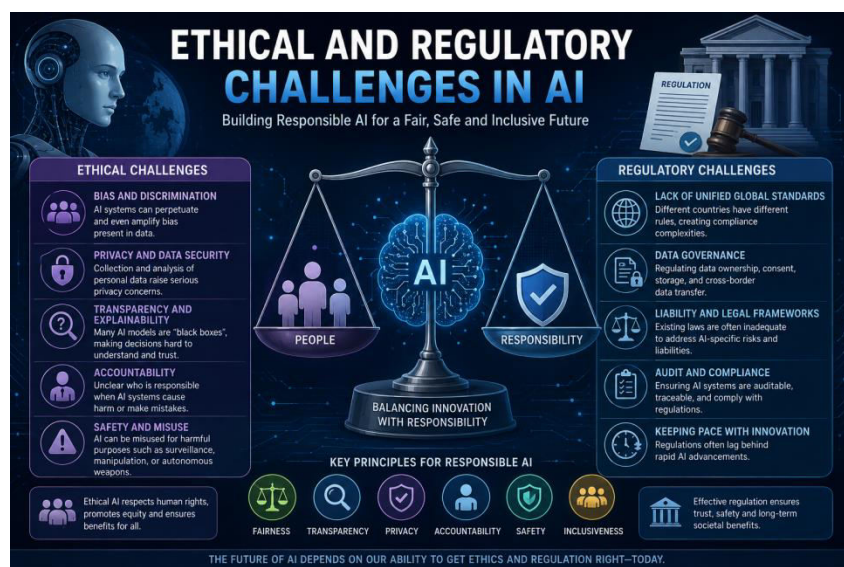


Figure 1.4: Ethical and Regulatory Challenges in AI

1.4.1 Data Privacy and Protection

One of the most important ethical concerns in life-science AI is the protection of sensitive data. AI models frequently require large datasets containing

electronic health records, medical images, genomic information, laboratory results, clinical-trial data, and other personal information. Such datasets may contain information that can identify individuals or reveal highly sensitive characteristics. Improper collection, storage, sharing, or secondary use of these data can therefore create substantial privacy risks. Ethical AI requires appropriate data governance, secure storage, controlled access, anonymization or pseudonymization where appropriate, and clearly defined purposes for data use. UNESCO emphasizes that privacy and data protection should be safeguarded throughout the entire AI lifecycle rather than being treated as a one-time requirement (UNESCO, 2021).

A related issue is informed consent. Traditional consent procedures may not adequately explain future AI applications when biological or clinical data are collected for research. Data originally collected for one purpose may later be reused for algorithm development, validation, or commercial applications. Researchers therefore need governance mechanisms that clarify permitted secondary uses and protect participants' rights.

AI-based clinical trials can make consent particularly complex because participants may not fully understand how their data will contribute to model development or future algorithmic decision-making (Schoenherr et al., 2024).

1.4.2 Algorithmic Bias and Fairness

AI systems learn patterns from historical data; consequently, biased or incomplete datasets can produce biased predictions. In life sciences, this problem can have serious consequences because an algorithm that performs well for one population may perform poorly for another. Differences in age, sex, ethnicity, socioeconomic background, geographic location, disease prevalence, or healthcare access can affect the representativeness of training data. Research has shown that AI systems may maintain high overall accuracy while producing substantially different performance across population groups (Meskó & Görög, 2020).

Bias can influence disease diagnosis, risk prediction, treatment recommendations, clinical-trial recruitment, and allocation of healthcare resources. Therefore, fairness should be considered throughout the AI lifecycle. Developers should use diverse datasets, evaluate model performance

across relevant demographic groups, conduct external validation, and continuously monitor systems after deployment. WHO identifies inclusiveness, equity, and protection against discrimination as fundamental principles for AI in health.

1.4.3 Transparency and Explainability

Many advanced AI systems, particularly deep-learning models, operate as complex "black boxes." Their predictions may be highly accurate while the reasoning behind individual outputs remains difficult for clinicians and patients to understand. This creates ethical difficulties when AI recommendations influence diagnosis or treatment. Healthcare professionals may be reluctant to rely on an algorithm when they cannot understand why it produced a particular result, especially when the recommendation conflicts with clinical judgment. Lack of transparency can also make it difficult to identify errors, bias, or inappropriate use (Meskó & Görög, 2020).

Explainability does not necessarily require revealing every technical detail of an algorithm.

Instead, users should receive meaningful information about the system's purpose, limitations, training context, performance, and appropriate use. UNESCO therefore recognizes transparency and explainability as important principles of responsible AI while also acknowledging that they must be balanced against privacy, safety, and security requirements (UNESCO, 2021).

1.4.4 Accountability and Human Oversight

Another major challenge concerns responsibility when an AI-supported decision causes harm. If an AI system incorrectly identifies a disease or recommends an inappropriate treatment, responsibility may potentially involve the developer, healthcare institution, clinician, data provider, or other stakeholders. The complexity of AI development can therefore make conventional concepts of accountability difficult to apply.

Human oversight remains particularly important in high-stakes life-science applications. AI should generally support rather than automatically replace professional judgment where decisions may significantly affect human health. WHO emphasizes that humans should remain responsible for decisions involving healthcare and that AI governance should establish clear

accountability among developers, healthcare professionals, institutions, and policymakers. UNESCO similarly recommends that AI systems should not displace ultimate human responsibility and accountability.

1.4.5 Safety, Reliability, and Model Drift

AI systems can behave differently when deployed in real-world environments compared with controlled research settings. Changes in patient populations, clinical practices, equipment, disease patterns, or data quality can reduce model performance. This phenomenon is often associated with dataset shift or model drift. Consequently, validation before deployment is insufficient by itself; continuous post-deployment monitoring is necessary.

The U.S. Food and Drug Administration (FDA) has recognized this challenge in its regulatory approach to AI/ML-based Software as a Medical Device. Its AI/ML Action Plan emphasizes lifecycle-based oversight, good machine-learning practices, transparency, and real-world performance monitoring (FDA, 2021).

More recent regulatory thinking also addresses predetermined change-control plans so that certain planned AI modifications can be managed while maintaining safety and effectiveness.

1.4.6 Regulatory Complexity

The regulatory environment for AI in life sciences is evolving rapidly. AI applications may fall under different legal frameworks depending on their intended purpose, risk level, jurisdiction, and whether they function as medical devices, research tools, clinical decision-support systems, or components of pharmaceutical development.

In the European Union, the EU Artificial Intelligence Act establishes a risk-based regulatory framework. Certain AI systems associated with medical devices and in-vitro diagnostic medical devices can fall within the high-risk category when specified regulatory conditions are met. High-risk systems are subject to requirements concerning risk management, data governance, documentation, transparency, human oversight, and other safeguards (European Union, 2024).

In the United States, the FDA has developed a framework for AI/ML-enabled medical devices and has recognized that conventional medical-device

regulation can be challenging for adaptive AI systems. The FDA's approach increasingly considers the entire product lifecycle, including development, validation, modification, and post-market monitoring.

1.4.7 Intellectual Property and Data Ownership

AI-driven life-science research also raises questions concerning ownership of datasets, algorithms, AI-generated outputs, biological discoveries, and commercially valuable models. Pharmaceutical and biotechnology companies may use proprietary clinical, genomic, or molecular datasets to train AI systems, while academic researchers may rely on publicly available or shared databases. Disputes can arise over who owns the resulting model, discoveries, or inventions and whether the original data contributors have appropriate rights or recognition.

These issues become especially important in collaborative research involving universities, hospitals, technology companies, pharmaceutical firms, and public institutions.

Clear data-sharing agreements, intellectual-property policies, attribution requirements, and contractual responsibilities are therefore essential.

1.4.8 Toward Responsible AI in Life Sciences

Ethical and regulatory challenges should not be viewed simply as barriers to technological innovation. Instead, they provide a framework for ensuring that AI delivers scientific and societal benefits without creating unacceptable risks. Responsible AI requires multidisciplinary collaboration among AI developers, clinicians, researchers, ethicists, patients, regulators, lawyers, and policymakers. Risk assessment should begin during system design and continue throughout deployment and subsequent updates.

A responsible governance framework should incorporate privacy protection, fairness, transparency, explainability, human oversight, accountability, cybersecurity, scientific validation, continuous monitoring, and meaningful stakeholder participation. WHO's guidance emphasizes that AI in health should place ethics and human rights at the center of design, development, and implementation (WHO, 2021). Similarly, UNESCO's global recommendation promotes human rights, fairness, privacy, accountability, transparency, safety,

and sustainability as core elements of ethical AI governance (UNESCO, 2021).

The successful integration of AI into life sciences depends not only on developing more powerful algorithms but also on creating trustworthy systems that respect human autonomy, protect sensitive information, reduce discrimination, and remain scientifically and clinically reliable. As AI becomes increasingly embedded in healthcare and biomedical research, ethical principles and regulatory requirements must evolve alongside technological capabilities. A balanced approach—one that encourages innovation while maintaining safety, equity, accountability, and human oversight—will be essential for realizing the long-term potential of AI in life sciences.

CHAPTER 2

AI in Genomics and Bioinformatics

2.1 AI APPLICATIONS IN GENOMIC DATA ANALYSIS

The rapid development of next-generation sequencing (NGS) and other high-throughput genomic technologies has generated enormous volumes of biological data. Whole-genome sequencing, whole-exome sequencing, RNA sequencing, single-cell sequencing, and other genomic technologies can produce millions to billions of sequence observations from a single experiment. Conventional bioinformatics pipelines remain essential for processing these data, but the increasing complexity, scale, and dimensionality of genomic datasets have created a strong need for more advanced computational approaches. Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), has therefore become an important component of modern genomic data analysis. AI techniques can identify complex patterns, relationships, and biological signals that may be difficult to detect using conventional statistical methods (Eraslan et al., 2019; Olawade et al., 2025). (Figure 2.1)

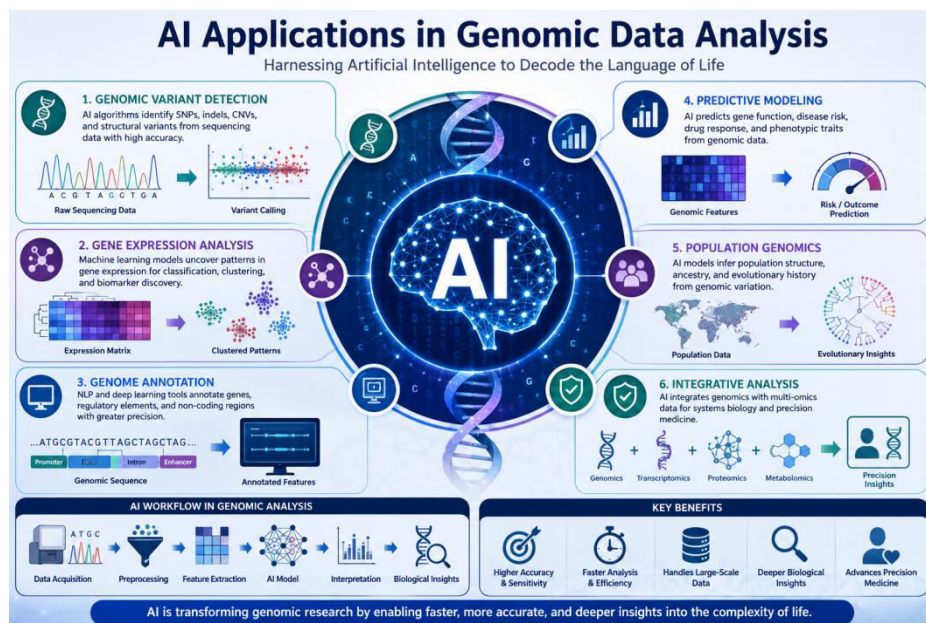


Figure 2.1: AI Applications in Genomic Data Analysis

2.1.1 AI for Sequence and Variant Analysis

One of the most established applications of AI in genomics is the identification of genetic variants from sequencing data. A genomic variant

may involve a single nucleotide substitution, a small insertion or deletion (indel), or larger structural changes.

Accurate identification of these variants is essential for studying genetic diseases, cancer, population diversity, and inherited disorders.

Traditional variant-calling approaches generally use statistical models to distinguish genuine genetic variation from sequencing errors. Deep learning has introduced an alternative approach in which models learn patterns directly from large collections of sequencing data. A prominent example is DeepVariant, which uses a deep convolutional neural network to analyze sequencing read information and determine whether a genomic position contains a true variant. Poplin et al. (2018) demonstrated that DeepVariant could outperform several existing variant-calling methods and could be adapted to different sequencing technologies and experimental designs.

AI-based variant analysis is particularly valuable because sequencing data contain substantial noise. Machine learning models can learn relationships between sequencing quality, read alignment, base-calling errors, and genuine genetic variation. This can improve the sensitivity and specificity of variant detection and reduce the amount of manual review required.

2.1.2 Gene Expression and Transcriptomic Analysis

AI is also increasingly used to analyse gene expression data. Gene expression studies examine how actively genes are transcribed under particular biological conditions. RNA sequencing (RNA-seq) can generate large datasets containing expression measurements for thousands of genes across numerous samples. AI algorithms can classify samples, identify expression patterns, detect abnormal expression, and predict relationships between genes.

Machine learning can be used to distinguish healthy and diseased tissues based on gene-expression signatures. For example, supervised learning models can be trained using labelled datasets in which samples are already associated with particular diseases or biological conditions. Unsupervised learning methods can subsequently identify naturally occurring groups or molecular subtypes without predefined labels. Such approaches are especially relevant to cancer genomics, where tumors may contain molecularly distinct subgroups that respond differently to treatment. Recent research highlights AI applications in

gene-expression profiling, transcriptomic analysis, and disease prediction (Sultana et al., 2026).

2.1.3 Prediction of Gene Function and Regulatory Elements

Another important application is predicting the functional consequences of DNA sequences. Although protein-coding regions have traditionally received considerable attention, a large proportion of the human genome consists of non-coding regions that participate in gene regulation. Identifying promoters, enhancers, transcription-factor binding sites, splice sites, and other regulatory elements is therefore a major challenge.

Deep learning models are capable of learning sequence patterns associated with regulatory activity. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), transformers, and related architectures can process DNA sequences and predict biological properties. These models can help researchers understand how particular DNA sequences influence gene expression and cellular function (Zou et al., 2019).

Recent genomic foundation models have expanded this capability by considering much longer stretches of DNA simultaneously. For example, AlphaGenome has been developed to predict multiple molecular consequences of DNA sequence variation, including gene expression, RNA splicing, chromatin accessibility, transcription-factor binding, and other regulatory signals. Such models illustrate the movement toward integrated AI systems capable of analysing multiple genomic processes within a single framework.

2.1.4 AI in Disease-Associated Variant Interpretation

Finding genetic variants is only the first step; researchers must also determine whether a variant is biologically meaningful or potentially associated with disease. The human genome contains millions of variants, and only small proportions have well-established clinical significance. AI can assist in prioritising variants by combining sequence information with population data, gene annotations, functional measurements, and other biological features.

Machine learning models can assign scores to candidate variants and estimate their potential functional impact. This is particularly useful in rare-disease research, where researchers may need to examine numerous variants in a patient's genome to identify the most plausible disease-causing candidate. AI-

assisted interpretation can therefore reduce the search space and help researchers focus experimental validation on the most promising variants.

2.1.5 AI for Cancer Genomics and Precision Medicine

Cancer is another major area in which AI-based genomic analysis has demonstrated considerable potential. Cancer genomes contain combinations of mutations, copy-number alterations, structural variants, and changes in gene expression. AI can integrate these molecular characteristics to classify tumors, identify biomarkers, predict disease progression, and support treatment selection.

In precision medicine, genomic information can potentially be combined with clinical characteristics to estimate how an individual may respond to a particular therapy. AI models can identify patterns across large patient datasets that may not be apparent through single-variable analysis. Reviews of AI applications in cancer genomics have reported growing use of machine learning and deep learning for genomic classification, biomarker discovery, prognosis, and treatment-related prediction (Khan et al., 2024).

2.1.6 Multi-Omics Integration

Modern life-science research increasingly combines genomics with transcriptomics, epigenomics, proteomics, metabolomics, and clinical information. These datasets differ in scale, structure, and biological meaning, making their integration challenging. AI provides computational approaches for identifying relationships across multiple molecular layers.

Deep learning and other machine learning methods can integrate heterogeneous datasets to construct more comprehensive representations of biological systems. Such approaches may reveal relationships between DNA variants, gene expression, cellular states, and disease phenotypes. Current research identifies multi-omics integration as one of the major emerging applications of AI in genomic analysis (Olawade et al., 2025; Sultana et al., 2026).

2.1.7 Challenges and Future Directions

Despite its potential, AI-based genomic analysis faces several limitations. Genomic datasets may contain batch effects, missing information, sequencing errors, population biases, and inconsistent data standards. Deep learning

models can also require substantial computational resources and large, high-quality training datasets. Another important concern is interpretability: highly complex models may produce accurate predictions without providing a sufficiently clear biological explanation of how those predictions were generated. This issue is particularly important when AI is used in clinical genomic interpretation (Eraslan et al., 2019; Olawade et al., 2025).

Future developments are likely to focus on explainable AI, multimodal and multi-omics foundation models, privacy-preserving learning, and models capable of analysing increasingly long genomic sequences. The integration of AI with experimental genomics may also accelerate the discovery and validation of disease-associated genetic mechanisms. However, AI predictions should complement rather than replace biological experimentation and expert interpretation.

Overall, AI is transforming genomic data analysis from a predominantly rule-based computational process into a data-driven and predictive discipline. Its applications extend from variant calling and gene-expression analysis to regulatory genomics, disease prediction, cancer research, and multi-omics integration. As genomic datasets continue to expand, AI is likely to become increasingly important for converting large-scale genomic information into biologically meaningful knowledge and clinically relevant insights.

2.2 MACHINE LEARNING IN DNA AND RNA SEQUENCING

The rapid development of high-throughput sequencing technologies has generated enormous volumes of DNA and RNA data. Although next-generation sequencing (NGS) and long-read sequencing platforms can produce millions to billions of sequence reads, converting these raw data into biologically meaningful information remains computationally challenging. Machine learning (ML), particularly deep learning (DL), has therefore become an important component of modern genomics and bioinformatics. ML algorithms can learn complex patterns from sequencing data and assist in base calling, sequence classification, variant detection, transcript quantification, splicing analysis, and error correction. Unlike conventional rule-based approaches, ML models can learn relationships directly from large datasets and can be adapted to different sequencing technologies and biological problems (Eraslan et al., 2019; Zou et al., 2019). (Figure 2.2)

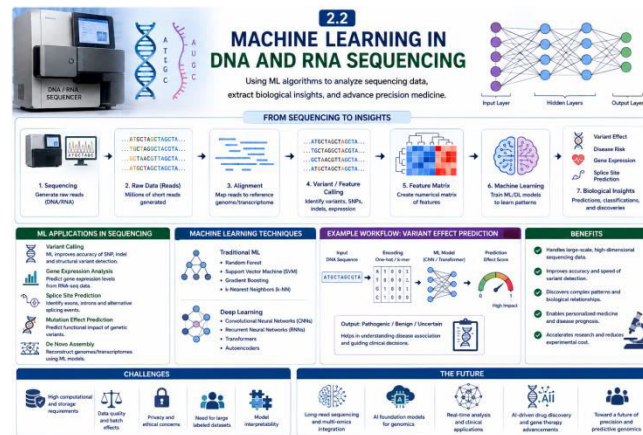


Figure 2.2: Machine Learning in DNA and RNA Sequencing

2.2.1 Machine Learning for DNA Sequencing

One of the earliest and most important applications of ML in sequencing is base calling. Base calling refers to the conversion of raw signals generated by a sequencing instrument into nucleotide sequences such as adenine (A), cytosine (C), guanine (G), and thymine (T). This is particularly important for nanopore sequencing, where DNA or RNA molecules pass through a nanopore and produce changes in electrical current. Neural-network-based algorithms analyze these electrical signals and predict the corresponding nucleotide sequence. Current nanopore basecalling systems use neural-network architectures, including recurrent neural networks and transformer-based approaches, to capture information across sequencing signals and improve prediction accuracy (Oxford Nanopore Technologies, 2026).

ML is also widely used in variant calling, which involves identifying differences between an individual's sequence and a reference genome. These differences may include single-nucleotide variants (SNVs), single-nucleotide polymorphisms (SNPs), insertions and deletions (InDels), and larger structural variants. Traditional variant callers generally rely on statistical models and manually defined rules. In contrast, deep-learning systems can learn sequence and read-alignment patterns associated with true genetic variants.

A prominent example is DeepVariant, which uses a deep convolutional neural network (CNN) to identify variants from aligned sequencing reads. The

system converts information surrounding candidate variants into representations resembling images and then uses a neural network to classify the likely genotype. Poplin et al. (2018) demonstrated that DeepVariant could achieve high accuracy across different sequencing technologies and experimental designs.

ML-based variant calling is particularly valuable when sequencing data contain technical errors, low coverage, repetitive genomic regions, or complex sequence patterns. Recent reviews indicate that AI-based callers such as DeepVariant, Clair, Clairvoyante, DNAscope, Medaka, and related systems are increasingly being investigated for accurate SNP, InDel, and structural-variant detection across short- and long-read sequencing platforms (Sultana et al., 2025).

2.2.2 Machine Learning for RNA Sequencing

RNA sequencing (RNA-seq) generates information about the transcripts expressed in a biological sample. ML can support several stages of RNA-seq analysis, including read processing, transcript quantification, isoform identification, splicing prediction, expression analysis, and biological classification.

One important task is transcript quantification, where sequencing reads are used to estimate the abundance of different transcripts or isoforms. RNA molecules originating from the same gene may undergo alternative splicing, producing multiple transcript variants. This creates ambiguity because a sequencing read may be compatible with more than one transcript. Computational methods such as Salmon use statistical inference and bias-aware models to estimate transcript abundance efficiently from RNA-seq reads (Patro et al., 2017). Although Salmon is primarily a statistical computational method rather than a deep-learning system, it illustrates the broader role of computational learning and probabilistic inference in extracting quantitative information from sequencing data.

Deep learning is particularly useful for RNA splicing analysis. Alternative splicing determines which exons are included or excluded from mature RNA molecules and can substantially influence protein production and cellular function. SpliceAI, developed by Jaganathan et al. (2019), uses a deep neural

network to predict splice junctions directly from pre-mRNA sequence. The model can identify sequence changes that may disrupt normal splicing, including variants located outside conventional splice-site regions. This demonstrates how ML can connect DNA sequence information with downstream RNA-processing consequences.

ML can also assist in distinguishing biologically meaningful expression patterns from technical noise in RNA-seq datasets. Modern genomic datasets may contain substantial variation caused by sequencing depth, batch effects, sample preparation, and other technical factors. Proper model design and validation are therefore essential because biological datasets often violate assumptions commonly used in conventional statistical machine-learning workflows (Whalen et al., 2022).

2.2.3 Advantages and Challenges

The major advantage of ML in DNA and RNA sequencing is its ability to analyze high-dimensional data and recognize complex patterns that may be difficult to define using manually designed rules. ML can improve sequencing accuracy, accelerate variant identification, support interpretation of regulatory and splicing signals, and facilitate the analysis of increasingly large genomic datasets. Deep learning models can also integrate sequence context over long genomic regions, providing opportunities for more comprehensive genome interpretation (Eraslan et al., 2019; Zou et al., 2019).

However, several limitations remain. ML models require large, high-quality and representative training datasets. Biases in training data can lead to poor performance in underrepresented populations or sequencing technologies. Model interpretability is another concern because complex neural networks may provide accurate predictions without clearly explaining the biological reasoning behind them. Computational requirements, data standardization, reproducibility, and independent validation are additional challenges. Consequently, ML predictions should generally be evaluated alongside established bioinformatics pipelines and experimental evidence rather than being treated as unquestionable biological conclusions.

Overall, machine learning is transforming DNA and RNA sequencing from a largely rule-driven analytical process into a data-driven predictive discipline.

Its applications now extend from converting raw sequencing signals into nucleotide sequences to detecting genetic variants and interpreting RNA processing. As sequencing technologies continue to generate larger and more complex datasets, integration of ML with conventional bioinformatics, experimental validation, and explainable AI is likely to become increasingly important for reliable genomic research and precision medicine.

2.3 PROTEIN STRUCTURE PREDICTION AND BIOMARKER DISCOVERY

Protein structure prediction and biomarker discovery are two important areas in which artificial intelligence (AI) is transforming genomics and bioinformatics. Proteins perform most of the functional activities within cells, and their three-dimensional (3D) structures determine how they interact with DNA, RNA, metabolites, drugs, and other proteins. At the same time, biomarkers-measurable biological characteristics associated with a disease, prognosis, or treatment response-are essential for disease diagnosis and precision medicine. The rapid growth of genomic, transcriptomic, proteomic, and clinical datasets has created an environment in which AI can identify complex biological patterns that may be difficult to detect using conventional computational approaches. (Figure 2.3)

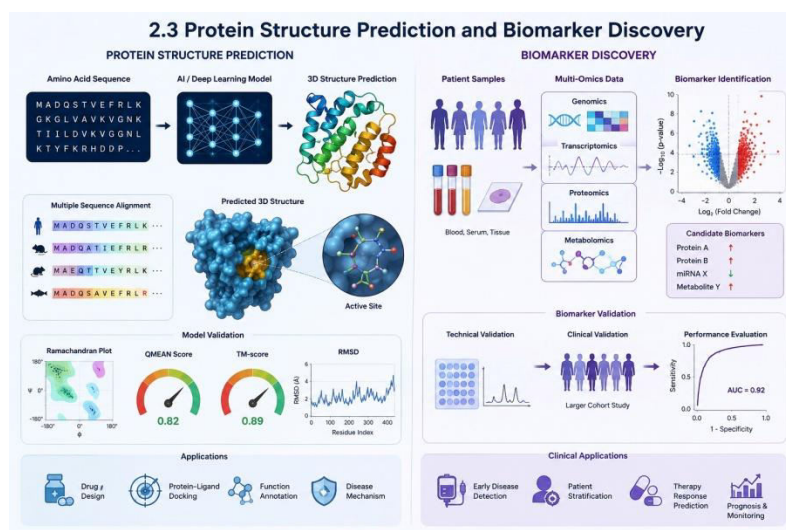


Figure 2.3: Protein Structure Prediction and Biomarker Discovery

2.3.1 AI-Based Protein Structure Prediction

Protein structure prediction involves determining the three-dimensional configuration of a protein from its amino acid sequence. Traditionally, experimental techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy have been used to determine protein structures. Although these methods can provide highly detailed structural information, they can be expensive, technically demanding, and time-consuming. Consequently, the number of experimentally determined structures has historically been much smaller than the number of known protein sequences (Cramer, 2021).

Deep learning has significantly changed this situation. One of the most influential developments has been AlphaFold2, developed by DeepMind. The system uses deep neural networks to infer structural relationships from amino acid sequences and evolutionary information. Its performance in the Critical Assessment of Structure Prediction (CASP) demonstrated that AI could generate highly accurate predictions for many protein structures, bringing computational prediction much closer to experimental structural biology (Jumper et al., 2021).

The importance of this development extends beyond simply obtaining a predicted protein shape. Structural information can help researchers investigate protein function, identify potential binding sites, understand disease-associated mutations, and generate hypotheses for drug development. Large-scale protein structure resources have also made structural information available for proteins for which experimental structures are unavailable, thereby expanding the practical scope of bioinformatics research.

More recently, AlphaFold3 has extended AI-based prediction beyond individual protein structures. It uses a diffusion-based architecture to model complexes containing proteins, nucleic acids, small molecules, ions, and modified residues. This capability is particularly important because biological functions usually depend on molecular interactions rather than isolated proteins. AlphaFold3 therefore represents a movement from protein structure prediction toward broader biomolecular structure and interaction prediction (Abramson et al., 2024).

However, AI-generated structures should not automatically be considered equivalent to experimentally determined structures. Predictions may be less reliable for intrinsically disordered regions, alternative conformations, unusual protein states, or situations that are poorly represented in training data. Therefore, structural predictions should be interpreted using confidence scores and, where necessary, validated experimentally.

2.3.2 AI in Biomarker Discovery

Biomarker discovery involves identifying biological features that can distinguish healthy individuals from patients, classify disease subtypes, predict disease progression, or indicate whether a patient is likely to respond to a particular treatment. Genomics and bioinformatics provide enormous quantities of potential biomarker candidates through technologies such as whole-genome sequencing, RNA sequencing, microarrays, proteomics, and metabolomics.

The major challenge is the high dimensionality of these datasets. A genomic or transcriptomic dataset may contain thousands or millions of variables while the number of patient samples may be comparatively small. Conventional statistical methods can therefore struggle with complex interactions and multiple-testing problems. Machine learning (ML) provides an alternative by learning relationships between molecular features and clinical outcomes. Nevertheless, ML models can overfit, meaning that a model may perform well on its training data but fail to generalize to independent patient populations (Ng et al., 2023).

AI-based biomarker discovery commonly involves several stages. First, biological data are collected and preprocessed to remove technical noise and inconsistencies. Second, relevant molecular features are selected using statistical methods, regularization, dimensionality reduction, or feature-importance techniques. Third, machine-learning or deep-learning models are trained to identify relationships between molecular features and a target clinical outcome. Finally, candidate biomarkers require independent validation using external datasets and, ideally, experimental or clinical studies.

An important emerging direction is multi-omics biomarker discovery, where genomics, transcriptomics, proteomics, metabolomics, imaging, and clinical

information are integrated. Such integration can provide a more comprehensive representation of disease biology than any single data type. Graph machine learning, for example, can help represent biological entities as interconnected networks and integrate information across different omics layers (Valous et al., 2024).

Deep learning can also support biomarker discovery by identifying complex patterns in high-dimensional biological data. In cancer research, multimodal AI approaches can combine molecular information with pathology images, radiological data, and clinical records to identify predictive or prognostic biomarkers. Such approaches are increasingly relevant to personalized medicine because biomarkers can potentially help determine which patients are at higher risk, which therapies may be effective, and how a disease is likely to progress (Steyaert et al., 2023).

2.3.3 Integration of Protein Structure Prediction and Biomarker Discovery

Protein structure prediction and biomarker discovery can complement one another. Genomic variants may alter the amino acid sequence of a protein, potentially affecting its structure, stability, or molecular interactions. AI-generated structural models can therefore help researchers investigate how disease-associated mutations might influence protein function. Conversely, biomarker analysis can identify genes or proteins associated with particular diseases, after which structural prediction can be used to explore their biological roles and potential therapeutic relevance.

This integrated approach is particularly promising in cancer and rare-disease research. AI can identify disease-associated genes or molecular signatures from large datasets and subsequently use structural information to investigate the corresponding proteins. In precision medicine, the combination of genomic information, molecular biomarkers, protein structures, and clinical characteristics may ultimately support more individualized diagnosis and treatment strategies (Mumtaz et al., 2023).

Despite its potential, AI-driven biomarker discovery requires careful validation. Reproducibility, dataset bias, population differences, model interpretability, data privacy, and the risk of false discoveries remain

important concerns. A computationally identified biomarker should therefore be regarded as a candidate until it has been independently validated. Similarly, AI-generated protein structures should support rather than replace experimental investigation.

Overall, AI is changing protein structure prediction and biomarker discovery from relatively slow, specialized activities into increasingly scalable computational processes. AlphaFold and related systems have demonstrated the ability of deep learning to predict biologically meaningful molecular structures, while machine learning and multi-omics approaches are expanding the discovery of disease-associated biomarkers. The future integration of structural biology, genomics, proteomics, and AI is likely to strengthen drug discovery, disease characterization, and precision medicine. The greatest impact will come not from AI operating independently, but from computational predictions being combined with biological expertise and rigorous experimental validation.

2.4 PERSONALIZED MEDICINE AND PRECISION GENOMICS

Personalized medicine represents a major shift from the traditional “one-size-fits-all” approach to healthcare toward strategies that consider the biological and clinical characteristics of individual patients. Precision genomics is an important component of this transformation because it uses information contained within an individual's genome to improve disease diagnosis, risk assessment, prevention, treatment selection, and therapeutic monitoring. Rather than treating patients solely according to disease classification, precision medicine seeks to identify biologically meaningful differences between individuals and disease subtypes. Genomic information can therefore help clinicians select treatments that are more likely to be effective and less likely to cause adverse reactions (Ashley, 2016; Sadee et al., 2023). (Figure 2.4)

2.4 Personalized Medicine and Precision Genomics

The right treatment for the right patient at the right time

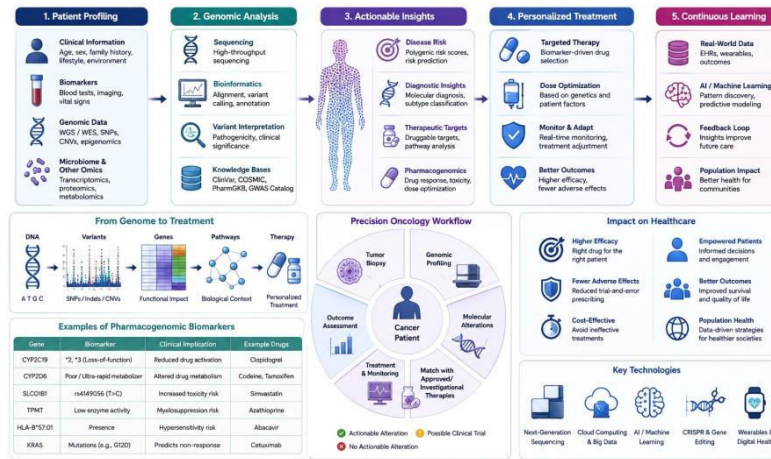


Figure 2.4: Personalized Medicine and Precision Genomics

2.4.1 Role of Genomics in Personalized Medicine

Genomics examines the complete set of genetic information within an organism and investigates how genes interact with one another and with environmental factors. Advances in next-generation sequencing (NGS), whole-genome sequencing (WGS), whole-exome sequencing (WES), and high-throughput molecular profiling have made it increasingly feasible to obtain large-scale genomic information for clinical applications. According to the World Health Organization (WHO), genomics has contributed to advances in genetic medicine, including pharmacogenomics and targeted therapies, while the development of the human pangenome is expanding the representation of human genetic diversity (WHO, 2026).

Precision genomics generally involves several stages: biological sample collection, DNA or RNA sequencing, bioinformatic processing, identification of genetic variants, interpretation of their clinical significance, and integration of genomic findings with clinical information. Variants may include single-nucleotide variants, insertions and deletions, copy-number variations, and larger structural changes. The interpretation of these variants is particularly important because the presence of a genetic variation does not necessarily indicate that an individual will develop a disease or respond to a particular therapy. Clinical interpretation therefore requires integration of genomic

evidence with phenotype, family history, environmental exposure, and other molecular information (Ashley, 2016).

2.4.2 Pharmacogenomics and Treatment Selection

One of the most established applications of precision genomics is pharmacogenomics-the study of how genetic variation influences an individual's response to medicines. Genetic differences can affect drug absorption, metabolism, transport, target interaction, efficacy, and toxicity. For example, variation in genes encoding drug-metabolizing enzymes can cause individuals to process the same medication at different rates. Consequently, a standard dose may be ineffective for one patient but produce toxicity in another.

Pharmacogenomic testing can help identify patients who are more likely to benefit from a particular medication or who may require an alternative drug or modified dosage.

Pirmohamed (2023) notes that the availability of relatively accessible genotyping technologies has made the clinical implementation of pharmacogenomics an important pathway toward mainstream genomic medicine. Sadee et al. (2023) further emphasize that personalized medicine increasingly combines genomic information with other biological and environmental factors rather than relying on a single genetic marker.

2.4.3 Precision Oncology

Cancer is one of the most prominent areas in which precision genomics has demonstrated clinical value. Tumors contain genetic alterations that influence their growth, progression, and response to therapy. Genomic profiling of tumor tissue or circulating tumor DNA can identify molecular alterations that may be associated with particular targeted therapies. This approach allows cancer to be classified not only according to its anatomical location but also according to its molecular characteristics.

Artificial intelligence (AI) is increasingly important in this area because genomic and clinical datasets can contain millions of data points. Machine-learning models can assist in identifying patterns linking genetic alterations with treatment response, disease progression, and potential therapeutic targets. Recent work highlights the convergence of machine learning and genomics in

precision oncology, particularly as large clinicogenomic datasets and molecular diagnostic workflows become increasingly integrated into clinical decision-making (Reardon et al., 2026). AI can therefore act as an analytical layer connecting genomic data with clinically relevant predictions.

2.4.4 AI-Enabled Precision Genomics

The integration of AI with precision genomics is particularly valuable because genomic datasets are high-dimensional and difficult to interpret manually. Machine-learning and deep-learning approaches can analyze relationships among genetic variants, gene expression, molecular pathways, patient characteristics, and treatment outcomes. AI can support variant prioritization, disease-risk prediction, biomarker discovery, drug-response prediction, and identification of potential therapeutic targets.

AI-based approaches may also contribute to multimodal precision medicine by integrating genomic information with electronic health records, medical imaging, pathology, proteomics, and other clinical datasets. Such integration can produce a more comprehensive representation of the patient and disease state. Research on AI and pharmacogenomics indicates that machine-learning approaches can help identify relationships between genetic characteristics and drug responses, potentially supporting more individualized treatment decisions (Intelligent Pharmacy, 2024).

2.4.5 Benefits of Personalized Medicine

Precision genomics offers several potential benefits. First, it may improve diagnostic accuracy, particularly for rare and genetically heterogeneous disorders. Second, genomic biomarkers can help identify patients who are more likely to respond to specific therapies. Third, pharmacogenomics may reduce adverse drug reactions by identifying individuals with altered drug metabolism or sensitivity. Fourth, genomic risk assessment can support earlier surveillance and preventive interventions for individuals with inherited susceptibility to particular diseases.

However, precision medicine should not be interpreted as predicting every aspect of an individual's health from DNA alone. Disease development is often influenced by complex interactions among genes, lifestyle, environment, age,

and social determinants of health. Consequently, genomic information should generally be interpreted alongside other clinical and biological evidence.

2.4.6 Challenges and Ethical Considerations

Despite its potential, precision genomics faces significant scientific, clinical, economic, and ethical challenges. A major difficulty is the interpretation of variants whose clinical significance remains uncertain. Sequencing technologies can identify enormous numbers of genetic differences, but determining which variants are clinically actionable remains challenging (Ashley, 2016).

Another concern is population representation. Genomic databases have historically contained unequal representation of different populations, which can reduce the accuracy of genetic risk prediction and variant interpretation for underrepresented groups. Sadee et al. (2023) identify population bias and limitations in real-world validation as important barriers to the wider application of pharmacogenomics.

Privacy and governance are also critical because genomic data are uniquely personal and potentially identifiable. The WHO's guidance on human genome data emphasizes responsible collection, access, use, and sharing, with particular attention to transparency, equity, privacy, and individual and collective rights (WHO, 2024). Healthcare systems must therefore establish strong data-security practices, informed-consent procedures, appropriate governance frameworks, and safeguards against discrimination.

2.4.7 Future Directions

The future of personalized medicine is likely to involve increasingly integrated genomic, clinical, environmental, and real-world data. AI will play an important role in transforming these large datasets into clinically useful predictions. Improvements in sequencing technologies, population-scale genomic databases, multimodal AI, and clinical decision-support systems may enable more precise disease prevention and treatment.

Ultimately, the objective of precision genomics is not simply to sequence more genomes but to convert genomic information into reliable and equitable clinical benefits. Successful implementation will require collaboration among

geneticists, clinicians, bioinformaticians, AI researchers, policymakers, and patients. As genomic technologies become more accessible, ensuring that their benefits are scientifically valid, clinically useful, ethically responsible, and available across diverse populations will be essential for realizing the full promise of personalized medicine.

CHAPTER 3

Artificial Intelligence in Drug Discovery and Development

3.1 AI-BASED DRUG DESIGN AND TARGET IDENTIFICATION

Artificial intelligence (AI) has emerged as an important technology for improving the efficiency and precision of modern drug discovery. Traditional drug discovery is a lengthy and expensive process involving disease understanding, target identification, hit discovery, lead optimization, preclinical evaluation, and clinical development. AI-based approaches can assist researchers by analyzing large and heterogeneous datasets, identifying relationships that may be difficult to detect manually, predicting molecular properties, and prioritizing promising drug candidates. Machine learning (ML), deep learning (DL), natural language processing, graph neural networks, and generative AI are increasingly being integrated into computer-aided drug discovery workflows (Singh et al., 2024). (Figure 3.1)

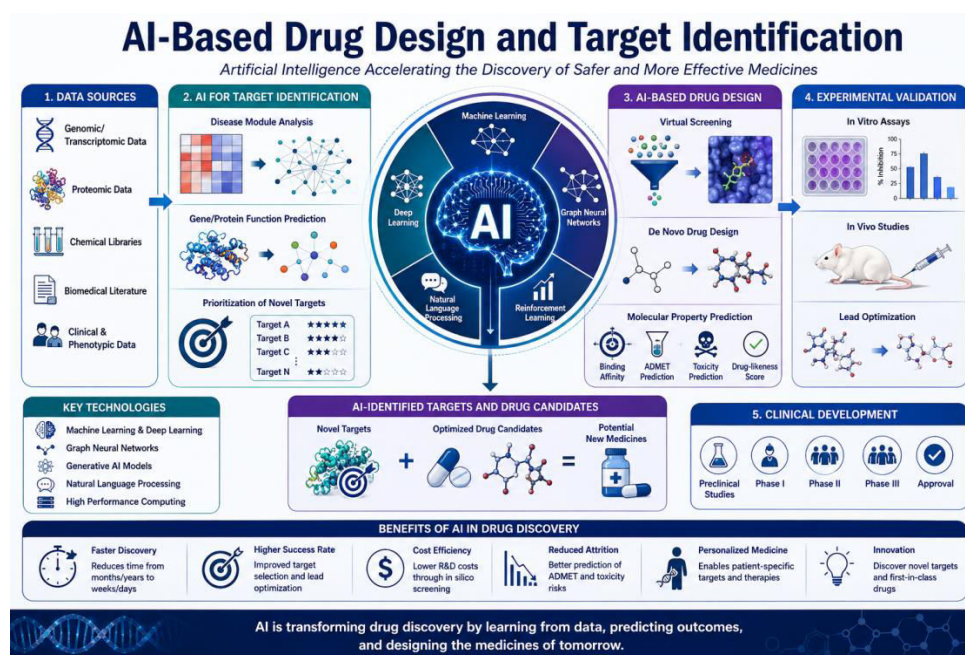


Figure 3.1: AI-Based Drug Design and Target Identification

3.1.1 AI-Based Target Identification

Target identification is one of the earliest and most critical stages of drug discovery. A drug target is generally a biological molecule, such as a protein, enzyme, receptor, nucleic acid, or signaling component, whose modulation can produce a desired therapeutic effect. Selecting an inappropriate target can

result in significant failure during later stages of development. AI can improve target identification by integrating information from genomics, transcriptomics, proteomics, metabolomics, electronic health records, biomedical literature, molecular interaction databases, and clinical datasets.

Machine learning models can identify associations between genes, proteins, diseases, and phenotypes. Network-based approaches are particularly useful because diseases are often caused by complex interactions among multiple biological pathways rather than by a single molecular component. AI can therefore analyze biological networks and identify highly connected or functionally important molecules that may represent potential therapeutic targets. Recent approaches increasingly combine network biology, multimodal data integration, and perturbation-aware modelling to identify context-specific and functionally actionable targets (Patne et al., 2024).

Natural language processing (NLP) also contributes to target discovery by extracting information from scientific publications, patents, clinical reports, and biomedical databases. Millions of scientific articles contain information about gene–disease relationships, protein interactions, mechanisms of disease, and experimental findings. AI-based text-mining systems can process this information rapidly and identify previously unrecognized relationships. This allows researchers to construct knowledge graphs connecting diseases, genes, proteins, pathways, drugs, and phenotypic outcomes. Such approaches can help prioritize targets for experimental validation.

Another important contribution of AI is the prediction of protein structure. The development of AlphaFold demonstrated the ability of deep-learning methods to predict highly accurate three-dimensional protein structures from amino-acid sequences (Jumper et al., 2021). Structural information is highly valuable for target assessment because understanding the three-dimensional architecture of a protein can reveal binding pockets, active sites, conformational features, and opportunities for interaction with drug molecules. However, predicted structures do not automatically establish that a protein is therapeutically relevant; biological validation and experimental evidence remain essential.

3.1.2 AI-Based Drug Design

Once a promising target has been identified, AI can support the design and prioritization of molecules capable of interacting with that target. Conventional computer-aided drug design commonly relies on molecular docking, quantitative structure–activity relationships (QSAR), molecular dynamics, and pharmacophore modelling. AI extends these approaches by learning relationships between molecular structures and biological properties from large datasets.

One important application is virtual screening. AI models can predict the probability that a compound will interact with a particular target, allowing researchers to prioritize a smaller number of molecules for laboratory testing. This can reduce the number of compounds requiring expensive experimental screening. Deep-learning architectures, including graph neural networks and attention-based models, can represent molecules as graphs or sequences and predict properties such as binding affinity, activity, toxicity, solubility, and pharmacokinetic characteristics (Lavecchia, 2024).

Generative AI has further expanded the possibilities of drug design. Instead of simply selecting compounds from an existing chemical library, generative models can create new molecular structures according to predefined objectives. Variational autoencoders, generative adversarial networks, recurrent neural networks, normalizing flows, and Transformer-based models have been investigated for de novo molecular design. These systems can generate molecules optimized for multiple characteristics, such as target affinity, selectivity, molecular stability, and drug-like properties (Lavecchia, 2024).

Structure-based generative design is particularly promising because AI models can incorporate information about the three-dimensional structure of a target protein while generating candidate molecules. Recent research has focused on protein-structure-based molecular generation, enabling models to explore chemical structures compatible with specific binding sites (Xu et al., 2024).

3.1.2 Advantages and Challenges

AI-based drug design can potentially reduce the time required for hit identification, improve molecular prioritization, explore larger chemical

spaces, and support more systematic decision-making. It can also facilitate drug repurposing by identifying new relationships between existing drugs, molecular targets, and diseases. Nevertheless, AI does not eliminate the fundamental uncertainties of drug discovery.

Model performance depends heavily on the quality, quantity, and representativeness of training data. Data scarcity, experimental bias, inconsistent measurements, model interpretability, and difficulties in translating computational predictions into biological effects remain important challenges (Gangwal et al., 2024).

Therefore, AI should be viewed as an augmentation of experimental drug discovery rather than a complete replacement for laboratory research. The most effective workflow combines computational prediction with biochemical assays, cellular experiments, structural studies, animal models, and ultimately clinical evaluation. AI can narrow the search space and prioritize hypotheses, while experimental science determines whether those predictions are biologically and clinically meaningful.

Overall, AI-based target identification and drug design represent a major evolution in pharmaceutical research. By integrating biological knowledge with advanced computational models, AI enables researchers to move from conventional trial-and-error approaches toward more data-driven and hypothesis-guided discovery. Continued improvements in multimodal data integration, protein structure prediction, generative modelling, explainable AI, and experimental validation are expected to further strengthen the role of AI in the development of safer, more selective, and potentially more effective therapeutics.

3.2 VIRTUAL SCREENING AND MOLECULAR MODELING

Virtual screening and molecular modeling have become important components of modern computer-aided drug discovery. Instead of experimentally testing thousands or millions of chemical compounds individually, computational methods can prioritize molecules that are more likely to interact with a selected biological target. This approach helps researchers reduce the size of experimental libraries, lower screening costs, and focus laboratory resources on the most promising candidates. Virtual screening is broadly divided into

ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS). LBVS relies primarily on information from known active compounds, whereas SBVS uses the three-dimensional structure of a biological target to identify molecules capable of fitting into its binding site (Cheng et al., 2012; Kučera, 2016). (Figure 3.2)

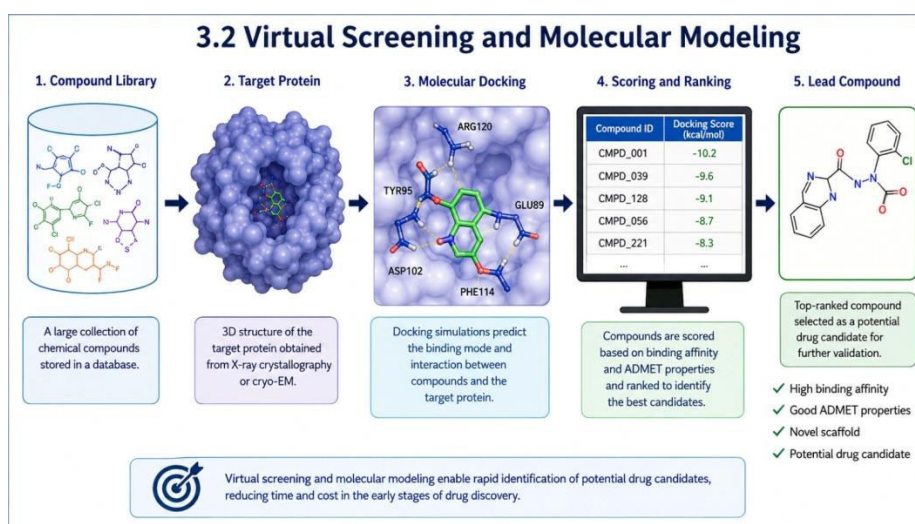


Figure 3.2: Virtual Screening and Molecular Modeling

3.2.1 Principles of Virtual Screening

The central objective of virtual screening is to identify potential “hit” compounds from a large chemical library. In a typical structure-based workflow, researchers first identify an appropriate protein target and determine or predict its three-dimensional structure. Structural information may be obtained using experimental techniques such as X-ray crystallography, nuclear magnetic resonance, or cryo-electron microscopy, or through computational protein-structure prediction and molecular modeling. The target is then prepared by assigning appropriate protonation states, removing or retaining water molecules as appropriate, adding hydrogen atoms, and defining the binding pocket (Cheng et al., 2012).

The compound library is similarly prepared. Molecules may be filtered according to molecular weight, physicochemical properties, chemical stability, or undesirable structural features. Pharmacophore filters, similarity searches,

and drug-likeness rules can further reduce the number of compounds that require computationally intensive analysis. This preprocessing is important because the quality of the input library can substantially influence the performance of a virtual-screening campaign.

3.2.2 Molecular Docking and Modeling

Molecular docking is one of the most widely used approaches in structure-based virtual screening. Docking attempts to predict how a small molecule, or ligand, can fit within the binding site of a protein and estimates the strength or favorability of the resulting interaction using a scoring function. The docking process generally involves two major components: pose generation and scoring. Pose generation explores possible orientations and conformations of the ligand, while scoring functions rank the resulting complexes according to predicted interaction quality (Pinzi & Rastelli, 2019).

Molecular modeling extends beyond docking by providing a broader computational representation of molecular structure, dynamics, and interactions. Molecular mechanics, molecular dynamics (MD), pharmacophore modeling, homology modeling, and free-energy calculations can all contribute to drug-design workflows. MD simulations, for example, can provide information about protein flexibility and ligand–protein interactions over time. This is particularly relevant because proteins are not rigid structures; they undergo conformational changes that can influence ligand binding. Consequently, ensemble docking, in which multiple protein conformations are used during screening, can sometimes improve the representation of receptor flexibility (Gorgulla et al., 2023).

3.2.3 Integration of Artificial Intelligence

Artificial intelligence (AI) is increasingly being integrated into virtual screening and molecular modeling. Machine-learning and deep-learning algorithms can learn relationships between molecular structures, biological activities, and target interactions from large datasets. Molecules can be represented using molecular fingerprints, descriptors, graphs, three-dimensional coordinates, or learned embeddings. Deep-learning approaches can then predict properties such as binding affinity, activity, toxicity, or drug–target interactions and use these predictions to prioritize compounds (Jiménez-Luna et al., 2021).

One important advantage of AI-assisted screening is its ability to process chemical and biological datasets at a scale that would be difficult to handle manually. AI models can be used as an initial filtering stage before computationally expensive docking or molecular simulation. Alternatively, docking results can be combined with machine-learning models to improve compound ranking. Recent research has increasingly explored hybrid workflows combining molecular docking, pharmacophore modeling, machine learning, and interpretable AI techniques for lead identification and optimization (Taha & Daoud, 2026).

3.2.3 Workflow of AI-Assisted Virtual Screening

A modern AI-assisted virtual-screening workflow can be summarized as:

Target identification → Protein-structure preparation → Chemical-library preparation → Molecular/property filtering → AI-based prioritization → Molecular docking → Rescoring → Molecular dynamics/free-energy analysis → Selection of candidate hits → Experimental validation

This multi-stage approach is generally more effective than relying on a single computational method. Recent developments have also enabled ultra-large virtual screening, in which computational workflows can examine hundreds of millions or even billions of compounds.

Such approaches combine highly efficient docking, machine learning, distributed computing, and intelligent compound prioritization to explore chemical space on an unprecedented scale (Chee et al., 2026).

3.2.4 Advantages and Limitations

Virtual screening offers several advantages. It can examine very large chemical libraries rapidly, prioritize compounds before laboratory testing, support the identification of structurally diverse molecules, and provide hypotheses about ligand–target interactions. It is particularly useful during the early stages of drug discovery, where selecting appropriate compounds for experimental testing can substantially influence the efficiency of subsequent development.

However, virtual screening is not a substitute for experimental validation. Docking scores do not necessarily correspond directly to experimentally

measured binding affinities, and scoring functions can produce false positives and false negatives. Protein flexibility, solvent effects, protonation states, metal ions, ligand entropy, and induced-fit effects can complicate accurate prediction of molecular recognition (Cheng et al., 2012). Moreover, AI models can inherit biases from their training datasets and may perform poorly when applied to chemical spaces or biological targets that differ substantially from those represented in the training data. Recent evaluations have therefore emphasized problems such as dataset bias, data leakage, limited generalizability, and the complex geometric nature of molecular recognition (Suay-García & Falcó, 2026).

3.2.5 Future Perspectives

The future of virtual screening is likely to involve deeper integration between AI, molecular simulation, structural biology, and experimental chemistry. Deep-learning-based docking and scoring functions, protein-structure prediction, generative molecular design, ultra-large-scale screening, and physics-based simulations are increasingly being combined into integrated computational pipelines. Recent reviews indicate that docking-based virtual screening continues to evolve from conventional library screening toward large-scale and AI-enhanced workflows (Zou et al., 2026).

Overall, virtual screening and molecular modeling provide a computational bridge between biological target identification and experimental drug discovery. Their greatest value lies not in replacing laboratory experiments, but in making those experiments more selective and information-rich. When combined with high-quality structural data, appropriate molecular modeling, AI-based prioritization, and rigorous experimental validation, these approaches can accelerate the identification and optimization of promising drug candidates.

3.3 AI IN CLINICAL TRIALS AND DRUG REPURPOSING

Artificial intelligence (AI) is increasingly becoming an important component of modern drug development, particularly in clinical trials and drug repurposing. Traditional clinical research is often associated with lengthy timelines, high costs, complex protocols, difficulties in patient recruitment, and substantial rates of trial failure. AI and machine learning (ML) can address several of these limitations by processing large and heterogeneous datasets,

identifying patterns that may not be readily visible to researchers, and supporting evidence-based decisions throughout the clinical development process (Zhang et al., 2025). (Figure 3.3)

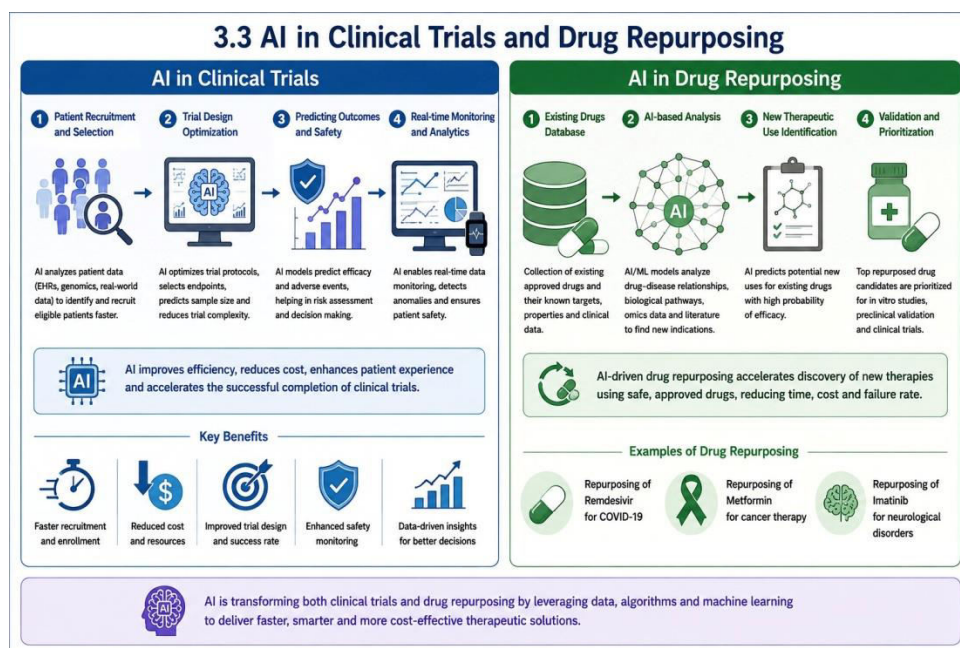


Figure 3.3: AI in Clinical Trials and Drug Repurposing

3.3.1 AI in Clinical Trials

Clinical trials represent one of the most important and resource-intensive stages of drug development. AI can be applied across almost the entire clinical-trial lifecycle, beginning with protocol development and continuing through patient recruitment, trial monitoring, data analysis, and post-trial evaluation. During trial design, machine-learning algorithms can analyze historical clinical-trial data to identify factors associated with successful or unsuccessful studies. Such analysis can support the selection of appropriate endpoints, sample sizes, eligibility criteria, treatment arms, and trial sites. AI-based approaches may also facilitate adaptive trial designs in which aspects of a study can be modified in response to accumulating evidence while maintaining appropriate statistical and regulatory controls.

One of the most significant applications of AI is patient recruitment and selection. Clinical trials frequently experience delays because eligible participants are difficult to identify and enroll. Natural language processing (NLP) can analyze electronic health records (EHRs), medical notes, laboratory results, and other clinical information to identify individuals who potentially satisfy trial eligibility criteria. Machine-learning models can additionally stratify participants according to demographic, clinical, genetic, or disease characteristics. This can improve the identification of suitable participants and potentially reduce recruitment time and screening costs.

AI can also support patient stratification and personalized clinical trials. By integrating genomic, proteomic, imaging, clinical, and real-world data, AI systems can identify subgroups that are more likely to respond to a particular intervention or experience adverse effects. This approach is particularly valuable in diseases with substantial biological heterogeneity, such as cancer, neurological disorders, and autoimmune diseases. More precise patient selection can improve the probability of detecting treatment effects and may contribute to more efficient clinical development.

Another important application is trial monitoring and data management. AI can continuously analyze information generated through EHRs, laboratory systems, wearable devices, imaging platforms, and electronic patient-reported outcomes. Algorithms can identify missing or inconsistent data, detect unusual patterns, and support the early identification of safety signals. AI can also assist in endpoint assessment and automate certain repetitive data-processing tasks. Nevertheless, these systems require rigorous validation because errors, biased datasets, or inappropriate algorithmic assumptions can affect clinical conclusions.

AI also has potential in predicting clinical-trial outcomes. By learning from previous trials, disease characteristics, drug properties, and patient-level information, predictive models can estimate the probability of trial success or failure. Such predictions can help pharmaceutical companies prioritize promising candidates and allocate research resources more efficiently. However, prediction should be considered a decision-support function rather than a substitute for clinical judgment, experimental evidence, or regulatory assessment.

3.3.2 AI in Drug Repurposing

Drug repurposing, also called drug repositioning, involves identifying a new therapeutic application for an existing drug that was originally developed or approved for another indication. Because many repurposed drugs already have information regarding pharmacokinetics, toxicity, manufacturing, dosing, or human safety, repurposing can potentially reduce some of the time, cost, and uncertainty associated with developing an entirely new medicine (Fu et al., 2026).

AI provides powerful methods for identifying potential repurposing opportunities. Conventional approaches may depend heavily on individual discoveries or experimental screening, whereas AI can systematically integrate information from drug databases, disease databases, molecular networks, genomic datasets, scientific publications, EHRs, and real-world evidence. Machine-learning and deep-learning models can examine relationships among drugs, proteins, genes, diseases, pathways, and phenotypes to predict previously unrecognized drug–disease associations.

Knowledge graphs and graph neural networks (GNNs) are particularly promising for drug repurposing. These approaches represent biological and medical entities as interconnected networks, allowing algorithms to investigate complex relationships between drugs and diseases. For example, Huang et al. (2024) developed TxGNN, a graph-based foundation model designed to predict therapeutic candidates for diseases, including conditions with limited available treatments. The model illustrates how AI can extend drug-repurposing analysis beyond diseases for which well-established treatment options already exist.

Generative AI is another emerging approach. Large language models can analyze and synthesize information from extensive scientific and biomedical literature to generate hypotheses about potential therapeutic relationships. In one study, generative AI was used to prioritize potential drugs for Alzheimer’s disease, after which selected candidates were examined using large clinical datasets. The findings demonstrated how generative AI can contribute to hypothesis generation and candidate prioritization, although computational predictions still require rigorous experimental and clinical validation.

AI-supported repurposing can therefore establish a computational-to-clinical pipeline: first, AI identifies candidate drugs; second, molecular and laboratory experiments evaluate biological plausibility; third, observational or real-world data can provide supporting evidence; and finally, appropriately designed clinical trials determine efficacy and safety for the new indication. This integration is important because an AI prediction alone does not establish therapeutic effectiveness.

Despite its considerable potential, AI-based clinical trials and drug repurposing face several challenges. Data quality, missing information, population bias, privacy, algorithmic transparency, reproducibility, cybersecurity, and regulatory acceptance remain important concerns. AI models trained on limited or nonrepresentative datasets may generate misleading predictions. Furthermore, complex models can be difficult to interpret, creating challenges when researchers or regulators need to understand why a particular recommendation was produced. Current discussions involving regulators therefore emphasize validation, human oversight, transparency, and appropriate governance when AI is incorporated into drug development.

Overall, AI is not expected to replace clinical researchers, physicians, pharmacologists, or regulatory authorities. Instead, its greatest value lies in augmenting human expertise by rapidly processing complex information and identifying patterns or hypotheses that can guide further investigation. When combined with high-quality biomedical data, rigorous validation, and appropriate regulatory oversight, AI can make clinical trials more efficient and expand the search for new therapeutic applications of existing medicines. Consequently, AI-driven clinical research and drug repurposing are likely to become increasingly important components of precision medicine and next-generation drug development.

3.4 CHALLENGES AND FUTURE SCOPE IN PHARMACEUTICAL AI

Artificial intelligence (AI) is increasingly being integrated into pharmaceutical research, covering target identification, virtual screening, molecular design, lead optimization, toxicity prediction, clinical trial design, drug repurposing, manufacturing, and post-market surveillance. The technology offers the potential to reduce the time and cost associated with conventional drug

development while enabling researchers to explore chemical and biological spaces that are difficult to investigate experimentally. However, the practical impact of pharmaceutical AI remains constrained by limitations in data quality, model reliability, biological complexity, validation, regulation, and ethical governance. Recent research emphasizes that improvements in computational performance alone are insufficient; successful pharmaceutical AI requires integration with experimental science and clinical decision-making (Ghislat et al., 2024; Su et al., 2026).

3.4.1 Challenges in Pharmaceutical AI

1. Data quality, availability, and heterogeneity: The effectiveness of AI models depends heavily on the quality and representativeness of training data. Pharmaceutical datasets are frequently heterogeneous, incomplete, biased, noisy, or relatively small. Experimental measurements generated by different laboratories may use different protocols, instruments, assay conditions, and reporting standards, making direct comparison difficult. In addition, negative experimental results are often underreported, producing datasets that may not accurately represent the complete chemical or biological landscape. Such problems can lead to biased models and unreliable predictions (Ghislat et al., 2024). Data scarcity is particularly important for rare diseases and novel biological targets, where only limited experimental observations may be available (Gangwal et al., 2024).

2. Limited interpretability and explainability: Many advanced AI systems, particularly deep-learning and generative models, operate as complex "black boxes." Although a model may accurately predict molecular activity, toxicity, or treatment response, researchers may not always understand why a particular prediction was generated. Lack of interpretability can reduce confidence among medicinal chemists, clinicians, regulators, and patients. Explainable AI (XAI) approaches are therefore becoming increasingly important for identifying the biological or chemical factors responsible for predictions and for supporting scientifically defensible decisions (Alam et al., 2024).

3. Translational and experimental validation: Strong performance on retrospective datasets does not necessarily translate into successful drug candidates. AI models may identify promising molecules computationally, but

these candidates must subsequently undergo synthesis, biochemical testing, pharmacokinetic studies, toxicological assessment, and clinical evaluation. Consequently, the gap between *in silico* prediction and real-world biological performance remains a major obstacle. Recent analyses highlight unrealistic benchmarks, inadequate prospective testing, and uncertainty estimation as important reasons why impressive computational results do not always produce clinically meaningful outcomes (Ghislat et al., 2024).

4. Regulatory uncertainty: Pharmaceutical products are subject to stringent regulatory requirements, and AI introduces additional questions concerning model credibility, validation, data provenance, reproducibility, change management, and accountability. Regulatory agencies must determine how AI-generated evidence should be assessed without unnecessarily restricting innovation. In January 2025, the U.S. Food and Drug Administration (FDA) issued draft guidance proposing a risk-based credibility assessment framework for AI models used to support regulatory decisions concerning drug and biological product safety, effectiveness, or quality. This represents an important step toward establishing clearer expectations for pharmaceutical AI, although regulatory frameworks will continue to evolve as technologies mature (FDA, 2025).

5. Privacy, security, and ethical concerns: Pharmaceutical AI increasingly uses clinical records, genomic information, real-world evidence, and other sensitive datasets. Improper data handling may create risks related to privacy, cybersecurity, consent, and unauthorized secondary use. AI systems may also reproduce existing demographic or socioeconomic biases if their training data are not representative. The World Health Organization emphasizes that AI in health should be developed around principles including transparency, accountability, inclusiveness, equity, and protection of human autonomy (WHO, 2021).

6. Lack of interdisciplinary expertise: Effective pharmaceutical AI requires collaboration among pharmacologists, medicinal chemists, biologists, clinicians, data scientists, computational scientists, and regulatory experts. A technically sophisticated model may still be unsuitable if it does not address a meaningful pharmaceutical problem. Therefore, future development must

focus not simply on building increasingly complex algorithms but on designing AI systems around clearly defined scientific and clinical needs.

3.4.2 Future Scope of Pharmaceutical AI

The future of pharmaceutical AI is likely to move from isolated prediction tools toward integrated, multimodal, and experimentally connected platforms. Generative AI is expected to support the design of novel molecules with desired potency, selectivity, physicochemical characteristics, and safety profiles. Instead of screening millions of existing compounds alone, generative systems can propose previously unexplored molecular structures for experimental evaluation (Gangwal & Lavecchia, 2024).

Another important direction is the integration of multimodal biological data, including genomics, transcriptomics, proteomics, imaging, clinical records, chemical structures, and real-world evidence. Combining these data sources may improve understanding of disease mechanisms and enable more accurate target identification and patient stratification. AI may consequently contribute to precision medicine by identifying which patient subgroups are most likely to respond to particular therapies.

Foundation models and large language models (LLMs) also have considerable future potential. These systems can assist researchers in extracting knowledge from scientific literature, generating hypotheses, interpreting experimental results, designing molecules, and integrating information from diverse biological databases. Recent developments indicate that AI is expanding beyond individual stages of drug discovery toward broader workflows spanning target identification, molecular design, lead optimization, clinical research, and drug repurposing (Su et al., 2026).

Future pharmaceutical AI will also increasingly incorporate human-in-the-loop systems, in which AI provides predictions and recommendations while scientists retain decision-making authority. This approach can combine computational scalability with human scientific judgment and experimental expertise. Similarly, AI-enabled automated laboratories may create closed-loop systems in which algorithms design experiments, robotic platforms conduct them, experimental data are automatically analyzed, and subsequent experiments are selected based on the results.

Regulatory science is another major area of future development. Risk-based validation frameworks, standardized reporting practices, model monitoring, and international regulatory harmonization will be essential for trustworthy adoption. The FDA has already recognized increasing AI use throughout nonclinical, clinical, manufacturing, and post-marketing stages of the drug product life cycle and is working toward responsible integration of AI into drug development (FDA, 2026).

Overall, the future of pharmaceutical AI is not likely to involve the complete replacement of researchers. Rather, AI is expected to function as an advanced scientific partner capable of processing large datasets, generating hypotheses, prioritizing experiments, and identifying patterns beyond conventional human analysis. The greatest value will emerge when high-quality data, explainable algorithms, experimental validation, interdisciplinary expertise, and responsible regulation are developed together. Thus, the long-term success of pharmaceutical AI will depend less on algorithmic sophistication alone and more on its ability to produce reproducible, biologically meaningful, clinically relevant, and regulatory acceptable outcomes.

CHAPTER 4

AI Applications in Healthcare and Medical Diagnostics

4.1 AI IN DISEASE DIAGNOSIS AND PREDICTION

Artificial Intelligence (AI) has emerged as an important technology in modern healthcare, particularly in the areas of disease diagnosis, risk assessment, prognosis, and clinical decision support. Unlike conventional statistical approaches that often depend on a limited set of predefined variables, AI and machine learning (ML) systems can process large and complex datasets, including electronic health records (EHRs), medical images, laboratory results, genomic information, physiological signals, and clinical notes. By identifying patterns within these datasets, AI systems can assist healthcare professionals in detecting diseases earlier, estimating disease risk, and supporting more personalized clinical decisions (Rajkomar et al., 2018). (Figure 4.1)

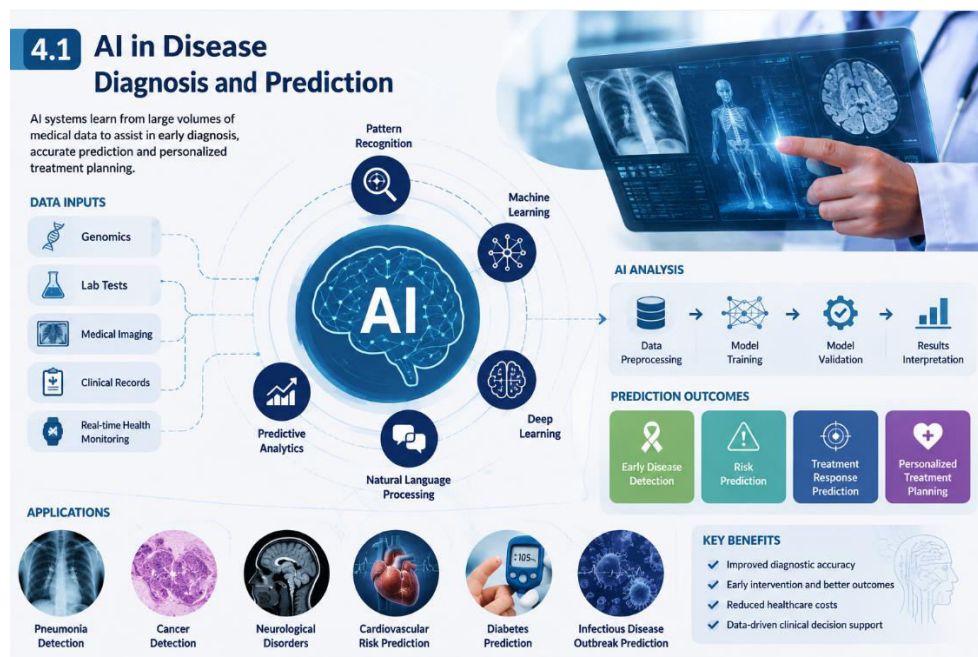


Figure 4.1: AI in Disease Diagnosis and Prediction

4.1.1 Role of AI in Disease Diagnosis

Disease diagnosis traditionally depends on a combination of patient history, physical examination, laboratory investigations, imaging, and clinical expertise. AI can complement these processes by rapidly analysing large volumes of information and identifying subtle patterns that may be difficult to recognize consistently through manual assessment. Machine learning

algorithms, particularly deep learning and neural networks, have demonstrated considerable potential in image-based diagnosis.

Medical imaging is one of the most developed applications of AI in diagnostics. Deep learning models can analyse X-rays, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, mammograms, and dermatological images to identify abnormalities. For example, Esteva et al. (2017) demonstrated that a deep convolutional neural network trained on more than 129,000 clinical images could classify skin lesions at a level comparable to dermatologists in a specific experimental setting. Such systems illustrate how AI can function as a screening and decision-support tool, particularly when large numbers of images must be reviewed.

AI has also demonstrated significant potential in cancer detection. McKinney et al. (2020) evaluated an AI system for breast cancer screening using mammographic images from the United Kingdom and United States. The system reduced false-positive and false-negative results compared with conventional screening approaches and outperformed the average performance of six radiologists in an independent reader study. These findings indicate that AI may help improve screening efficiency and diagnostic consistency, although performance in controlled research settings should not automatically be interpreted as equivalent to effectiveness in every clinical environment.

Another important application is ophthalmology. AI-based image-analysis systems can examine retinal photographs for signs of diabetic retinopathy and other eye diseases. Such technology can potentially expand screening capacity, particularly in settings where specialist ophthalmologists are limited. AI can identify patients requiring further examination and thereby support early intervention.

4.1.2 AI for Disease Prediction and Risk Stratification

The role of AI extends beyond identifying diseases that are already present. Predictive models can estimate the probability that a patient will develop a disease or experience an adverse clinical event. This distinction is important because prediction allows healthcare systems to move from a reactive model of treatment toward a more preventive and proactive approach.

Electronic health records provide a particularly valuable source of predictive information. Patient records contain longitudinal information such as diagnoses, medications, laboratory measurements, procedures, clinical observations, and hospital admissions. Deep learning models can analyse these data to identify relationships between historical patient information and future outcomes. Rajkomar et al. (2018), for example, developed deep learning models using raw EHR data from more than 216,000 adult patients across two academic medical centres. Their models demonstrated strong predictive performance for outcomes including in-hospital mortality, unplanned readmission, prolonged hospital stay, and discharge diagnoses.

AI-based prediction can also be applied to chronic diseases. Models may estimate an individual's risk of conditions such as cardiovascular disease, diabetes, kidney disease, or complications associated with existing illnesses. Risk scores generated from AI systems can help clinicians identify high-risk patients who may benefit from closer monitoring, additional testing, lifestyle interventions, or preventive treatment.

4.1.3 Multimodal Diagnosis and Clinical Decision Support

The future of AI-driven diagnosis increasingly involves multimodal systems capable of combining different forms of medical information. Instead of analysing an X-ray independently, for example, an AI system could potentially combine imaging data with laboratory results, symptoms, medication history, genetic information, and previous diagnoses. This approach may provide a more comprehensive representation of a patient's clinical condition.

AI can also support differential diagnosis by generating possible explanations for a patient's symptoms based on available clinical information. In this role, AI should be viewed primarily as a decision-support technology rather than an independent replacement for medical professionals. The final diagnosis may require contextual understanding, physical examination, additional investigations, and professional judgment.

4.1.4 Benefits of AI-Based Diagnosis and Prediction

Several potential advantages explain the growing interest in AI-enabled healthcare. First, AI can process large datasets rapidly, reducing the time required for certain analytical tasks. Second, it can provide consistent

application of learned patterns, which may reduce some forms of human variability. Third, AI can support early detection by identifying risk signals before symptoms become severe.

Fourth, automated screening may improve access to diagnostic services in areas where specialist resources are limited. Finally, predictive analytics can assist healthcare providers in prioritizing patients according to estimated clinical risk.

The regulatory landscape also demonstrates the increasing integration of AI into clinical technologies. The U.S. Food and Drug Administration (FDA) maintains an AI-enabled medical device list containing devices authorized for marketing in the United States, with applications across areas such as radiology and cardiovascular care.

4.1.5 Challenges and Limitations

Despite its potential, AI-based diagnosis is associated with important limitations. The quality of an AI system depends heavily on the quality, representativeness, and completeness of its training data. If datasets underrepresent particular populations, AI models may perform differently across demographic or clinical groups. Privacy and security are also major concerns because healthcare AI requires access to sensitive patient information.

Another challenge is explainability. Some deep learning systems operate as complex "black boxes," making it difficult for clinicians to understand why a particular prediction was generated. This can create problems involving trust, accountability, and clinical responsibility. AI systems may also produce false-positive or false-negative results, both of which can have significant consequences in medical settings.

The World Health Organization emphasizes that AI in healthcare should be developed and deployed with attention to ethics, human rights, transparency, accountability, safety, and equity (WHO, 2021). Therefore, high diagnostic accuracy alone is not sufficient. AI systems require rigorous clinical validation, appropriate regulatory oversight, continuous monitoring, and meaningful involvement of healthcare professionals.

AI is transforming disease diagnosis and prediction by enabling the analysis of medical images, electronic health records, physiological signals, and other complex healthcare data. Its greatest potential lies not simply in replacing conventional diagnostic procedures, but in augmenting clinical expertise through faster analysis, earlier detection, risk stratification, and personalized decision support. Evidence from dermatology, breast cancer screening, and EHR-based prediction demonstrates the potential of AI to improve specific diagnostic and predictive tasks (Esteva et al., 2017; McKinney et al., 2020; Rajkomar et al., 2018). However, responsible implementation requires high-quality datasets, clinical validation, transparency, privacy protection, and human oversight. Consequently, the future of AI-enabled diagnosis is best understood as a collaborative model in which intelligent computational systems and healthcare professionals work together to improve the accuracy, efficiency, accessibility, and quality of patient care.

4.2 DEEP LEARNING IN MEDICAL IMAGING

Medical imaging is one of the most important areas in which artificial intelligence (AI) has demonstrated significant potential. Modern healthcare generates large volumes of images through X-ray, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, mammography, positron emission tomography (PET), retinal photography, digital pathology, and other imaging technologies. Traditionally, interpretation of these images has depended heavily on the expertise and experience of radiologists, pathologists, ophthalmologists, and other specialists. Deep learning, a major branch of machine learning, has introduced computational methods capable of learning complex patterns directly from medical images and supporting clinicians in detection, classification, segmentation, and diagnosis. A major review of the field identified convolutional neural networks (CNNs) as particularly influential across image classification, object detection, segmentation, registration, and related medical imaging tasks (Litjens et al., 2017). (Figure 4.2)

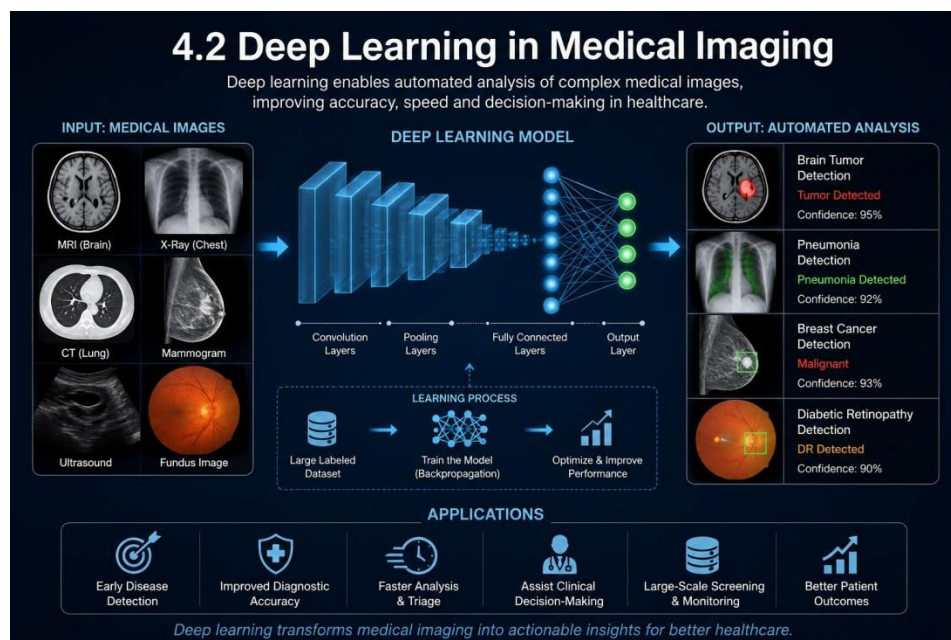


Figure 4.2: Deep Learning in Medical Imaging

4.2.1 Fundamentals of Deep Learning for Medical Images

Deep learning uses multilayer neural networks to learn representations from data. Unlike conventional computer-aided diagnosis systems, which often require manually designed features such as texture, shape, intensity, or edge characteristics, deep neural networks can learn relevant features automatically during training. CNNs are particularly suitable for medical imaging because convolutional operations can identify spatial patterns at different levels of abstraction. Early layers may detect simple structures such as edges and textures, while deeper layers can learn more complex anatomical or pathological patterns.

The general workflow involves image acquisition, preprocessing, model training, validation, and clinical evaluation. During training, large collections of labelled images are provided to the model. The algorithm adjusts its internal parameters to minimize errors between predicted and known outcomes. Once trained, the model can process previously unseen images and generate classifications, probability scores, segmentation masks, or other clinical outputs. Depending on the application, architectures such as CNNs, U-Net,

three-dimensional CNNs, recurrent networks, and transformer-based models may be employed.

4.2.2 Disease Detection and Classification

One of the most established applications of deep learning is automated disease detection. Algorithms can classify medical images into categories such as normal or abnormal, benign or malignant, or specific disease classes. This capability is particularly useful in screening programs where healthcare professionals must examine large numbers of images.

For example, Gulshan et al. (2016) developed and validated a deep learning algorithm for detecting diabetic retinopathy from retinal fundus photographs. Their work demonstrated that deep learning could achieve high diagnostic performance and highlighted the potential of automated retinal screening for identifying patients requiring further ophthalmic assessment.

Deep learning has also been applied extensively to cancer detection. Esteva et al. (2017) trained a deep convolutional neural network using a large collection of skin disease images and reported performance comparable to dermatologists for classifying important categories of skin cancer. The study demonstrated how image-based AI could potentially support preliminary screening and referral, particularly where specialist expertise is limited.

4.2.3 Medical Image Segmentation

Segmentation involves identifying and separating specific anatomical structures, organs, lesions, or tumors from surrounding tissues. It is essential for treatment planning, disease monitoring, quantitative measurements, and surgical applications. Deep learning models can generate pixel-level or voxel-level maps that identify areas of clinical interest.

For example, segmentation algorithms can identify brain tumors in MRI scans, lung nodules in CT images, organs in abdominal scans, and lesions in retinal images. U-Net and related encoder-decoder architectures have become particularly important because they combine high-level semantic information with spatial information required for accurate localization. Automated segmentation can reduce the time required for manual delineation and improve consistency between assessments.

4.2.4 Deep Learning in CT, MRI, and X-Ray Imaging

Deep learning is increasingly being used across multiple imaging modalities. In chest radiography, AI systems can assist in identifying abnormalities such as pneumonia, tuberculosis, nodules, and other pulmonary conditions. In CT imaging, three-dimensional deep learning models can analyze volumetric data rather than treating each image slice independently.

Ardila et al. (2019) demonstrated an end-to-end three-dimensional deep learning approach for lung cancer screening using low-dose chest CT. Their work illustrated the potential of deep learning to analyze complex volumetric information and support cancer-risk assessment.

MRI presents another important application because it produces high-dimensional images containing detailed anatomical and functional information. Deep learning can assist with tumor identification, tissue segmentation, image reconstruction, and detection of neurological disorders. In addition, AI-based reconstruction methods may reduce image acquisition or processing requirements while maintaining clinically useful image quality.

4.2.5 Clinical Decision Support and Workflow Optimization

The role of deep learning is not limited to replacing manual image interpretation. Increasingly, AI is being developed as a clinical decision-support technology that works alongside healthcare professionals. Systems can prioritize urgent cases, highlight suspicious regions, provide quantitative measurements, and assist physicians in comparing current images with previous examinations.

Research has shown that the effect of AI assistance can vary among clinicians and clinical tasks. A large 2024 study involving radiologists found that AI assistance did not produce uniform improvements across all clinicians and that incorrect AI predictions could negatively influence diagnostic performance. This emphasizes that AI should generally be considered a support mechanism rather than an infallible substitute for clinical expertise.

4.2.6 Challenges and Limitations

Despite its considerable potential, deep learning in medical imaging faces important limitations. High-quality labelled datasets are expensive and time-consuming to create because expert clinicians often need to annotate images.

Differences in scanners, imaging protocols, hospitals, patient populations, and disease prevalence can also reduce model performance when an algorithm is transferred to a new environment.

Another major concern is generalizability. A model that performs well on its development dataset may perform less effectively on external populations. Recent research has demonstrated that medical imaging AI can encounter limitations related to fairness and real-world generalization, reinforcing the importance of diverse datasets and independent validation (Yang et al., 2024).

Interpretability is another challenge. Deep neural networks can produce highly accurate predictions while providing limited explanations for how those predictions were reached. In healthcare, where decisions can directly affect patient treatment, clinicians need information about model reliability, uncertainty, and limitations. Privacy, cybersecurity, regulatory compliance, and responsibility for errors are additional considerations.

Regulatory oversight is consequently becoming increasingly important. The U.S. Food and Drug Administration (FDA) maintains a growing list of AI-enabled medical devices and evaluates their safety and effectiveness before authorization. The agency also recognizes applications of AI in image acquisition, image processing, early disease detection, diagnosis, prognosis, and risk assessment.

4.2.7 Future Perspectives

The future of deep learning in medical imaging is likely to involve increasingly multimodal and clinically integrated systems. Future models may combine radiological images with electronic health records, laboratory results, genomic information, and clinical histories to produce more comprehensive assessments. Transformer-based architectures, self-supervised learning, foundation models, and federated learning may further reduce dependence on extensively labelled datasets while improving adaptability and privacy.

Overall, deep learning has transformed medical image analysis from predominantly rule-based computer assistance toward data-driven systems capable of learning sophisticated patterns directly from clinical images. Its greatest value is likely to emerge through human-AI collaboration, in which algorithms provide rapid and consistent computational analysis while

healthcare professionals retain responsibility for clinical interpretation and patient management. Continued progress will therefore depend not only on improving model accuracy but also on external validation, fairness, transparency, safety, regulatory oversight, and successful integration into real clinical workflows.

4.3 SMART HEALTHCARE SYSTEMS AND WEARABLE DEVICES

Smart healthcare represents a major transformation in the way health services are delivered, monitored, and managed. It combines technologies such as artificial intelligence (AI), machine learning (ML), the Internet of Things (IoT), cloud computing, big-data analytics, mobile applications, and connected medical devices to create healthcare systems that are more responsive, personalized, and data-driven. Within this ecosystem, wearable devices have become particularly important because they enable continuous collection of physiological and behavioral information outside conventional clinical environments. The World Health Organization (WHO) identifies smart wearables, remote monitoring, artificial intelligence, big-data analytics, and related digital technologies as having potential to improve diagnosis, treatment decisions, self-management, and person-centred care (WHO, 2021). (Figure 4.3)

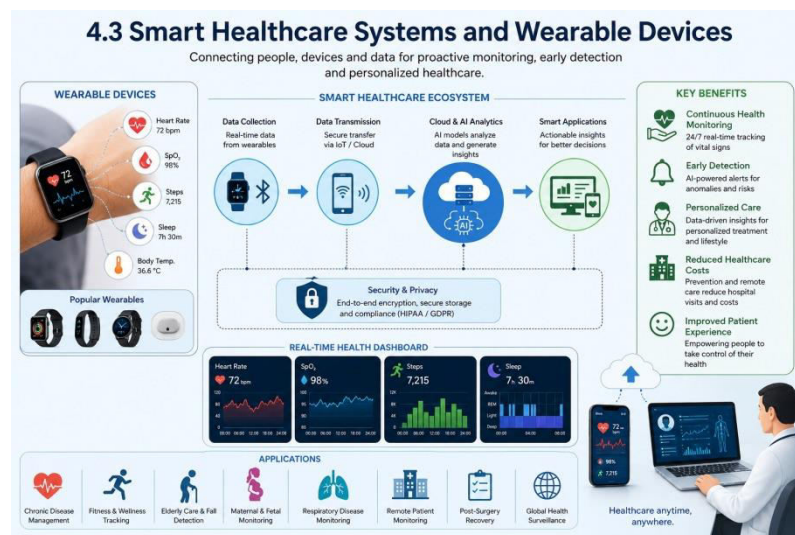


Figure 4.3: Smart Healthcare Systems and Wearable Devices

4.3.1 Concept of Smart Healthcare Systems

A smart healthcare system can be understood as an interconnected digital environment in which health-related data are continuously collected, transmitted, analyzed, and converted into clinically meaningful information. Traditional healthcare largely depends on periodic consultations and measurements taken during hospital or clinic visits. Smart healthcare, in contrast, enables healthcare professionals to obtain information about patients between clinical encounters. This is particularly valuable for individuals with chronic diseases, elderly patients, and people requiring long-term monitoring.

The basic architecture generally includes four interconnected layers: data acquisition, data communication, intelligent analysis, and clinical decision support. Sensors embedded in wearable devices collect physiological data such as heart rate, blood oxygen saturation, body temperature, physical activity, sleep patterns, and in appropriate devices electrocardiographic signals. The collected information can then be transferred through smartphones, wireless networks, or IoT platforms to cloud or edge-computing environments. AI and ML algorithms analyze these data to identify abnormalities, recognize patterns, estimate risk, and generate alerts. Such systems can therefore transform raw measurements into actionable information for patients and healthcare professionals.

The WHO emphasizes that digital health systems should be developed around principles including interoperability, accessibility, privacy, security, reliability, equity, and sustainability.

4.3.2 Role of Wearable Devices

Wearable devices are electronic technologies designed to be worn on or attached to the body. Examples include smartwatches, fitness bands, smart rings, patches, chest straps, continuous glucose monitoring systems, and specialized biosensor platforms. These devices can continuously or periodically capture physiological and behavioral signals.

Commonly monitored parameters include heart rate, heart-rate variability, ECG, blood oxygen saturation, respiratory rate, physical activity, sleep duration, body temperature, and movement patterns. Wearables can therefore generate longitudinal health data that are difficult to obtain through occasional

clinical measurements. Research on cardiovascular applications demonstrates that AI can analyze wearable sensor signals to support disease detection, prediction, and risk assessment (Huang et al., 2022).

Wearables are also increasingly relevant to remote patient monitoring (RPM). Instead of requiring patients to remain in hospitals or repeatedly visit clinics, connected devices can transmit selected health measurements to healthcare professionals. AI algorithms can detect deviations from an individual's normal physiological patterns and potentially generate early warnings. This changes healthcare from a predominantly reactive model toward a more proactive and preventive model.

4.3.3 Artificial Intelligence in Wearable-Based Healthcare

AI is the analytical component that gives smart wearable systems much of their intelligence. Wearable devices can generate large volumes of continuous data, but raw data alone have limited clinical value. ML and deep-learning algorithms can identify relationships and patterns that may not be apparent through manual observation.

For example, an AI system can examine heart-rate and ECG patterns to identify possible arrhythmias, analyze movement patterns to assess mobility or neurological changes, or evaluate sleep and activity data to identify changes in an individual's usual behavior. AI can also combine multiple data streams rather than relying on a single measurement. Such multimodal analysis may provide a more comprehensive representation of an individual's health status.

A systematic review of AI approaches for smart health devices found that ML and deep learning are increasingly used to analyze data generated by wearable IoT devices, while also highlighting privacy, security, methodological, and implementation challenges (Aversano et al., 2024).

4.3.4 Applications in Healthcare

One important application is cardiovascular monitoring. Smart wearables can continuously measure heart rate and, in certain devices, obtain ECG signals. AI-based models can analyze these signals to identify abnormalities and estimate cardiovascular risk. Reviews of smart wearable research have reported growing evidence for applications in cardiovascular disease detection

and prediction, although further validation is required before widespread clinical adoption (Moshawrab et al., 2023).

A second major application is chronic disease management. Patients with diabetes, hypertension, cardiovascular disorders, and other long-term conditions can benefit from continuous or repeated monitoring. Rather than evaluating health only during scheduled appointments, healthcare providers can examine trends over time and potentially identify deterioration earlier.

Wearables are also being investigated for elderly care and neurological disorders. Changes in gait, movement, activity, and sleep can provide digital indicators of changes in functional health. These measurements may support monitoring of conditions in which behavioral or motor changes are clinically relevant.

Another important area is preventive and personalized healthcare. Longitudinal wearable data can contribute to the development of digital biomarkers-measurable characteristics derived from digital devices that may help characterize health or disease. Digital biomarkers can provide information about behavior and physiological functioning in everyday environments, supporting more continuous and personalized approaches to healthcare (Powell, 2024).

4.3.4 Benefits and Challenges

Smart healthcare systems offer several potential advantages. Continuous monitoring can improve the availability of patient information, support early detection, reduce unnecessary hospital visits, facilitate personalized interventions, and encourage patient participation in self-management. The integration of digital technologies can also strengthen continuity of care when data are securely exchanged across healthcare settings.

However, important challenges remain. Wearable measurements may be affected by sensor quality, body movement, device placement, environmental conditions, and differences between individuals. A 2025 scoping review found that although wearable-based remote monitoring research is expanding, evidence of clinical effectiveness remains limited and studies vary substantially in devices, conditions, monitoring protocols, and outcomes (Wearables research for continuous monitoring of patient outcomes, 2025).

Privacy and cybersecurity are equally important because wearable systems can continuously generate highly sensitive personal health information. AI models may also reproduce biases present in their training data, potentially producing unequal performance across populations. Furthermore, interoperability with electronic health records, regulatory approval, clinical validation, patient adherence, battery limitations, and the need for healthcare professionals to interpret AI-generated alerts remain significant implementation barriers.

4.3.5 Future Perspectives

The future of smart healthcare is likely to involve closer integration between wearable sensors, AI, edge computing, cloud platforms, electronic health records, and telemedicine. Advances in biosensing may allow wearable devices to measure increasingly diverse physiological signals, while AI may improve the interpretation of multimodal and longitudinal data. Federated learning and privacy-preserving technologies may also help organizations develop AI systems while reducing the need to centralize sensitive patient information.

Ultimately, the objective should not be to replace healthcare professionals but to augment their capabilities. WHO emphasizes that AI for health should be implemented responsibly, with appropriate governance, safety, ethics, equity, and regulation. Smart healthcare systems will therefore be most effective when technological innovation is combined with clinical evidence, human oversight, patient consent, data protection, and equitable access. The convergence of AI and wearable technology has the potential to move healthcare toward a more continuous, predictive, preventive, and personalized model of care, but responsible implementation and rigorous clinical validation will determine whether this potential can be translated into sustainable improvements in health outcomes.

4.4 ROBOTIC SURGERY AND INTELLIGENT CLINICAL SYSTEMS

Artificial intelligence (AI) is increasingly transforming healthcare by combining machine learning, computer vision, natural language processing, predictive analytics, and robotics with clinical knowledge. One of the most advanced areas of this transformation is the integration of AI with robotic surgery and intelligent clinical systems. Traditional robotic surgery primarily provides surgeons with enhanced visualization, dexterity, motion control, and

minimally invasive access, whereas AI adds capabilities such as pattern recognition, surgical workflow analysis, predictive assistance, and decision support. Consequently, the modern concept of robotic surgery is gradually shifting from simple computer-assisted manipulation toward intelligent, data-driven surgical systems (Topol, 2019; Cella et al., 2026). (Figure 4.4)

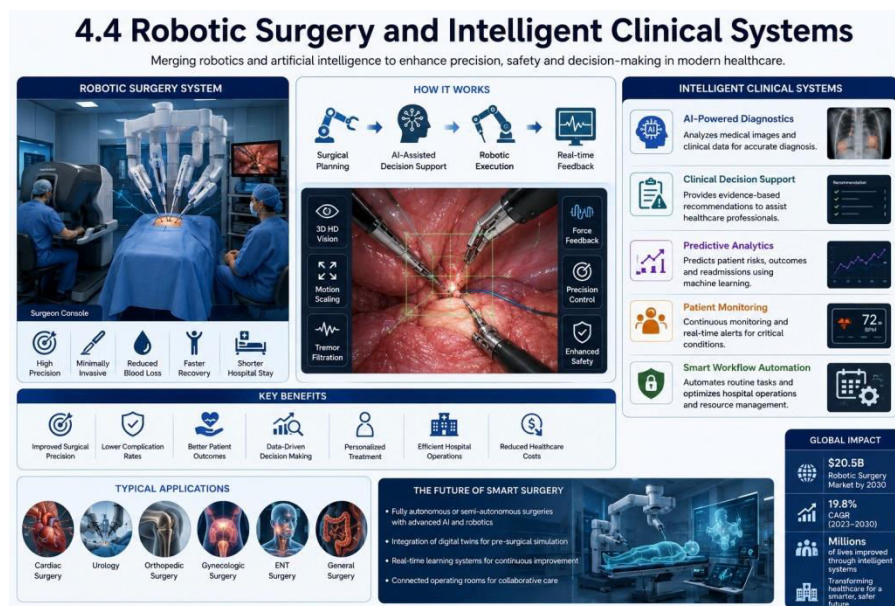


Figure 4.4: Robotic Surgery and Intelligent Clinical Systems

4.4.1 AI in Robotic Surgery

Robotic surgical systems generally consist of a surgeon-controlled console, robotic arms, specialized instruments, high-definition imaging, and computer-based control systems. AI can enhance these components by interpreting surgical images, identifying anatomical structures, recognizing surgical phases, predicting the next surgical action, and providing real-time assistance. Computer vision algorithms, for example, can analyze intraoperative video to identify tissues, blood vessels, organs, instruments, and pathological structures. Such capabilities may help surgeons maintain situational awareness and reduce cognitive workload during complex procedures (Vasey et al., 2023).

One important application is automated anatomical identification and surgical guidance. AI models can recognize anatomical landmarks and potentially highlight important structures during an operation. This capability may be particularly useful when anatomical structures are difficult to distinguish visually or when surgical anatomy varies between patients. AI can also support tumor localization and augmented-reality visualization, allowing clinically relevant information to be superimposed onto the surgeon's operative view (Cella et al., 2026).

Another important area is surgical skill assessment and training. Robotic systems generate large amounts of data concerning instrument movements, tissue interaction, operation time, camera movements, and surgical gestures. Machine-learning algorithms can analyze these data to evaluate surgical performance and identify patterns associated with expert or inexperienced performance. AI-enabled simulation platforms can therefore provide personalized feedback to trainees, potentially making surgical education more objective and continuous (Vasey et al., 2023).

AI can also contribute to intraoperative decision support. During surgery, intelligent systems may detect potentially dangerous situations, identify deviations from expected surgical workflows, and provide alerts or recommendations. Rather than replacing the surgeon, the near-term objective is generally to create assistive systems that improve perception, anticipate risks, and support human decision-making while maintaining surgeon control (Cella et al., 2026).

4.4.2 Towards Autonomous Robotic Surgery

A major long-term objective of surgical AI is increasing the level of robotic autonomy. Current systems largely operate under direct human supervision, but research is exploring progressively more autonomous tasks. These include automated tissue segmentation, instrument positioning, trajectory planning, suturing, tissue manipulation, and other repetitive subtasks. Research has demonstrated increasing interest in supervised autonomous systems, although fully autonomous surgery remains largely experimental and presents substantial technical, clinical, ethical, and regulatory challenges (Saeidi et al., 2022; Vasey et al., 2023).

The future development of autonomous surgery is likely to follow a gradual pathway rather than an immediate transition to completely independent robotic surgeons. AI systems may initially perform narrowly defined tasks under continuous human supervision. As reliability, validation, safety mechanisms, and regulatory frameworks improve, more complex tasks may become possible. Recent research identifies imitation learning, reinforcement learning, hybrid control systems, and vision-language-action models as potential technologies for increasing surgical autonomy (Cella et al., 2026).

4.4.3 Intelligent Clinical Systems

Beyond the operating room, AI is becoming an important component of intelligent clinical systems. These systems integrate patient records, laboratory results, medical images, physiological signals, clinical notes, and other sources of healthcare data to support diagnosis, prognosis, treatment planning, and patient monitoring. AI can process large and heterogeneous datasets more rapidly than conventional manual approaches, allowing clinicians to identify patterns that may otherwise be difficult to recognize. The broader impact of AI in medicine includes improved image interpretation, workflow optimization, clinical decision support, and patient-level data analysis (Topol, 2019).

Clinical decision-support systems (CDSS) are particularly important. An intelligent CDSS may generate alerts for abnormal laboratory values, estimate disease risk, identify potential drug interactions, prioritize patients, or assist clinicians in selecting appropriate diagnostic and therapeutic options. Predictive models can also be used for early identification of clinical deterioration and postoperative complications.

However, these systems should function as decision-support tools rather than unquestionable replacements for clinical judgment.

The effectiveness of intelligent clinical systems depends heavily on data quality and model reliability. Healthcare datasets can contain missing information, measurement errors, demographic imbalance, and institutional biases. A model trained primarily on data from one population or healthcare system may not perform equally well in another environment. Therefore, external validation, continuous monitoring, transparency, and appropriate human oversight are essential before clinical deployment (WHO, 2021).

4.4.4 Challenges and Ethical Considerations

The integration of AI into robotic surgery and clinical decision-making introduces important ethical and safety concerns. Errors in an AI-assisted diagnostic system or surgical robot can have direct consequences for patients. Questions concerning responsibility and liability also become more complex when clinical decisions involve clinicians, software developers, hospitals, and autonomous or semi-autonomous machines.

Explainability and transparency are equally important. Clinicians need sufficient information to understand the limitations and reliability of AI recommendations. In addition, patient privacy must be protected because intelligent clinical systems frequently depend on large quantities of sensitive health data. The World Health Organization emphasizes that AI in healthcare should be developed and deployed with human autonomy, safety, transparency, accountability, equity, and sustainability in mind (WHO, 2021).

Another challenge is the translation of promising laboratory research into routine clinical practice. Although numerous AI technologies demonstrate technical performance in controlled environments, real-world healthcare involves complex workflows, diverse patient populations, changing clinical conditions, and regulatory requirements. Recent research on AI-enabled surgical devices highlights the growing potential of AI for preoperative assessment, surgical planning, intraoperative decision-making, postoperative monitoring, and complication prediction, while also emphasizing the need for robust clinical validation (2026).

4.4.4 Future Outlook

The convergence of AI, robotics, medical imaging, and intelligent clinical systems is likely to create a more connected and adaptive healthcare environment. Future operating rooms may combine robotic platforms with computer vision, real-time analytics, augmented reality, predictive models, and intelligent clinical decision support. Similarly, hospitals may increasingly use AI systems that connect diagnosis, treatment planning, surgery, monitoring, and follow-up into a continuous data-driven workflow.

Nevertheless, the most realistic near-term model is human-AI collaboration, rather than complete replacement of healthcare professionals. AI can provide

computational speed, pattern recognition, and continuous data analysis, while clinicians contribute contextual understanding, ethical judgment, communication, and responsibility for patient-centered decisions. If these technologies are subjected to rigorous validation, appropriate regulation, continuous monitoring, and meaningful human oversight, robotic surgery and intelligent clinical systems can become important components of safer, more precise, and personalized healthcare.

CHAPTER 5

Artificial Intelligence in Agriculture and Environmental Life Sciences

5.1 AI IN PRECISION AGRICULTURE AND CROP MONITORING

Artificial Intelligence (AI) is increasingly transforming agriculture from a conventional, experience-based activity into a data-driven and precision-oriented production system. Precision agriculture refers to the use of digital technologies to observe, measure, analyze, and manage agricultural variability at a field or sub-field level. AI strengthens this approach by processing large volumes of information collected through satellites, unmanned aerial vehicles (UAVs), Internet of Things (IoT) sensors, weather stations, farm machinery, and field observations. By converting these data into actionable insights, AI can support farmers in making timely decisions related to irrigation, fertilization, crop protection, weed management, and harvesting. The Food and Agriculture Organization (FAO) identifies soil and crop monitoring and predictive analytics among important areas where AI can contribute to agricultural productivity and efficiency (FAO, 2017). (Figure 5.1)



Figure 5.1: AI in Precision Agriculture and Crop Monitoring

5.1.1 Role of AI in Precision Agriculture

Traditional agricultural management often applies the same quantity of water, fertilizers, pesticides, and other inputs across an entire field. However, crop

conditions can vary considerably because of differences in soil fertility, moisture, topography, pests, diseases, and microclimatic conditions. Precision agriculture attempts to identify these variations and provide site-specific management. AI algorithms can analyze spatial and temporal data to identify patterns that may not be easily recognized through conventional field inspection.

Machine learning (ML) algorithms such as support vector machines (SVM), random forests (RF), artificial neural networks (ANN), and gradient-based models are widely used for agricultural classification, prediction, and decision support. A 2024 review of remote sensing and machine learning applications in precision agriculture found that SVM and random forest were among the most frequently applied algorithms, while hyperspectral and UAV-based remote sensing were particularly important data sources (Zhang et al., 2024).

Deep learning has further expanded these capabilities. Convolutional neural networks (CNNs) can automatically extract visual features from agricultural images, making them useful for crop classification, disease recognition, weed detection, and plant phenotyping. More recently, transformer-based architectures and self-supervised learning approaches have gained attention because they can learn complex representations from large volumes of remote-sensing data (Silva et al., 2025).

5.1.2 AI-Based Crop Monitoring

Crop monitoring involves continuous observation of crop growth, health, water status, nutrient conditions, and environmental stress throughout the growing season. AI enables this monitoring to move beyond periodic manual inspection toward automated and near-real-time assessment.

Satellite imagery is one of the most important sources of information for large-scale crop monitoring. Platforms such as Sentinel-2 provide multispectral imagery that can be used to calculate vegetation indices, including the Normalized Difference Vegetation Index (NDVI). These indices provide information about vegetation vigor and can be combined with AI models to estimate crop growth, biomass, and yield. Recent research shows that AI combined with Sentinel-2 imagery is increasingly being used for field-level crop monitoring and yield estimation, although cloud cover, limited ground-

truth observations, and differences between geographical regions remain important challenges (Demissie et al., 2026).

UAVs or drones provide another important source of high-resolution crop information. Unlike satellites, UAVs can collect imagery at very fine spatial resolution and can be deployed when specific fields require inspection. RGB, multispectral, hyperspectral, and thermal cameras mounted on UAVs can capture information about plant structure, chlorophyll activity, water stress, disease symptoms, and temperature variation. AI-based image-processing systems can then identify individual plants, classify crop conditions, and detect abnormal areas. UAV imagery therefore provides a useful balance between high spatial resolution and relatively flexible field-level deployment (Pérez-Ortiz et al., 2024).

5.1.3 Detection of Crop Diseases, Pests, and Stress

One of the most significant applications of AI in crop monitoring is the early identification of diseases and pests. Conventional disease identification depends heavily on farmer experience and physical inspection, which can be time-consuming and difficult over large farms. AI-based computer vision can analyze leaf, stem, fruit, and canopy images and identify visual symptoms associated with different diseases.

CNN-based models can detect characteristics such as discoloration, lesions, fungal spots, abnormal growth, and insect damage. AI systems can potentially detect symptoms at an early stage, allowing farmers to take corrective action before the problem spreads throughout a field. A recent systematic review found CNNs to be among the dominant approaches for AI-based crop disease detection, while satellite and drone-based sensing are increasingly being combined with machine learning for monitoring both biotic and abiotic stresses (2025).

AI can also support the detection of abiotic stresses, including drought, salinity, nutrient deficiency, excessive temperature, and water stress. Deep learning models can integrate RGB, thermal, spectral, UAV, satellite, and IoT sensor data to identify and quantify stress conditions. Such systems can help farmers distinguish between healthy and stressed crops and prioritize areas requiring intervention (2025).

5.1.4 AI for Irrigation, Fertilization, and Resource Optimization

Crop monitoring becomes more valuable when monitoring information is connected to farm management decisions. AI can combine crop images with soil moisture, weather, temperature, humidity, and historical yield information to recommend appropriate irrigation and fertilization strategies. For example, a field can be divided into management zones according to crop vigor and soil moisture. Areas showing signs of water stress may receive additional irrigation, while healthy areas can receive less water.

This approach can reduce unnecessary use of agricultural inputs and contribute to environmental sustainability. FAO notes that agricultural automation and digital technologies can improve resource-use efficiency, productivity, resilience, and environmental sustainability, although unequal access to technology remains an important concern (FAO, 2022).

5.1.5 Yield Prediction and Decision Support

Another important application is crop yield prediction. AI models can integrate historical yield records, weather conditions, soil properties, vegetation indices, crop growth characteristics, and satellite imagery to estimate expected production before harvesting. Early yield estimates can assist farmers, agricultural cooperatives, food processors, traders, and policymakers in planning storage, transportation, marketing, and food-security strategies.

Recent research indicates that the integration of AI with remote sensing is moving toward field-level yield prediction, multimodal data fusion, and integration with crop-growth models. Combining different information sources can improve the reliability of predictions under changing environmental conditions (Demissie et al., 2026).

5.1.6 Challenges and Future Directions

Despite its potential, AI-based precision agriculture faces several limitations. High-quality labelled datasets are often difficult to obtain, and models trained under one climatic, geographic, or crop condition may perform poorly in another. Recent systematic reviews emphasize problems related to dataset diversity, external validation, model transferability, computational requirements, and the cost of field deployment (2026).

Smallholder farmers may also face barriers related to the cost of drones, sensors, connectivity, computing infrastructure, and technical expertise. Consequently, the future of AI in agriculture should focus not only on improving model accuracy but also on developing affordable, explainable, scalable, and locally adaptable systems. Multimodal AI, edge computing, self-supervised learning, foundation models, and integration of satellite, UAV, IoT, and ground-based observations are likely to become increasingly important.

Overall, AI is changing crop monitoring from periodic observation to continuous, predictive, and site-specific management. Its integration with remote sensing and precision agriculture can improve early stress detection, resource efficiency, yield forecasting, and agricultural sustainability. However, successful adoption will depend on reliable datasets, robust validation, farmer accessibility, appropriate infrastructure, and responsible implementation. The long-term objective should therefore be to develop AI systems that complement farmers' knowledge and decision-making rather than simply replacing human judgment.

5.2 INTELLIGENT SYSTEMS FOR SOIL AND WATER MANAGEMENT

Soil and water are two of the most critical resources determining agricultural productivity, ecosystem stability, and food security. Increasing water scarcity, soil degradation, irregular rainfall, groundwater depletion, and climate variability have created a need for more precise and efficient resource-management approaches. Traditional soil and irrigation management often depends on periodic field observations, fixed irrigation schedules, and generalized recommendations. Although these methods remain useful, they may not adequately capture the spatial and temporal variability of soil and crop conditions. Artificial intelligence (AI), machine learning (ML), Internet of Things (IoT), and remote sensing are therefore emerging as important technologies for developing intelligent soil and water management systems. FAO emphasizes that improving agricultural water management is essential for food security and sustainable resource use, particularly because agriculture accounts for the largest share of global freshwater withdrawals (FAO, 2026). (Figure 5.2)

5.2 Intelligent Systems for Soil and Water Management

Smart Data. Informed Decisions. Sustainable Future.

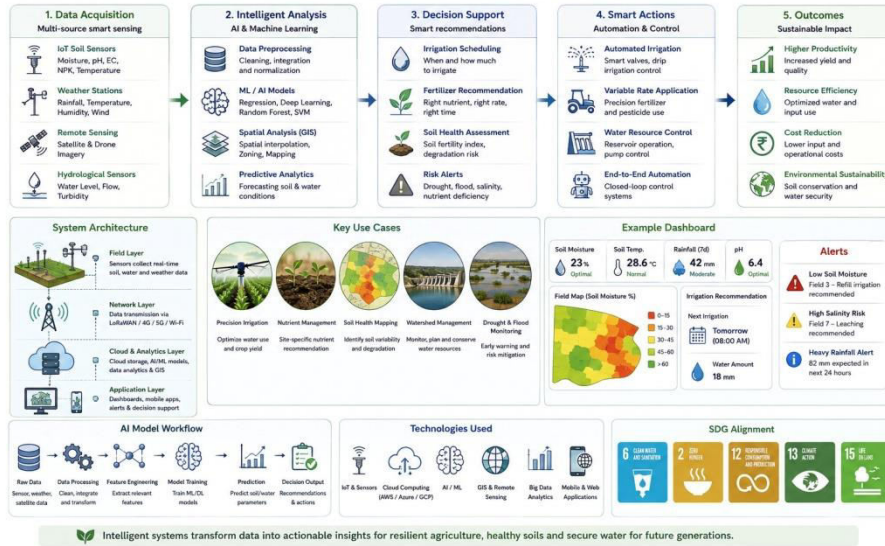


Figure 5.2: Intelligent Systems for Soil and Water Management

5.2.1 AI-Based Soil Monitoring and Assessment

Intelligent soil-management systems combine data from soil sensors, laboratory measurements, weather stations, satellite imagery, drones, and geographic information systems (GIS). AI and ML algorithms can analyse these heterogeneous datasets to identify relationships between soil properties, environmental conditions, crop performance, and management practices. Important soil parameters that can be monitored or estimated include soil moisture, temperature, pH, electrical conductivity, organic carbon, salinity, nutrient availability, and soil structure.

Soil-moisture monitoring is particularly important because it directly influences irrigation decisions and plant growth. Machine-learning models can estimate soil moisture from meteorological observations, soil characteristics, vegetation indices, topography, and remote-sensing information. Recent research shows that combining remote sensing with ML can improve the spatial monitoring of soil moisture and support more efficient agricultural water management (Ben Abbes et al., 2024). Instead of measuring moisture only at a limited number of physical sampling points, AI-supported systems

can generate spatially distributed estimates, helping farmers identify dry and adequately hydrated zones within the same field.

AI can also contribute to digital soil mapping. By integrating historical soil samples with satellite data, elevation, rainfall, temperature, land-use information, and other environmental variables, ML algorithms can predict soil characteristics over areas where direct sampling is unavailable. Such approaches can reduce the cost and time associated with extensive field sampling while providing information useful for site-specific fertilization, irrigation, and land-management decisions.

5.2.2 Intelligent Irrigation and Water Management

One of the most significant applications of AI in agriculture is intelligent irrigation. Conventional irrigation often follows fixed schedules, which can result in over-irrigation during wet periods or water stress during dry periods. Intelligent irrigation systems instead use real-time and predictive information to determine when, where, and how much water should be applied.

A typical AI-enabled irrigation system may integrate soil-moisture sensors, weather forecasts, crop-growth information, evapotranspiration estimates, satellite imagery, and irrigation-system data. ML models analyse these inputs and estimate future crop-water requirements. The irrigation controller can then automatically activate pumps, valves, or drip systems according to the predicted requirement. A systematic review of smart irrigation research found that ML techniques have demonstrated considerable potential for improving irrigation decision-making, although issues concerning generalizability and field-level implementation remain (Abioye et al., 2024).

AI can also support irrigation scheduling by predicting soil-moisture depletion and crop water demand. More advanced systems use reinforcement learning, in which an algorithm learns irrigation strategies from environmental feedback and attempts to balance crop requirements with water conservation. Such approaches can consider both present soil conditions and expected future moisture losses rather than responding only to current sensor readings.

5.2.3 Integration of Remote Sensing, IoT, and AI

The effectiveness of intelligent soil and water management increases when AI is combined with IoT and remote sensing. IoT networks allow sensors

distributed across agricultural fields to continuously collect information about soil moisture, temperature, humidity, rainfall, water flow, and other variables. This information can be transmitted to cloud-based or edge-computing platforms where AI models analyse the data and generate recommendations.

Satellite and drone-based remote sensing provide another important information layer. Multispectral and hyperspectral imagery can help identify variations in vegetation health, soil conditions, moisture levels, and water stress. When these observations are combined with ground-based measurements, AI models can create detailed maps showing areas requiring irrigation or further investigation. A recent review of digital agricultural water-management technologies identified AI and remote sensing among the most versatile technologies for improving irrigation efficiency, identifying water-related problems, and supporting resource management (Parra-López et al., 2025).

5.2.4 Environmental Benefits and Sustainability

Intelligent soil and water management can produce benefits beyond increasing agricultural productivity. More precise irrigation can reduce unnecessary water extraction, minimize runoff, and decrease the risk of waterlogging and salinity. Better soil monitoring can support conservation practices, improve nutrient-use efficiency, and help maintain soil organic matter. These outcomes contribute to the broader objectives of sustainable and climate-resilient agriculture.

FAO notes that modern irrigation management increasingly requires technical and managerial improvements that enhance water, labour, economic, and environmental resource utilization. Decision-support tools can help determine appropriate irrigation timing, water allocation, and crop-water requirements (FAO, 2026). In this context, AI should be viewed not as a replacement for agronomic knowledge but as a decision-support technology that strengthens farmers' ability to interpret complex information.

5.2.5 Challenges and Future Directions

Despite its potential, intelligent soil and water management faces several challenges. AI models require reliable, representative, and sufficiently large

datasets. Sensor failures, missing observations, differences in soil types, regional climatic variations, and limited field data can reduce model reliability. The cost of sensors, drones, connectivity, computing infrastructure, and technical maintenance can also restrict adoption, particularly among smallholder farmers. Recent research highlights data availability, model complexity, infrastructure limitations, and the need for more generalizable models as continuing challenges in AI-based agricultural applications (Ben Abbes et al., 2024; Abioye et al., 2024).

Future intelligent systems are likely to combine AI with physical soil-water models, satellite observations, IoT sensors, digital twins, and real-time weather forecasting. Hybrid models that integrate physical knowledge with data-driven learning may improve reliability and interpretability compared with purely data-driven approaches. Ultimately, the goal is to develop affordable, transparent, and locally adaptable systems capable of providing farmers with actionable recommendations while conserving soil and water resources.

5.3 AI-BASED PEST AND DISEASE DETECTION

Pests and diseases are among the major biological constraints affecting agricultural productivity, crop quality, and food security. Conventional identification generally depends on visual inspection by farmers, agricultural extension workers, or plant pathologists. Although expert diagnosis can be highly reliable, it is often time-consuming, subjective, and difficult to scale across large agricultural areas. Smallholder farmers may also lack timely access to specialists, particularly in remote regions. Artificial intelligence (AI), especially machine learning, deep learning, and computer vision, is increasingly being applied to automate the detection and identification of crop pests and diseases from images and other environmental data (Mohanty et al., 2016; Pacal et al., 2024). (Figure 5.3)

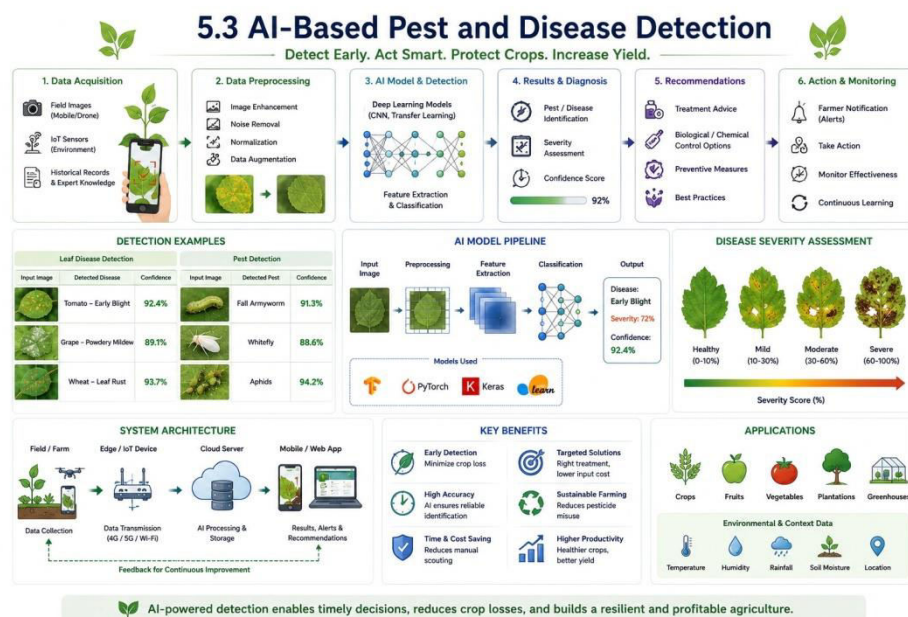


Figure 5.3: AI-Based Pest and Disease Detection

5.3.1 Role of AI in Pest and Disease Detection

AI-based detection systems generally operate by analyzing visual or sensor-based characteristics associated with crop stress. Images of leaves, stems, fruits, flowers, or insects can be captured using smartphones, digital cameras, drones, or field-based imaging systems. These images are processed by AI algorithms to identify symptoms such as discoloration, lesions, spots, wilting, deformation, holes, fungal growth, or the physical presence of insects. The system can subsequently classify the crop condition as healthy or affected by a particular pest or disease.

The development of deep learning, particularly convolutional neural networks (CNNs), has significantly improved automated plant disease recognition. CNNs can automatically learn relevant visual features from images instead of relying entirely on manually designed features. A landmark study by Mohanty et al. (2016) trained a deep convolutional neural network using 54,306 images representing 14 crop species and 26 diseases or healthy conditions. The model achieved 99.35% accuracy on a controlled test dataset, demonstrating the potential of deep learning for automated plant disease diagnosis. However, its

performance dropped substantially when tested on images collected under different real-world conditions, highlighting the importance of diverse and representative training data.

5.3.2 AI Techniques Used in Detection

Several AI approaches are currently used for agricultural pest and disease detection. Image classification models determine whether an image belongs to a particular disease category. Object-detection models, including YOLO-based architectures, can identify and localize multiple pests or disease symptoms within an image. Image segmentation techniques go one step further by identifying the precise region of an affected leaf or fruit. Recent systematic reviews show that CNNs, transfer learning, object detection, segmentation, and transformer-based approaches are among the major techniques being investigated for plant disease recognition (Pacal et al., 2024).

Transfer learning is particularly useful when agricultural datasets are relatively small. In this approach, a model previously trained on a large image dataset is adapted to recognize crop-specific symptoms. This reduces the amount of training data and computational resources required. Lightweight architectures can also be deployed on smartphones and edge devices, allowing farmers to receive diagnostic assistance directly in the field.

5.3.3 Data Sources and Detection Workflow

An AI-based pest and disease detection system generally follows a sequence of steps:

Image/Data Acquisition → pre-processing → Feature Learning → Classification/Detection → Diagnosis → Decision Support

During data acquisition, images may be collected under laboratory, greenhouse, or field conditions. Pre-processing can include resizing, noise reduction, normalization, background removal, and image augmentation. The processed images are then supplied to a trained AI model. Depending on the application, the model may classify the disease, identify an insect, estimate disease severity, or locate affected regions.

The quality and diversity of datasets are critical. Images collected only under controlled backgrounds and lighting conditions may produce excellent laboratory results but fail when exposed to complex field environments. Variations in sunlight, shadows, leaf orientation, soil background, crop varieties, camera quality, and disease severity can substantially affect model performance. Recent reviews emphasize dataset quality, generalization, and realistic field validation as continuing challenges in AI-based plant disease detection (Ngugi et al., 2024; Bagga, 2024).

5.3.4 Integration with Precision Agriculture

AI-based detection becomes more valuable when integrated with precision agriculture technologies. A farmer can use a smartphone to photograph an affected crop, while the AI system identifies a probable pest or disease and provides a confidence score. Drone-based systems can survey large fields and identify areas showing abnormal vegetation patterns. Satellite imagery and hyperspectral sensors can provide additional information about crop stress that may not be visible to the human eye.

AI can also combine image-based observations with weather, soil, crop-growth, and historical disease data. Such multimodal systems can move beyond simple diagnosis toward disease-risk prediction. For example, temperature, humidity, rainfall, and leaf-wetness conditions can be analyzed together with visual symptoms to estimate the likelihood of disease development. This can support earlier intervention and more targeted crop-management decisions.

5.3.5 Benefits for Sustainable Agriculture

The application of AI to pest and disease detection offers several potential benefits. Early identification can allow farmers to isolate affected plants or apply treatment before an outbreak spreads. Accurate detection can also reduce unnecessary pesticide application by enabling targeted interventions rather than treating entire fields indiscriminately. This may reduce input costs, environmental contamination, and risks associated with excessive chemical use.

AI can additionally improve agricultural monitoring by providing rapid and standardized assessments over large areas. For agricultural researchers and

extension agencies, digital disease records can contribute to surveillance systems and help identify emerging patterns. In this sense, AI-based detection is not simply a replacement for human diagnosis; it can function as a decision-support technology that extends the reach of agricultural expertise.

5.3.6 Challenges and Future Directions

Despite considerable progress, several challenges remain. Models trained on limited datasets may perform poorly on unfamiliar crops, disease stages, geographical regions, or environmental conditions. Some diseases also produce visually similar symptoms, making accurate classification difficult. Furthermore, AI predictions can be difficult for farmers to interpret when models operate as “black boxes.” Explainable AI (XAI) techniques are therefore becoming increasingly important because they can indicate which portions of an image influenced a model's prediction.

Future systems are likely to combine computer vision with multimodal AI, edge computing, drones, Internet of Things (IoT) sensors, and geospatial data. Vision transformers and other advanced architectures are also expanding the range of possible detection methods. However, technological accuracy must be accompanied by field validation, affordability, multilingual interfaces, farmer training, and reliable agricultural recommendations. The ultimate objective should be an accessible AI-assisted crop protection ecosystem that enables earlier diagnosis, more precise intervention, reduced resource use, and improved agricultural sustainability.

5.4 ENVIRONMENTAL MONITORING AND CLIMATE PREDICTION

Environmental monitoring and climate prediction have become increasingly important as climate change, extreme weather events, biodiversity loss, water scarcity, and land degradation create complex challenges for agriculture and ecosystem management. Traditional environmental monitoring depends heavily on field observations, weather stations, laboratory measurements, and periodic surveys. Although these approaches provide valuable information, they can be expensive, spatially limited, and difficult to scale. Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), is transforming this landscape by enabling the rapid analysis of large and heterogeneous environmental datasets generated by satellites, drones, Internet

of Things (IoT) sensors, weather stations, and geographic information systems (GIS). AI-based approaches can identify patterns, detect anomalies, estimate environmental variables, and generate predictions that support timely decision-making (Tuia et al., 2023; Han et al., 2024). (Figure 5.4)

5.4 Environmental Monitoring and Climate Prediction



Figure 5.4: Environmental Monitoring and Climate Prediction

5.4.1 AI for Environmental Monitoring

Environmental monitoring involves continuous or periodic observation of variables such as temperature, precipitation, soil moisture, vegetation health, air quality, water quality, land use, forest cover, and ecosystem conditions. AI improves this process by combining information from multiple sources and automatically identifying meaningful patterns.

Remote sensing is one of the most important data sources for AI-enabled environmental monitoring. Satellite imagery provides repeated observations over large geographical areas, while unmanned aerial vehicles (UAVs) can provide high-resolution information at the field or local scale. Machine learning algorithms can process multispectral, hyperspectral, thermal, and radar imagery to classify land cover, identify vegetation stress, monitor forest

changes, and estimate soil and water characteristics. The integration of remote sensing and machine learning has therefore become an important component of precision agriculture and environmental assessment (Han et al., 2024; Wang et al., 2024).

AI is particularly useful for monitoring vegetation and crop conditions. Algorithms such as Random Forest, Support Vector Machines, convolutional neural networks (CNNs), and other deep-learning architectures can identify changes in vegetation indices and imagery associated with drought stress, disease, nutrient deficiencies, or changes in crop growth. This enables agricultural managers to detect problems earlier and apply water, fertilizers, or other interventions more efficiently. Recent research indicates that CNN-based approaches have become particularly prominent in automated crop monitoring, although limitations related to annotated datasets and model generalization remain important challenges (Jabed et al., 2024).

Soil and water monitoring also benefit from AI. Soil moisture is a critical environmental variable because it affects crop growth, irrigation requirements, hydrological processes, and drought assessment. Machine learning models can estimate soil moisture from satellite observations and other environmental variables, potentially providing more detailed spatial information than conventional measurements alone. Recent research highlights the value of combining physical models with machine learning to improve the accuracy, scalability, and robustness of remote-sensing-based soil-moisture estimation (Ben Abbes et al., 2024).

AI can similarly support water-quality monitoring by analysing measurements such as temperature, turbidity, dissolved oxygen, nutrient concentrations, and other indicators. Predictive models can identify abnormal conditions and provide early warnings of pollution or ecological deterioration. When integrated with sensor networks, AI can support near-real-time environmental surveillance rather than relying exclusively on periodic manual sampling.

5.4.2 AI for Climate Prediction

Climate prediction involves estimating future atmospheric and environmental conditions using historical observations, physical relationships, and numerical models. Conventional climate models are based primarily on physical

equations describing the atmosphere, oceans, land surface, and cryosphere. AI does not necessarily replace these models; instead, it can complement them by learning complex relationships from large datasets and improving computational efficiency.

Machine learning can be applied to predict variables such as temperature, precipitation, drought probability, extreme rainfall, soil moisture, and agricultural climate risks. Time-series models can analyse historical weather observations to identify trends and recurring patterns, while deep-learning models can process high-dimensional spatial and temporal datasets. In agroclimatic research, ML and DL are increasingly being investigated for understanding climate impacts and developing adaptation strategies (Hassan et al., 2024).

One important application is extreme-event prediction. Droughts, floods, heatwaves, storms, and excessive rainfall can severely affect agricultural productivity and ecosystems. AI systems can analyse historical weather records, satellite observations, soil characteristics, topography, and atmospheric variables to estimate the probability of such events. Early-warning systems based on these predictions can help farmers and environmental authorities take preventive measures.

AI can also contribute to local-scale weather and microclimate prediction. Conventional weather forecasts may not capture environmental differences between individual farms or small geographical areas. Combining IoT sensors with AI models can provide high-resolution information about temperature, humidity, rainfall, and other local conditions. Such systems can support decisions concerning irrigation, planting, harvesting, and protection against weather-related risks.

5.4.3 Applications in Agriculture and Environmental Management

The combination of environmental monitoring and climate prediction creates several practical applications. Farmers can use AI-generated information to determine irrigation schedules, identify drought-prone areas, select appropriate planting periods, and anticipate climate-related crop losses. Environmental agencies can use AI to monitor deforestation, land degradation, ecosystem changes, and water resources. AI-supported soil-carbon mapping can also

reduce dependence on extensive physical sampling by using environmental and remote-sensing variables to estimate soil organic carbon across larger areas.

Another important application is climate-resilient agriculture. Crop-yield prediction models can integrate historical yields, weather conditions, soil characteristics, satellite imagery, and management information to estimate expected production. Such predictions can assist farmers, governments, insurers, and food-supply organizations in preparing for climate variability. Recent reviews demonstrate the growing use of ML, DL, ensemble learning, and explainable AI for crop-yield prediction under abnormal climatic conditions (Jabed et al., 2024).

5.4.4 Challenges and Future Directions

Despite its potential, AI-based environmental monitoring faces several challenges. Environmental datasets may contain missing observations, measurement errors, inconsistent spatial resolutions, and geographical biases. Models trained in one region may perform poorly in another because climate, soil, vegetation, and agricultural practices differ. The lack of high-quality labelled datasets and difficulties in combining satellite, sensor, meteorological, and field data can also limit model performance.

Interpretability is another important concern. Environmental decisions can have significant economic and ecological consequences, making it necessary to understand why an AI system produces a particular prediction. Explainable AI, physics-informed machine learning, and hybrid models that combine physical knowledge with data-driven algorithms are therefore likely to become increasingly important.

Overall, AI is emerging as a powerful component of environmental monitoring and climate prediction. Its greatest value lies not simply in generating predictions but in transforming large volumes of environmental data into actionable information. When combined with remote sensing, IoT, GIS, physical climate models, and expert knowledge, AI can support more accurate monitoring, earlier warnings, efficient resource management, and climate-resilient agriculture. Future systems are likely to emphasize multimodal data integration, explainability, physics-aware learning, edge computing, and real-

time decision support, making AI an increasingly important technology for sustainable management of agricultural and environmental life sciences.

CHAPTER 6

AI in Biotechnology and Industrial Life Sciences

6.1 AI APPLICATIONS IN BIOTECHNOLOGY AND BIOENGINEERING

Artificial Intelligence (AI) has emerged as an important enabling technology in biotechnology and bioengineering, where biological systems generate complex, high-dimensional, and often heterogeneous datasets. Conventional biological experimentation is frequently time-consuming, expensive, and dependent on repeated laboratory trials. AI and machine learning (ML) provide computational approaches for identifying patterns in genomic, proteomic, metabolomic, imaging, and bioprocess datasets and using these patterns to support prediction, design, optimization, and decision-making. Consequently, AI is increasingly being integrated into areas such as protein engineering, metabolic engineering, synthetic biology, drug discovery, fermentation, bioprocess optimization, and biomaterial development (Dunn et al., 2020; Peng et al., 2026). (Figure 6.1)

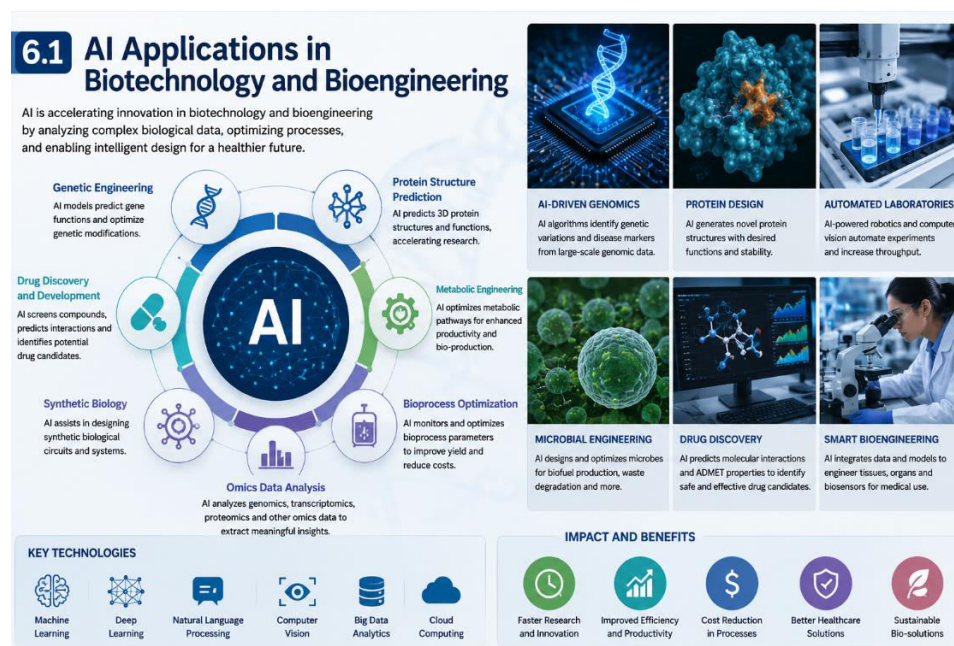


Figure 6.1: AI Applications in Biotechnology and Bioengineering

One of the most significant applications of AI in biotechnology is protein structure prediction and protein engineering. Proteins perform essential biological functions, and understanding their three-dimensional structures is

critical for designing enzymes, antibodies, therapeutics, and industrial catalysts.

Traditional experimental techniques for determining protein structures can require substantial time and resources. Deep-learning systems such as AlphaFold demonstrated that AI can predict protein structures with remarkable accuracy from amino acid sequences, substantially improving computational structural biology (Jumper et al., 2021). More recent machine-learning approaches have expanded from structure prediction toward functional protein design, allowing researchers to explore protein sequences and identify candidate variants with desirable characteristics. Such approaches can support the development of enzymes with improved stability, catalytic activity, specificity, or resistance to harsh industrial conditions (Notin et al., 2024).

AI is also transforming metabolic engineering and synthetic biology. Metabolic engineering involves modifying microorganisms or cells so that they produce valuable compounds such as biofuels, pharmaceuticals, amino acids, organic acids, enzymes, and other industrial chemicals. Biological pathways can contain numerous interacting genes, enzymes, metabolites, and regulatory mechanisms, making it difficult to predict the consequences of individual genetic modifications. Machine-learning models can analyze omics data and experimental results to identify relationships between genetic modifications and production characteristics. These models can subsequently guide the selection of metabolic pathways, predict promising gene combinations, and prioritize experiments. Machine learning has therefore been applied to pathway construction, pathway optimization, strain engineering, and scale-up activities (Dunn et al., 2020).

Another important application is enzyme engineering. Enzymes are widely used in food processing, detergents, pharmaceuticals, agriculture, biofuels, and chemical manufacturing. Conventional enzyme engineering often relies on directed evolution, in which large numbers of variants are experimentally generated and screened. AI can reduce the number of experimental candidates by learning relationships between amino acid sequences, structures, and functional properties. Protein language models and generative AI can identify potentially useful sequences, while predictive models can estimate properties such as enzyme activity, thermostability, and substrate specificity before

laboratory testing. Recent research indicates that AI-assisted protein design is moving toward integrated models that consider protein sequences, structures, metabolic pathways, and cellular phenotypes simultaneously (Notin et al., 2024).

In bioprocess engineering, AI can improve the production of biological products at laboratory, pilot, and industrial scales. Fermentation and cell-culture processes depend on numerous parameters, including temperature, pH, dissolved oxygen, nutrient concentration, agitation speed, feed rate, and biomass concentration. Small changes in these variables can significantly influence yield and product quality. Machine-learning models can learn from historical process data and predict production outcomes under different operating conditions. AI can consequently support process optimization, anomaly detection, predictive maintenance, and real-time control. This is particularly valuable in biopharmaceutical manufacturing, where upstream processes involving living cells can generate complex datasets while experimental data may remain relatively limited (Peng et al., 2026).

AI additionally supports automated experimentation and closed-loop biotechnology. In conventional research, scientists design an experiment, perform laboratory work, analyze the results, and then design the next experiment. AI can partially automate this cycle by selecting experiments that are expected to provide the greatest amount of useful information. When combined with laboratory robotics, automated instruments, and high-throughput screening, AI systems can iteratively generate hypotheses, conduct experiments, analyze results, and recommend subsequent experiments. This approach has the potential to reduce experimental workload and accelerate optimization of biological systems (Duong-Trung et al., 2022).

Drug discovery and biopharmaceutical development represent another major area in which biotechnology and AI intersect. AI can analyze molecular structures, biological targets, genomic information, and experimental data to support target identification, virtual screening, molecular property prediction, and candidate optimization. Generative models can also propose new molecular structures with selected characteristics. In protein and antibody engineering, machine-learning systems can help identify candidate binders and optimize molecular properties. These capabilities can reduce the number of

compounds that must be experimentally evaluated and help researchers prioritize promising candidates.

Nevertheless, AI-generated candidates still require laboratory validation, preclinical studies, and clinical evaluation because computational predictions cannot completely reproduce the complexity of biological systems.

Beyond pharmaceuticals, AI has applications in industrial biotechnology, including biofuel production, sustainable chemical manufacturing, food biotechnology, agricultural biotechnology, and biomaterial development. AI can help identify microorganisms capable of producing useful compounds, optimize fermentation conditions, predict material properties, and improve biological production efficiency. In energy biotechnology, for example, AI-assisted enzyme and metabolic engineering can contribute to the development of microbial systems capable of producing fuels and other valuable products more efficiently (Dunn et al., 2020; 2026).

Despite these opportunities, AI adoption in biotechnology faces several challenges. Biological datasets can be incomplete, noisy, biased, or difficult to standardize. Many biotechnology experiments are expensive, resulting in small datasets that can limit model performance. Furthermore, biological systems are highly context-dependent, meaning that a model trained under one experimental condition may not perform equally well under another. Interpretability, reproducibility, data privacy, computational requirements, and experimental validation also remain important concerns. Therefore, AI should generally be viewed as a decision-support and discovery technology rather than a complete replacement for biological experimentation.

Overall, AI is creating a shift from predominantly trial-and-error biotechnology toward data-driven, predictive, and increasingly automated biological engineering. Its greatest value lies in combining computational prediction with laboratory experimentation. As protein-language models, generative AI, automated laboratories, multi-omics integration, and digital bioprocessing continue to develop, AI is likely to become increasingly important in designing biological systems, optimizing industrial processes, and accelerating the development of biotechnology products.

6.2 SYNTHETIC BIOLOGY AND INTELLIGENT BIOPROCESSING

Synthetic biology and artificial intelligence (AI) are increasingly converging to transform the design, engineering, and production of biological systems. Synthetic biology applies engineering principles to biological components, enabling researchers to construct or redesign cells, genetic circuits, metabolic pathways, and enzymes for specific purposes. AI and machine learning (ML), in contrast, provide computational methods for identifying patterns in large biological datasets, predicting system behaviour, and selecting promising designs. Their integration is creating a more data-driven approach to biotechnology in which biological systems can be designed, tested, analysed, and optimized through iterative computational and experimental cycles (Goshisht, 2024). (Figure 6.2)

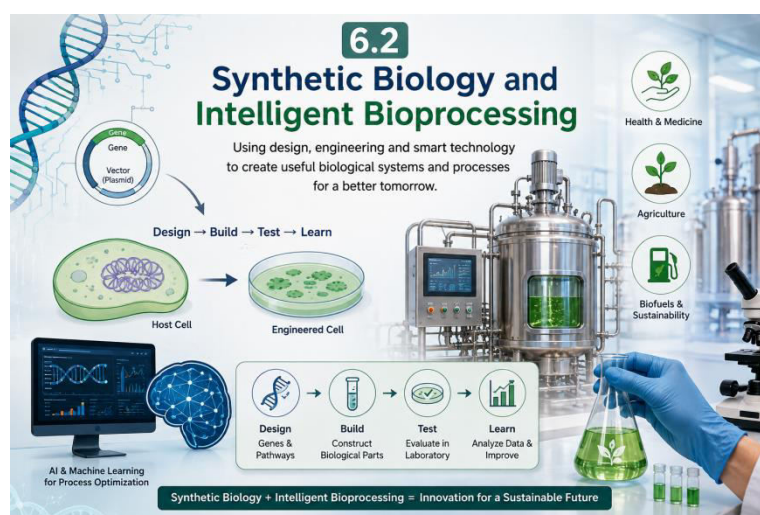


Figure 6.2: Synthetic Biology and Intelligent Bioprocessing

6.2.1 AI-Enabled Synthetic Biology

A central framework in synthetic biology is the Design–Build–Test–Learn (DBTL) cycle. In the design stage, researchers identify biological components or pathways that may achieve a desired function. The build stage involves constructing the selected genetic or cellular system, while the test stage evaluates its performance experimentally. Finally, the learn stage uses experimental data to identify successful designs and guide the next cycle. AI

can strengthen each stage, particularly by helping researchers predict which designs are most likely to work before laboratory resources are committed.

Machine learning can analyse genomic, transcriptomic, proteomic, metabolomic, and experimental datasets to identify relationships between genotype, phenotype, and environmental conditions. For example, ML models can predict gene expression, enzyme activity, protein function, pathway performance, and cellular responses. Such predictions can reduce the number of experimental combinations that need to be tested. In protein engineering, recent advances in machine learning have enabled computational design and optimization of proteins with desired structural or functional properties, expanding the search space beyond what can practically be explored through conventional laboratory evolution (Notin et al., 2024).

AI is also being used for the design of genetic regulatory elements and metabolic pathways. Instead of relying exclusively on trial-and-error experimentation, researchers can use predictive models to estimate how changes in promoters, enzymes, genes, or pathway structures may influence cellular performance. This is particularly valuable because biological systems contain complex interactions and nonlinear relationships that are difficult to model using conventional approaches alone. Resource competition and cellular burden, for instance, can cause engineered genetic systems to behave differently from their theoretical designs, making quantitative and data-driven approaches increasingly important (Di Blasi et al., 2024).

6.2.2 Intelligent Bioprocessing

Synthetic biology produces engineered biological systems, but successful industrial biotechnology also requires efficient processes for growing cells and converting biological substrates into useful products. This is the domain of intelligent bioprocessing, where AI, sensors, automation, and advanced process models are combined to monitor and optimize biological production.

Industrial bioprocesses such as fermentation and cell culture involve numerous interacting variables, including temperature, pH, dissolved oxygen, nutrient concentration, agitation, aeration, biomass concentration, metabolite levels, and feeding rates. These variables can change dynamically during production. Traditional control systems may struggle to capture all these nonlinear

interactions. ML models can instead learn relationships between process variables and important outcomes such as product yield, productivity, biomass growth, and product quality (Mondal et al., 2023).

One important application is the development of soft sensors. Many important biological variables cannot be measured continuously because direct measurement is expensive, technically difficult, or requires laboratory sampling. AI models can estimate such variables indirectly by combining available sensor measurements. This allows operators to obtain near-real-time information about the biological state of a process and make faster control decisions. AI can also support fault detection, process monitoring, parameter optimization, and predictive control.

6.2.3 AI for Metabolic Pathway and Strain Optimization

An important connection between synthetic biology and intelligent bioprocessing is the development of microbial cell factories. Microorganisms such as *Escherichia coli*, yeast, and other industrial hosts can be genetically engineered to produce pharmaceuticals, enzymes, biofuels, organic acids, specialty chemicals, and other valuable products.

AI can assist in identifying metabolic bottlenecks and predicting how modifications to genes or enzymes may affect product formation. ML-based pathway optimization can support genome-scale metabolic modelling, enzyme engineering, regulatory-element design, and selection of promising pathway configurations. Data-driven approaches are particularly useful when experimental datasets are sufficiently large to capture complex biological relationships (Jin et al., 2023).

Another emerging approach is autonomous metabolic regulation, in which engineered cells dynamically adjust metabolic activity according to environmental or intracellular signals. Such systems can redirect cellular resources between growth and product formation, potentially improving production efficiency while reducing metabolic burden (Ream & Prather, 2024).

6.2.3 Digital Twins and Autonomous Bioprocesses

The next stage of intelligent bioprocessing involves digital twins-virtual representations of biological production systems that combine process data,

mathematical models, sensors, and AI. A digital twin can continuously receive information from a physical bioreactor and use computational models to predict future process behaviour. The system can then support optimization or, in advanced applications, automated process control.

Recent research indicates that hybrid digital-twin approaches combining mechanistic models with ML can provide advantages over relying exclusively on either approach. Mechanistic models incorporate known biological and engineering principles, whereas ML models can capture complex patterns that are difficult to describe mathematically. Their combination can therefore improve monitoring, optimization, and control in bioprocessing (On digital twins in bioprocessing, 2025).

6.2.4 Challenges and Future Perspectives

Despite its potential, AI-enabled synthetic biology faces several challenges. High-quality biological datasets are often limited, heterogeneous, and difficult to standardize. Models trained on one organism, laboratory, or process may not generalize effectively to another. Interpretability is another concern because highly complex models may produce accurate predictions without providing clear biological explanations. In industrial environments, scalability, sensor integration, cybersecurity, regulatory compliance, and validation are also important considerations.

Future systems are likely to combine AI, synthetic biology, robotics, high-throughput experimentation, multi-omics, IoT sensors, and digital twins into integrated autonomous platforms. Such systems could continuously design biological variants, conduct experiments, analyse results, and update predictive models. This would move biotechnology from predominantly human-guided experimentation toward increasingly automated and adaptive biological engineering. AI-enabled programming of biology is already extending from nucleic acids and proteins toward increasingly complex cellular systems (Abudayyeh & Gootenberg, 2024).

Overall, the integration of synthetic biology and intelligent bioprocessing represents a major transition in biotechnology. Synthetic biology provides the ability to engineer biological systems, while AI provides the computational intelligence required to explore, predict, and optimize those systems. Together,

they can accelerate the development of engineered organisms, improve industrial production efficiency, reduce experimental costs, and support more sustainable biomanufacturing. The long-term objective is not simply to automate individual laboratory tasks, but to establish intelligent biological production systems capable of learning from data and continuously improving their performance.

6.3 AI IN FOOD BIOTECHNOLOGY AND SAFETY ANALYSIS

Artificial intelligence (AI) is increasingly becoming an important enabling technology in food biotechnology, food quality assessment, and food safety management. The food sector generates large and complex datasets from spectroscopy, chromatography, mass spectrometry, biosensors, genomic analysis, imaging systems, production processes, and supply-chain monitoring. Conventional analytical approaches often require substantial laboratory time, specialized equipment, and expert interpretation. AI and machine learning (ML) can complement these approaches by identifying patterns in multidimensional datasets, predicting quality and safety outcomes, and supporting faster decision-making (Yi et al., 2024). Recent research indicates that AI is moving beyond experimental applications toward integrated food-safety systems involving sensing, prediction, traceability, and risk management. (Figure 6.3)



Figure 6.3: AI in Food Biotechnology and Safety Analysis

6.3.1 AI-Based Food Quality and Safety Assessment

One of the major applications of AI in food biotechnology is the rapid assessment of food quality and safety indicators. Machine-learning algorithms can process data generated through near-infrared (NIR), hyperspectral imaging, Raman spectroscopy, nuclear magnetic resonance (NMR), gas chromatography, and mass spectrometry. These analytical techniques produce complex chemical and spectral profiles that may be difficult to interpret using conventional statistical methods alone. AI models, including artificial neural networks, support vector machines, random forests, and deep-learning architectures, can learn relationships between analytical signals and food characteristics such as freshness, moisture, composition, oxidation, contamination, and authenticity (Yi et al., 2024).

Computer vision and deep learning also provide important tools for automated inspection. Cameras and hyperspectral imaging systems can capture information about colour, texture, surface defects, foreign materials, and physical deterioration. Convolutional neural networks (CNNs) can subsequently classify food products or identify defects with limited manual intervention. Such systems are particularly useful in high-throughput processing environments where continuous human inspection can be expensive, inconsistent, and difficult to scale.

6.3.2 Detection of Foodborne Pathogens

Foodborne microorganisms, including *Salmonella*, *Listeria monocytogenes*, *Escherichia coli*, and other pathogens, represent a major food-safety concern. Traditional microbiological testing can require enrichment, culturing, biochemical identification, and laboratory confirmation, which may delay corrective action. AI-assisted detection combines machine learning with biosensors, microscopy, spectroscopy, molecular diagnostics, and other rapid analytical technologies to accelerate pathogen identification.

ML-powered biosensors can process signals generated by biological recognition elements and improve classification of microorganisms and other contaminants.

Recent research has demonstrated applications of ML-enabled biosensing for microorganisms, antibiotics, pesticides, mycotoxins, heavy metals, anions, and

persistent organic pollutants (Hassan et al., 2025). More recent systematic research has expanded this perspective by examining AI-assisted microscopy, spectroscopy, biosensors, and sensor-based systems as components of broader food-safety frameworks (Schirone et al., 2026).

AI can also contribute to foodborne outbreak investigation. Machine-learning models can analyse pathogen genomes, epidemiological information, geographic patterns, food-trade data, inspection records, and other heterogeneous datasets to identify relationships that may not be readily apparent through conventional analysis. Earlier research has highlighted applications of ML in pathogen source attribution, antibiotic-resistance prediction, outbreak detection, and food-safety risk assessment (Larkin et al., 2021).

6.3.3 Detection of Adulteration and Contaminants

Food adulteration is another area in which AI offers considerable potential. Adulteration may involve substitution of expensive ingredients, addition of unauthorized substances, dilution, mislabelling, or contamination. AI can compare the chemical, spectral, visual, or molecular fingerprint of a sample with reference datasets to identify deviations from expected characteristics.

For example, AI-assisted spectroscopy can be used to distinguish authentic and adulterated food products, while metabolomic and mass-spectrometric datasets can be analysed using ML algorithms to identify characteristic molecular patterns. This approach is particularly valuable for complex products such as spices, oils, dairy products, herbal materials, beverages, and processed foods. Reviews of AI and ML applications in food integrity indicate that these technologies can strengthen authenticity verification, quality control, and supply-chain resilience (Gbashi & Njobeh, 2024).

Mycotoxin detection is another promising application. Mycotoxins are toxic compounds produced by certain fungi and may contaminate cereals, nuts, spices, and other agricultural commodities. ML models can analyse spectral or imaging data to predict contamination and support rapid screening before products enter the food chain.

However, the reliability of these systems depends strongly on representative datasets, appropriate model validation, and transparent reporting of model parameters.

6.3.4 Predictive Food Safety and Risk Management

AI is increasingly being applied not only to detect existing contamination but also to predict potential food-safety risks. Predictive models can integrate environmental conditions, temperature, humidity, processing parameters, storage duration, microbial growth characteristics, and historical contamination data. These models can estimate the probability of spoilage or microbial growth and support preventive interventions.

Integration with the Internet of Things (IoT) can further improve this capability. Sensors installed throughout production and storage environments can continuously generate data, while AI models analyse these data in real time. Abnormal temperature patterns, unexpected changes in microbial indicators, or deviations from processing conditions can trigger early warnings. Consequently, food-safety management can become more proactive rather than relying exclusively on end-product testing.

6.3.5 AI in Food Biotechnology and Bioprocess Optimization

Beyond safety analysis, AI has applications in food biotechnology itself. Fermentation processes, enzyme production, microbial cultures, alternative proteins, and cultured-food technologies involve numerous interacting biological and environmental variables. ML can help identify relationships among parameters such as pH, temperature, nutrient composition, oxygen availability, inoculum concentration, and fermentation time.

In emerging cultured-meat biotechnology, for example, ML has been investigated for cell-line development, culture-media optimization, microscopy and image analysis, and bioprocess optimization. Such approaches may reduce the number of experimental iterations required to identify suitable conditions and thereby accelerate research and development.

6.3.6 Challenges and Future Perspectives

Despite its potential, AI should not be viewed as a replacement for laboratory science, food microbiologists, or regulatory expertise. The quality of an AI model is highly dependent on the quality, quantity, diversity, and

representativeness of its training data. Models trained on limited datasets may perform well under laboratory conditions but fail when applied to different food matrices, geographic regions, processing environments, or contamination levels. Issues of data standardization, model interpretability, reproducibility, cybersecurity, privacy, and regulatory acceptance therefore remain important.

Another challenge is the “black-box” nature of some deep-learning systems. In food-safety decisions, stakeholders may need to understand why a system classified a sample as contaminated or unsafe. Explainable AI can therefore become increasingly important for regulatory and industrial applications. Recent research also emphasizes the need for appropriate governance, open and high-quality datasets, validation across real-world conditions, and meaningful human oversight.

Overall, AI is transforming food biotechnology from primarily reactive testing toward intelligent, predictive, and data-driven management. Its integration with biosensors, spectroscopy, genomics, computer vision, IoT, and automated processing systems could enable faster detection of hazards, improved product authenticity, optimized bioprocesses, and more responsive food-safety systems. The future of AI in food biotechnology will therefore depend not only on increasingly sophisticated algorithms but also on reliable data, interdisciplinary collaboration, validation, regulatory frameworks, and responsible human supervision.

6.4 INDUSTRIAL AUTOMATION AND COMPUTATIONAL BIOTECHNOLOGY

Industrial biotechnology is increasingly shifting from manually controlled, experience-driven production toward data-driven, automated, and computationally optimized systems. The integration of artificial intelligence (AI), machine learning (ML), robotics, process sensors, computational biology, and digital twins is creating a new generation of intelligent biomanufacturing systems. These technologies can support the design of biological systems, optimize fermentation and cell culture, improve process control, reduce production variability, and accelerate the development of industrially valuable products. This convergence is often associated with Bioprocessing 4.0, in which industrial biotechnology adopts technologies

similar to those of Industry 4.0, including industrial Internet of Things (IIoT), AI, automation, and digital twins (Isoko et al., 2024). (Figure 6.4)

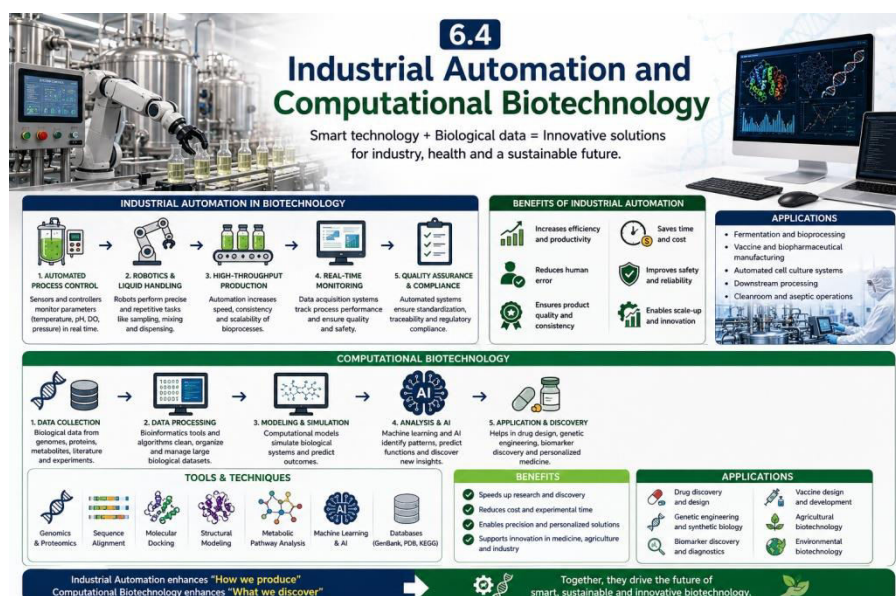


Figure 6.4: Industrial Automation and Computational Biotechnology

6.4.1 Role of AI in Industrial Biotechnology

Industrial biotechnology involves complex biological processes in which small changes in temperature, pH, oxygen concentration, nutrient availability, agitation, or cell physiology can substantially influence productivity. Conventional optimization often depends on repeated experiments and expert knowledge. Machine learning provides an alternative by identifying relationships between process variables and biological outcomes from historical and experimental data. Applications include strain selection, metabolic pathway optimization, fermentation optimization, scale-up, process monitoring, and quality prediction (Lawson et al., 2021).

For example, ML models can analyze fermentation datasets containing variables such as dissolved oxygen, glucose concentration, biomass, temperature, pH, and metabolite levels. Once trained, predictive models can estimate product yield or identify process conditions likely to maximize productivity. This allows researchers to reduce the number of physical

experiments required and focus experimental resources on the most promising conditions.

AI is also becoming important in metabolic and enzyme engineering. Computational models can predict how genetic modifications may influence metabolic pathways and cellular phenotypes. Recent developments include protein language models, generative AI approaches for enzyme design, and integrated models connecting protein function, metabolic pathways, and cellular states (Zhao et al., 2026).

6.4.2 Automation of Bioprocesses

Automation provides the physical infrastructure through which computational predictions can be translated into biological experiments and industrial operations. Automated bioreactors, liquid-handling systems, robotic sampling, online sensors, automated analytical instruments, and process-control systems can continuously collect and respond to process information.

A particularly important development is the emergence of closed-loop bioprocess automation. In such systems, sensors collect real-time information, AI models interpret the data, algorithms determine an appropriate action, and automated equipment implements the decision. The resulting biological response is measured again, creating a continuous feedback loop.

This approach can be extended from laboratory-scale experimentation to industrial fermentation and biopharmaceutical manufacturing. Instead of maintaining fixed process conditions throughout production, intelligent control systems can potentially adjust operating parameters according to the physiological state of the culture. Such adaptive control can improve consistency and reduce deviations from desired process conditions.

6.4.3 Digital Twins in Biomanufacturing

Digital twins represent another important component of computational biotechnology. A digital twin is a computational representation of a physical process or production system that is continuously informed by data from its real-world counterpart. In biopharmaceutical manufacturing, digital twins can be used to model process behavior, simulate operating conditions, predict deviations, and support process optimization. Their potential benefits include

improved productivity, reduced costs, consistent product quality, and support for Quality-by-Design principles (Ding et al., 2024).

For example, a digital twin of a fermentation process can integrate historical process data, real-time sensor readings, mechanistic models, and ML predictions. Engineers can use the virtual model to evaluate the likely consequences of changing agitation speed, feeding rate, temperature, or oxygen availability before implementing the change in the physical bioreactor. Consequently, digital twins can function as decision-support systems and reduce the risks associated with trial-and-error optimization.

6.4.4 Computational Biotechnology and Self-Driving Laboratories

The combination of AI with robotics has led to the development of self-driving laboratories (SDLs). These systems integrate automated laboratory equipment with AI algorithms capable of selecting and prioritizing experiments. Rather than researchers manually designing every experimental iteration, the system can analyze previous results and select subsequent experiments according to a defined objective.

A 2025 study demonstrated an autonomous laboratory for biotechnology that combined robotic equipment with Bayesian optimization to perform a closed experimental loop involving cell cultivation, sample preparation, measurement, analysis, and hypothesis formulation. The system was successfully used to optimize culture conditions for recombinant *Escherichia coli* producing glutamic acid (Fushimi et al., 2025).

Similarly, self-driving laboratory platforms have been applied to protein engineering. The SAMPLE platform used an AI agent to learn relationships between protein sequences and functions, design new protein variants, and send them to an automated robotic system for experimental testing. Experimental results were then returned to the computational model, enabling further optimization (Rapp et al., 2024).

Such systems demonstrate the transition from a traditional design–build–test–learn workflow toward an automated design–build–test–learn–redesign cycle. This can substantially increase experimental throughput while allowing researchers to investigate complex biological design spaces more efficiently.

6.4.5 Applications across Industrial Life Sciences

Industrial automation and computational biotechnology have applications across several sectors. In pharmaceutical biotechnology, AI-assisted process control can support the manufacture of monoclonal antibodies, vaccines, recombinant proteins, and other biologics. In industrial fermentation, computational models can optimize microbial growth and metabolite production. In enzyme biotechnology, AI can identify or redesign enzymes with desirable properties such as improved stability, catalytic activity, or temperature tolerance.

The technology is also relevant to biofuels, food biotechnology, agricultural biotechnology, biomaterials, and sustainable chemical production. ML-based metabolic engineering has already been investigated for pathway construction, genetic-editing optimization, cell-factory development, and production-scale optimization (Lawson et al., 2021).

6.4.6 Challenges and Future Directions

Despite its potential, intelligent automation in biotechnology faces important challenges. Biological systems generate highly variable and context-dependent data, while industrial datasets may be limited, noisy, heterogeneous, or poorly standardized.

ML models trained on insufficient or biased datasets may produce unreliable predictions. The complexity of biological processes also means that purely data-driven models may not always provide adequate mechanistic understanding.

Another challenge is integration and validation. AI models must interact reliably with sensors, robotic systems, laboratory information systems, and industrial control infrastructure. In regulated environments, traceability, data integrity, model validation, cybersecurity, and human oversight are particularly important. The increasing use of self-driving laboratories has therefore created a parallel need for safety frameworks and clearly defined human–machine responsibilities (Leong et al., 2025).

Future industrial biotechnology is likely to move toward hybrid intelligent systems that combine mechanistic models, machine learning, digital twins, robotics, and human expertise. Rather than completely replacing

biotechnologists, these systems are expected to automate repetitive activities and support complex decision-making. The result could be more adaptive, efficient, reproducible, and sustainable biomanufacturing. As AI models become capable of integrating multi-omics, protein, metabolic, process, and environmental data, computational biotechnology may increasingly become a central component of industrial life-science innovation.

CHAPTER 7

Ethical, Legal, and Social Implications of AI in Life Sciences

7.1 DATA PRIVACY AND SECURITY IN AI SYSTEMS

Artificial Intelligence (AI) has become increasingly important in life sciences, supporting applications such as medical diagnosis, drug discovery, genomic analysis, personalized medicine, clinical research, disease surveillance, and biomedical data analysis. These applications depend heavily on large and diverse datasets, including electronic health records, medical images, genomic sequences, laboratory results, wearable-device data, clinical-trial information, and patient-reported outcomes. Although access to such data can accelerate scientific discovery and improve healthcare delivery, it also creates significant challenges related to privacy, confidentiality, cybersecurity, and responsible data governance. The World Health Organization (WHO) emphasizes that AI in health should place ethics, human rights, accountability, and protection of individuals at the center of system design and deployment (WHO, 2021). (Figure 7.1)



Figure 7.1: Data Privacy and Security in AI Systems

7.1.1 Nature of Privacy Risks in Life-Science AI

Health and biological data are particularly sensitive because they can reveal information about an individual's physical condition, genetic characteristics,

reproductive history, lifestyle, disease risks, and family relationships. Unlike conventional identifiers such as names or addresses, certain biological characteristics cannot simply be changed after disclosure. Genetic information, for example, may provide information not only about an individual but also about biologically related family members.

AI introduces additional privacy risks because machine-learning systems can identify patterns that were not necessarily obvious when the data were originally collected. Even when datasets are de-identified, combinations of apparently harmless attributes may sometimes allow individuals to be re-identified. NIST notes that AI systems can create new privacy risks by enabling inference about individuals or previously private information. It therefore recommends privacy-enhancing approaches such as data minimization, de-identification, aggregation, and other privacy-enhancing technologies (NIST, 2023).

Another concern is the secondary use of data. Information initially collected for patient treatment may subsequently be used for research, AI model training, pharmaceutical development, or commercial applications. Such secondary use raises questions about whether individuals understood how their information would be used and whether their consent adequately covers future applications. Effective governance therefore requires clear rules concerning the purpose of data collection, permitted uses, retention periods, access rights, and data-sharing arrangements.

7.1.2 Data Security Threats

Privacy protection and cybersecurity are closely connected but are not identical. Privacy concerns primarily address appropriate collection, use, disclosure, and control of personal information, whereas security focuses on protecting information and systems against unauthorized access, alteration, destruction, or disruption.

AI systems used in life sciences may be exposed to conventional cybersecurity threats such as unauthorized access, phishing, ransomware, insider threats, insecure databases, and compromised cloud infrastructure. They may also face AI-specific attacks. For example, attackers may attempt to manipulate training

datasets, extract sensitive information from models, infer characteristics about individuals, or exploit vulnerabilities in AI-enabled applications.

The consequences of a security breach can be particularly serious in healthcare. Unauthorized access to medical records can result in financial harm, discrimination, reputational damage, or psychological distress. Manipulation of clinical or laboratory data could also affect medical decisions and research conclusions. In the United States, the HIPAA Security Rule requires covered entities and business associates to implement administrative, physical, and technical safeguards to protect the confidentiality, integrity, and availability of electronic protected health information (HHS, 2026).

7.1.3 Privacy-Preserving Approaches

Several technical and organizational measures can reduce privacy and security risks in AI-enabled life-science systems. Data minimization involves collecting and processing only the information necessary for a specific purpose. De-identification and pseudonymization can reduce direct identification risks, although they should not automatically be considered complete protection against re-identification.

Encryption should be applied to sensitive data both during transmission and while stored. Strong identity and access-management mechanisms should ensure that researchers, clinicians, developers, and administrators can access only the information required for their responsibilities. Multi-factor authentication, role-based access control, audit logs, vulnerability assessments, and continuous monitoring can further strengthen security.

Privacy-enhancing technologies can also support AI research. Techniques such as federated learning allow models to be trained across distributed datasets without necessarily transferring all raw data to a central repository. Differential privacy introduces controlled statistical noise to reduce the possibility of identifying individuals from analytical outputs. Secure multi-party computation and other privacy-preserving approaches can similarly enable collaboration while reducing direct exposure of sensitive datasets. However, these methods may involve trade-offs between privacy, computational cost, data utility, and model accuracy. NIST specifically recognizes that privacy-enhancing techniques can sometimes affect accuracy

and therefore require careful consideration within broader AI risk management.

7.1.4 Governance, Consent, and Accountability

Technical safeguards alone cannot guarantee responsible data use. Organizations need comprehensive data-governance policies covering data ownership, consent, access, retention, sharing, breach response, and accountability. The OECD recommends governance mechanisms that combine privacy-protective data use with transparent communication, appropriate authorization processes, de-identification, security practices, and continuous reassessment as technologies and risks evolve (OECD, 2015).

The NIST AI Risk Management Framework provides a broader risk-management approach that encourages organizations to consider trustworthy AI throughout the system lifecycle, including design, development, deployment, use, testing, and evaluation. Its trustworthiness characteristics include security, resilience, accountability, transparency, privacy enhancement, and fairness (NIST, 2023).

In life sciences, informed consent is particularly important. Individuals should, where applicable, be informed about what data are being collected, why they are being collected, who may access them, and whether they may be used for future research or commercial purposes. Consent procedures should also recognize that AI systems may create new uses for data that were difficult to anticipate when the information was originally collected. Consequently, governance frameworks should provide mechanisms for transparency, withdrawal where legally and technically feasible, and responsible secondary use.

Data privacy and security are fundamental requirements for trustworthy AI in life sciences. The enormous value of health, genomic, and biomedical datasets must be balanced against the rights and interests of individuals whose information makes AI research possible. Effective protection requires a combination of privacy-by-design principles, data minimization, encryption, access controls, de-identification, privacy-enhancing technologies, cybersecurity monitoring, informed consent, institutional governance, and continuous risk assessment. The WHO stresses that AI in health should be

developed around ethical and human-rights principles, while frameworks such as the NIST AI RMF provide practical approaches for managing emerging risks. Ultimately, protecting data privacy is not merely a technical or legal obligation; it is essential for maintaining public trust and ensuring that AI-driven advances in life sciences produce meaningful benefits without compromising individual rights.

7.2 ETHICAL ISSUES IN AI-BASED HEALTHCARE AND RESEARCH

Artificial intelligence (AI) is increasingly being integrated into healthcare and life-science research, including medical diagnosis, clinical decision-making, drug discovery, medical imaging, genomics, patient monitoring, and public-health surveillance. AI can process large and complex datasets at a speed that is difficult for humans to achieve and can potentially improve diagnostic accuracy, treatment planning, and research productivity. However, because healthcare involves human life, dignity, privacy, and fundamental rights, the use of AI creates significant ethical challenges. The World Health Organization (WHO) emphasizes that ethics and human rights should remain central to the design, development, deployment, and use of AI in health (WHO, 2021).

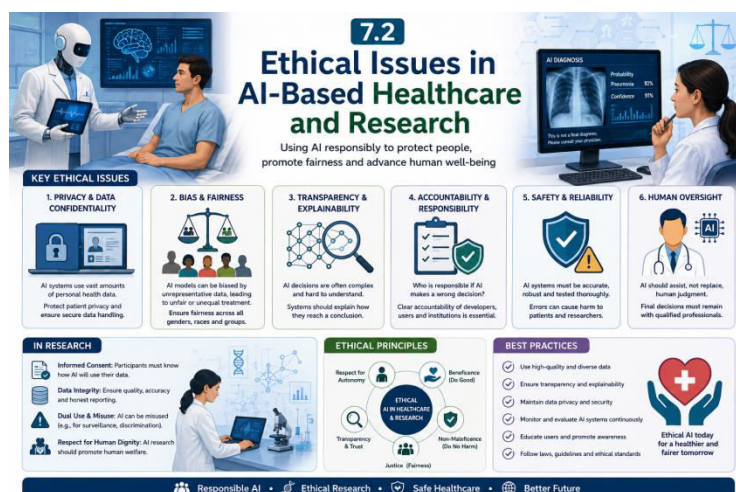


Figure 7.2: Ethical Issues in AI-Based Healthcare and Research

7.2.1 Privacy and Protection of Health Data

Healthcare AI depends heavily on large datasets containing sensitive information such as medical histories, diagnostic reports, genetic profiles, images, laboratory results, and behavioral information. The collection, storage, sharing, and secondary use of such data can create serious privacy concerns. Patients may not always understand how their information will be processed by AI systems or whether their data will be shared with researchers, technology companies, hospitals, or other institutions. In research, large datasets may also be combined from different sources, increasing the possibility of re-identification even when obvious identifiers have been removed. Therefore, strong data-governance mechanisms, appropriate consent procedures, access controls, encryption, and responsible data-sharing practices are essential. Privacy and security are recognized as major limitations and ethical concerns associated with healthcare AI (Topol, 2019).

7.2.2 Algorithmic Bias and Health Inequalities

One of the most important ethical concerns is algorithmic bias. AI systems learn patterns from historical data, and if the training data reflect existing social, economic, gender, racial, geographical, or healthcare inequalities, the resulting algorithm may reproduce or even amplify those inequalities. Bias can arise from under-representation of particular populations, poor-quality data, inappropriate selection of variables, or the use of biased outcomes as prediction targets.

A well-known example was reported by Obermeyer et al. (2019), who demonstrated racial bias in a healthcare algorithm used to identify patients requiring additional care. The algorithm used healthcare expenditure as a proxy for health needs, but because healthcare spending differed systematically between racial groups, the algorithm underestimated the healthcare needs of Black patients. This example demonstrates that removing race as an explicit variable does not necessarily eliminate discrimination. Ethical AI therefore requires systematic assessment of datasets, model performance, and outcomes across different demographic groups.

7.2.3 Transparency and Explainability

Many advanced AI models, particularly deep-learning systems, operate as complex “black boxes.” They may produce highly accurate predictions

without providing an understandable explanation of how a particular conclusion was reached. This creates a significant problem in healthcare because clinicians and patients may need to understand why a particular diagnosis, risk score, or treatment recommendation was generated.

Lack of explainability can also make it difficult to identify errors, challenge inappropriate decisions, or determine responsibility when harm occurs. UNESCO emphasizes that transparency and explainability are important conditions for protecting human rights and ensuring effective accountability in AI systems (UNESCO, 2021). Consequently, healthcare AI should provide explanations appropriate to the clinical context while avoiding unnecessary technical complexity.

7.2.4 Human Oversight and Clinical Responsibility

AI should support healthcare professionals rather than automatically replace human judgment in situations involving significant clinical consequences. An AI system may identify patterns that a clinician overlooks, but it can also generate incorrect predictions because of incomplete data, unusual cases, or differences between the population used for training and the population receiving care. Excessive dependence on AI may lead to automation bias, in which healthcare professionals accept algorithmic recommendations without sufficient independent evaluation.

The ethical principle of human oversight therefore requires clinicians to retain meaningful responsibility for decisions affecting patients. AI recommendations should be considered alongside clinical evidence, professional judgment, and patient preferences. The WHO similarly stresses accountability and the importance of ensuring that AI remains responsive to the healthcare workers and communities affected by its use (WHO, 2021).

7.2.5 Informed Consent and Patient Autonomy

AI also raises difficult questions about informed consent. Traditional consent generally involves explaining the purpose, benefits, risks, and alternatives associated with a medical intervention or research study. With AI, however, patients may not fully understand how their data will be used or how an algorithm will influence clinical decisions. In research, individuals may

consent to the collection of biological or medical data without anticipating that the information could later be used to train AI models for different purposes.

Respect for autonomy requires that patients and research participants receive meaningful information about significant uses of AI and their data. Where appropriate, consent processes should explain data reuse, sharing, potential risks, and the role of AI in decision-making.

7.2.6 Safety, Reliability, and Accountability

AI systems used in healthcare must be sufficiently accurate, reliable, and clinically validated. A model that performs well in a laboratory or controlled research environment may perform poorly in another hospital, geographical region, or demographic population. Changes in clinical practice, patient characteristics, equipment, and data quality can also reduce model performance over time.

Ethical deployment therefore requires continuous monitoring rather than relying solely on pre-deployment testing. Responsibility must also be clearly assigned among developers, healthcare institutions, clinicians, and other stakeholders.

Ethical discussions of healthcare AI emphasize not only technical performance but also questions of traceability, responsibility, and the consequences of system failures (Morley et al., 2020).

7.2.7 Ethical Issues in AI-Based Life-Science Research

AI-driven research introduces additional concerns related to research integrity, authorship, data ownership, reproducibility, and responsible innovation. Researchers may use AI to generate hypotheses, analyze datasets, identify drug candidates, or produce scientific text. However, AI-generated outputs can contain inaccurate information, hidden biases, or fabricated references. Researchers therefore remain responsible for verifying AI-generated results and ensuring that scientific conclusions are supported by reliable evidence.

Another concern is unequal access to AI technologies. Well-funded institutions may possess greater computational resources, high-quality datasets, and specialized expertise than institutions in low- and middle-income settings. Without deliberate efforts to promote inclusion, AI could widen existing global inequalities in healthcare and scientific research. WHO

identifies equitable access, inclusiveness, and public benefit as central considerations for responsible AI in health (WHO, 2021).

Ethical AI in healthcare and life-science research requires more than technical accuracy. Privacy, fairness, transparency, informed consent, human oversight, safety, accountability, and equitable access must be incorporated throughout the AI lifecycle. AI should be designed to complement human expertise while protecting patient autonomy and dignity. As AI becomes increasingly embedded in healthcare and research, multidisciplinary governance involving clinicians, researchers, ethicists, patients, policymakers, and technology developers will be essential. Ultimately, the ethical objective should not simply be to create more powerful AI systems, but to ensure that these systems contribute to safer, fairer, more transparent, and human-centered healthcare and scientific progress.

7.3 REGULATORY POLICIES AND LEGAL FRAMEWORKS

The rapid integration of artificial intelligence (AI) into life sciences has created significant opportunities in medical diagnosis, drug discovery, biomedical research, genomics, personalized medicine, clinical decision-making, and public health. At the same time, the use of AI in these sensitive domains raises complex legal and regulatory questions concerning patient safety, privacy, informed consent, accountability, discrimination, intellectual property, and the reliability of automated decisions. Consequently, regulatory policies are increasingly shifting from technology-neutral principles toward risk-based frameworks that consider the potential impact of AI on individuals and society. The World Health Organization (WHO) emphasizes that AI applications in health should place ethics, human rights, accountability, and public benefit at the centre of their design and deployment (WHO, 2021). (Figure 7.3)



Figure 7.3: Regulatory Policies and Legal Frameworks

7.3.1 Global Regulatory Approaches

There is currently no single international legal framework governing AI across all areas of life sciences. Instead, regulation is developing through a combination of AI-specific legislation, medical-device regulation, data-protection laws, research-ethics requirements, and professional standards. This fragmented regulatory environment reflects differences in national legal systems and healthcare structures.

The European Union (EU) has adopted one of the most comprehensive approaches through the EU Artificial Intelligence Act (Regulation [EU] 2024/1689). The legislation follows a risk-based model in which AI systems are subject to different requirements depending on their potential level of harm. AI applications associated with healthcare and medical devices may fall within the high-risk category, requiring stronger controls relating to risk management, data governance, technical documentation, transparency, human oversight, accuracy, robustness, and cybersecurity. This approach is particularly relevant to life sciences because AI systems can directly influence diagnosis, treatment, and other high-impact decisions.

The United States generally regulates AI-enabled medical technologies through existing sector-specific regulatory structures rather than through a single comprehensive AI law. The U.S. Food and Drug Administration (FDA) evaluates AI-enabled medical devices under its medical-device regulatory framework, with attention to safety and effectiveness and the intended use and technological characteristics of the device. The FDA maintains an AI-enabled medical-device list to provide greater transparency regarding authorized products.

This framework is especially important for machine-learning systems used in medical imaging, diagnosis, monitoring, risk assessment, and treatment support.

The WHO provides an important international governance framework even though its recommendations are not directly equivalent to national legislation. Its guidance calls for AI in healthcare to respect human autonomy, promote human well-being and safety, ensure transparency and explainability, establish responsibility and accountability, promote inclusiveness and equity, and ensure that AI technologies are sustainable and responsive to public needs (WHO, 2021). Similarly, UNESCO's Recommendation on the Ethics of Artificial Intelligence, adopted in 2021, establishes a broader global framework emphasizing human rights, fairness, transparency, accountability, privacy, and environmental and social considerations (UNESCO, 2021).

7.3.2 Data Protection and Privacy

Data protection represents one of the most significant legal issues associated with AI in life sciences. AI systems frequently require large datasets containing electronic health records, genomic information, medical images, laboratory results, and other sensitive personal information. Improper collection, processing, sharing, or re-identification of such data can cause substantial harm to individuals.

In India, the Digital Personal Data Protection Act, 2023 (DPDP Act) establishes a legal framework governing the processing of digital personal data. It defines obligations for data fiduciaries and provides rights and duties concerning personal data. The framework is particularly relevant to AI-based healthcare systems because developers and healthcare institutions may process

large quantities of personal information when training or operating AI models. Effective governance therefore requires appropriate notice, lawful processing, security safeguards, access controls, and mechanisms for addressing individual rights.

7.3.3 Indian Regulatory and Ethical Framework

India is developing a combination of statutory, regulatory, and ethical mechanisms for governing AI in biomedical research and healthcare. The Indian Council of Medical Research (ICMR) published its *Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare* in 2023. These guidelines provide a framework for ethical decision-making during the development, deployment, and adoption of medical AI. They address ethical principles, stakeholder responsibilities, ethics review, governance, and informed consent. Importantly, the guidelines apply to developers, researchers, clinicians, ethics committees, institutions, sponsors, and funding organizations, demonstrating that responsible AI governance requires shared accountability rather than responsibility being placed solely on technology developers (ICMR, 2023).

7.3.4 Accountability, Liability, and Human Oversight

A major unresolved legal question concerns responsibility when an AI system produces an incorrect or harmful recommendation. For example, if an AI diagnostic system fails to identify a serious disease, responsibility could potentially involve the software developer, healthcare institution, clinician, data provider, or manufacturer. Traditional liability frameworks may not always adequately address such situations because AI systems can be complex, continuously updated, and dependent on data from multiple sources.

Regulatory frameworks therefore increasingly emphasize human oversight. AI should generally support rather than completely replace professional judgment in high-stakes medical decisions. Healthcare professionals must be able to understand the intended function and limitations of an AI system and intervene when its recommendations appear unreliable. Developers and institutions should also maintain documentation, validation records, audit trails, and mechanisms for reporting adverse events.

7.3.5 Regulatory Challenges and Future Directions

AI regulation in life sciences remains an evolving process. One important challenge is the adaptive nature of machine-learning systems. Conventional medical-device regulation often evaluates a product at a particular point in time, whereas AI systems may change when algorithms are retrained or updated with new data. Regulators must therefore develop mechanisms for continuous monitoring, post-market surveillance, algorithmic auditing, and controlled modification.

Future regulatory policies are likely to emphasize risk-based governance, transparency, explainability, data quality, cybersecurity, continuous monitoring, human oversight, and international harmonization. Regulatory authorities will also need to balance safety with innovation so that excessive compliance requirements do not prevent beneficial AI applications from reaching patients and researchers.

Overall, effective regulation of AI in life sciences should not be viewed simply as a mechanism for restricting technological development. Rather, it provides the legal and institutional foundation necessary for trustworthy innovation. A strong regulatory ecosystem can protect patients and research participants while simultaneously encouraging responsible development of AI technologies. The long-term objective should therefore be a governance model in which innovation, safety, privacy, human rights, scientific integrity, and public benefit are treated as interconnected requirements.

7.4 SOCIAL IMPACT AND HUMAN-AI COLLABORATION

Artificial intelligence (AI) is increasingly influencing life sciences by transforming healthcare delivery, biomedical research, drug discovery, diagnostics, genomics, public health, and patient management. Its social impact, however, extends beyond technical performance. AI systems can influence how healthcare resources are distributed, how professionals make decisions, how patients interact with healthcare institutions, and how societies understand health and disease. Consequently, the integration of AI into life sciences should be approached as a human-AI collaboration rather than as a simple replacement of human expertise. The World Health Organization (WHO) emphasizes that AI in health should be developed and deployed with

ethics, human rights, accountability, and public benefit at its core (World Health Organization [WHO], 2021). (Figure 7.4)



Figure 7.4: Social Impact and Human-AI Collaboration

7.4.1 Social Impact of AI in Life Sciences

One of the most significant positive social impacts of AI is its potential to improve access to healthcare and scientific knowledge. AI-based diagnostic systems can analyze medical images, laboratory results, genomic information, and patient records rapidly, potentially supporting earlier detection of diseases and improving clinical decision-making. In regions where specialist healthcare professionals are limited, AI-assisted tools may provide additional decision support to frontline healthcare workers. Similarly, AI can accelerate drug discovery by identifying molecular relationships, predicting drug–target interactions, and assisting researchers in screening large numbers of potential compounds. These capabilities may reduce research time and support the development of more personalized approaches to treatment (WHO, 2021).

AI can also affect the social organization of healthcare work. Routine administrative activities, documentation, image analysis, and data processing can increasingly be supported by automated systems. This may allow healthcare professionals to devote more time to communication, complex

decision-making, and patient-centred care. The broader objective should therefore not be to remove human involvement but to redistribute human effort toward tasks requiring empathy, contextual understanding, ethical judgment, and interpersonal communication. Topol (2019) similarly argues that appropriately implemented AI can potentially strengthen rather than weaken the human dimensions of medicine.

However, these benefits are not automatically distributed equally. Access to advanced AI systems may be concentrated in wealthy healthcare institutions, countries, or populations with better digital infrastructure. This creates a potential digital health divide between technologically advanced and resource-constrained communities. If AI becomes an important component of diagnosis or treatment while certain populations remain poorly represented in training datasets, existing inequalities may be reproduced or intensified. UNESCO therefore emphasizes inclusion, fairness, human dignity, and the avoidance of discrimination as important principles for AI development and deployment (UNESCO, 2021).

Algorithmic bias represents another important social concern. AI systems learn from historical data, and those data may reflect existing inequalities in healthcare systems. A widely discussed example is the study by Obermeyer et al. (2019), which demonstrated racial bias in a healthcare algorithm used for managing patient populations. The algorithm relied on healthcare costs as a proxy for healthcare needs, a choice that reflected underlying inequalities in healthcare expenditure and consequently underestimated the needs of some Black patients.

This example illustrates that an algorithm does not need to explicitly use race or another protected characteristic to produce socially unequal outcomes.

7.4.2 Human-AI Collaboration

Human-AI collaboration provides an alternative to the idea that AI will simply replace scientists, physicians, researchers, and other professionals. AI systems are particularly effective at processing enormous datasets, identifying statistical patterns, generating predictions, and performing repetitive computational tasks. Humans, in contrast, contribute contextual knowledge, ethical reasoning, emotional intelligence, professional responsibility, and the

ability to interpret complex social circumstances. Combining these capabilities can produce better outcomes than relying exclusively on either humans or machines.

In clinical practice, for example, an AI system may identify a potentially abnormal radiological feature, estimate a patient's risk of developing a disease, or recommend possible treatment options. The clinician can then evaluate that information against the patient's symptoms, medical history, preferences, social circumstances, and other factors that may not be adequately represented in the algorithm. This creates a collaborative decision-making process in which AI provides analytical support while the human professional retains responsibility for interpretation and action. Char et al. (2018) emphasize that implementing machine learning in healthcare requires careful attention to ethical challenges rather than assuming that technical performance alone guarantees beneficial outcomes.

Effective human-AI collaboration requires meaningful human oversight. Professionals should understand the intended purpose, limitations, uncertainty, and potential failure modes of an AI system before incorporating its recommendations into practice. AI outputs should not automatically be treated as objective or infallible. In high-risk applications, mechanisms should exist for human review, intervention, correction, and, where necessary, rejection of an AI recommendation. The NIST AI Risk Management Framework similarly promotes approaches for identifying, assessing, and managing AI risks throughout the system lifecycle (Tabassi, 2023).

Another essential element is transparency and communication. Patients and research participants should, where appropriate, know when AI contributes substantially to decisions affecting them. Healthcare professionals should also receive adequate training to understand AI-generated recommendations and recognize situations in which an algorithm may be unreliable. Such practices are important for maintaining trust between patients, professionals, institutions, and technology developers.

Human-AI collaboration also requires multidisciplinary participation. AI developers cannot independently determine what constitutes an acceptable healthcare outcome. Physicians, nurses, biomedical researchers, ethicists,

patients, legal experts, policymakers, and affected communities should participate in the design, evaluation, and governance of AI systems. Such participation can help identify social risks that may remain invisible from a purely technical perspective.

7.4.3 Building a Human-Centred AI Future

The future of AI in life sciences should therefore focus on augmentation rather than substitution. AI should enhance human capabilities while preserving human autonomy, dignity, accountability, and professional judgment. Organizations should evaluate AI systems not only according to accuracy or efficiency but also according to fairness, accessibility, safety, explainability, privacy, and their effects on different social groups.

A human-centred approach also requires continuous monitoring after deployment. AI systems may perform differently when used with new populations, hospitals, laboratories, or datasets. Consequently, evaluation should continue throughout the system's operational life rather than ending when an algorithm achieves satisfactory performance during development. WHO (2021) and UNESCO (2021) both stress the importance of accountability, equity, human rights, and responsible governance in ensuring that AI produces social benefit.

Ultimately, the social impact of AI in life sciences will depend less on the technology alone and more on how societies choose to design, govern, and use it. Responsible human-AI collaboration can improve scientific productivity, healthcare quality, and access to services while preserving the uniquely human dimensions of care and judgment. The objective should not be to create systems in which humans become passive recipients of machine decisions, but systems in which AI provides powerful analytical capabilities and humans remain informed, responsible, and empowered decision-makers.

CHAPTER 8

Future Trends and Emerging Innovations in AI and Life Sciences

8.1 AI AND QUANTUM COMPUTING IN LIFE SCIENCES

The convergence of Artificial Intelligence (AI) and quantum computing (QC) represents one of the most promising emerging directions in life sciences. AI has already transformed areas such as drug discovery, biomedical imaging, genomics, clinical decision support, and personalized medicine by enabling computers to identify patterns and relationships in large and complex datasets. Quantum computing, in contrast, introduces a fundamentally different computational paradigm based on quantum-mechanical principles such as superposition and entanglement. The combination of these technologies could eventually enable researchers to address biological and chemical problems that remain difficult or computationally expensive for conventional computers. (Figure 8.1)

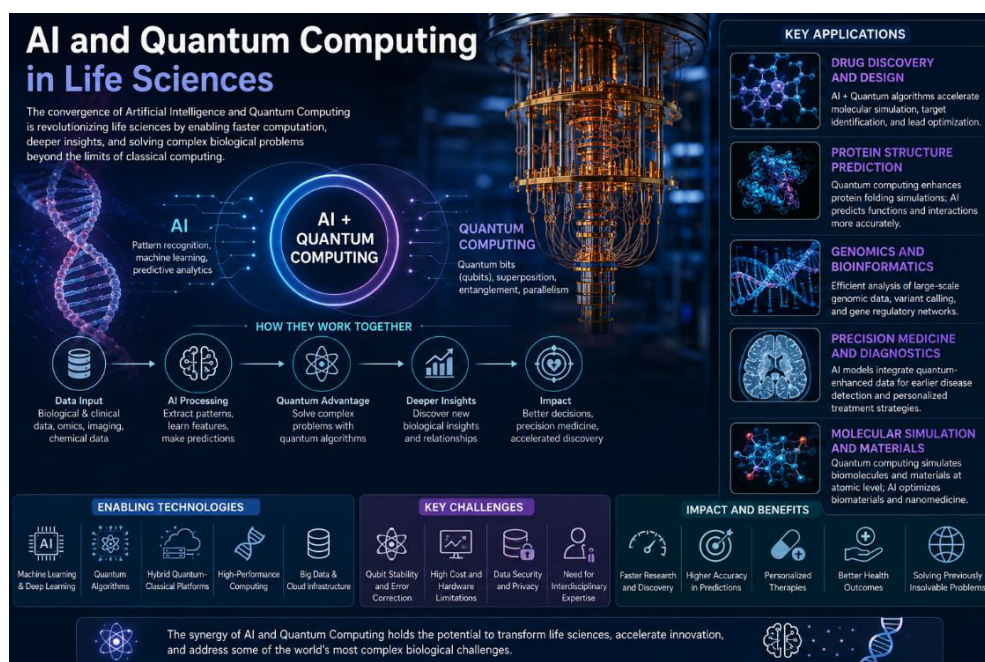


Figure 8.1: AI and Quantum Computing in Life Sciences

Modern life-science research generates enormous quantities of heterogeneous data, including genomic sequences, protein structures, molecular interactions, electronic health records, medical images, and clinical-trial data. Conventional AI and deep-learning methods have become increasingly effective at extracting useful information from these datasets. In drug discovery, for

example, deep learning is being used for molecular property prediction, virtual screening, molecular generation, and structure-based drug design (Tropsha et al., 2024).

Computational approaches have also expanded the ability of researchers to screen extremely large virtual libraries of drug-like molecules and prioritize promising candidates before laboratory testing (Sadybekov & Katritch, 2023). Nevertheless, many biological systems involve highly complex molecular interactions that remain challenging to model accurately.

Quantum computing could complement AI by providing new approaches for representing and solving certain computational problems. Unlike classical computers, which process information using bits represented as 0 or 1, quantum computers use qubits that can exist in quantum superpositions. Quantum algorithms may therefore provide computational advantages for selected classes of problems, although practical advantages in life-science applications have not yet been demonstrated broadly. The most important near-term opportunity is likely to involve hybrid quantum-classical computing, in which quantum processors perform particular computational tasks while classical computers and AI systems handle data preparation, optimization, and interpretation.

One particularly important application is molecular simulation and drug discovery. The behavior of molecules is fundamentally governed by quantum mechanics, making accurate simulation computationally demanding as molecular complexity increases. Quantum computers could potentially provide more efficient representations of molecular electronic structures and chemical interactions. At the same time, AI can learn relationships between molecular structures and properties from experimental and computational datasets. Combining AI with quantum methods could therefore improve molecular property prediction, binding-affinity estimation, virtual screening, and de novo drug design. Recent research on quantum machine learning (QML) has specifically explored quantum neural networks, variational quantum circuits, molecular property prediction, and molecular generation for drug discovery (Smaldone et al., 2025).

Another promising area is protein and biomolecular modelling. Proteins have highly complex three-dimensional structures and dynamic interactions with other molecules. AI-based protein-structure prediction has demonstrated the value of machine learning for understanding biological macromolecules, while quantum computing may eventually contribute to more detailed simulations of molecular interactions. The combination could support improved prediction of protein–ligand interactions and facilitate the development of therapeutic molecules with greater specificity. Neural-network approaches to quantum chemistry are also being investigated as a means of approximating complex electronic wavefunctions and molecular properties (Hermann et al., 2023).

Genomics and precision medicine represent another potential frontier. Genome-scale datasets contain enormous numbers of genetic variables, and AI can identify patterns associated with disease susceptibility, treatment response, and biological traits. Quantum algorithms could eventually be investigated for selected optimization, pattern-recognition, or genomic-analysis tasks. However, the current evidence does not support the assumption that quantum machine learning is already superior to classical machine learning for large-scale healthcare datasets. A systematic review of QML for digital health found limited rigorous evidence of empirical quantum advantage and highlighted important challenges related to data encoding, scalability, and realistic quantum-hardware constraints (Gupta et al., 2025). Therefore, quantum genomics should currently be viewed as an emerging research direction rather than an established clinical technology.

The AI–quantum combination may be particularly valuable because AI can help overcome some limitations of quantum systems. Quantum processors currently suffer from noise, limited qubit numbers, error rates, and difficulties in scaling. Machine-learning techniques can potentially assist with quantum error mitigation, circuit optimization, parameter tuning, and interpretation of quantum outputs. Conversely, quantum processors could eventually accelerate selected computational tasks involved in training or executing AI models. This creates a feedback loop in which AI improves quantum computing while quantum computing potentially expands the capabilities of AI.

Despite its promise, several challenges must be addressed before these technologies can become routine in life-science research. Quantum hardware

remains immature, and many proposed algorithms have only been demonstrated through simulations or small-scale experiments. Biological datasets are also highly heterogeneous and frequently contain missing, biased, or noisy information. Furthermore, transferring quantum algorithms from laboratory demonstrations to clinically meaningful applications requires reliable benchmarks and evidence of real-world benefit. Recent research emphasizes the importance of evaluating QML under realistic hardware conditions rather than relying solely on theoretical performance claims (Gupta et al., 2025). Similarly, current AI-driven drug-discovery research continues to face challenges in translating computational predictions into experimentally and clinically validated outcomes.

Looking ahead, the most realistic pathway is likely to be incremental integration rather than immediate replacement of classical computing. Classical AI will continue to dominate many applications because of its maturity, accessibility, and scalability. Quantum computing is more likely to enter specialized parts of the life-science pipeline where its computational characteristics provide a demonstrable advantage. Hybrid platforms could connect AI models, classical high-performance computing, quantum processors, laboratory automation, and real-world experimental data into integrated discovery systems.

Ultimately, the convergence of AI and quantum computing could contribute to a new generation of AI-driven scientific discovery. Potential long-term applications include faster drug development, improved molecular simulation, personalized treatment selection, advanced biomarker discovery, and more efficient modelling of complex biological systems. However, expectations should remain evidence-based. Quantum computing should not be regarded as a guaranteed solution to every computational problem in life sciences. Its significance will depend on advances in hardware, quantum algorithms, data integration, error correction, and rigorous validation. If these challenges are successfully addressed, AI and quantum computing could evolve from experimental technologies into complementary components of the future life-science research ecosystem.

8.2 DIGITAL HEALTH, IOT, AND SMART BIOLOGICAL SYSTEMS

The convergence of artificial intelligence (AI), digital health, the Internet of Things (IoT), wearable sensors, cloud computing, and biological data analytics is transforming modern life sciences. Rather than treating healthcare as a sequence of isolated clinical encounters, these technologies are creating continuously connected systems in which physiological, environmental, behavioral, and molecular information can be collected, analyzed, and translated into actionable insights. The World Health Organization (WHO) recognizes digital health as an important component of stronger, more equitable, and people-centred health systems. In 2025, the WHO extended its Global Strategy on Digital Health through 2027, reflecting the increasing importance of digital technologies in health-system transformation (WHO, 2025a). (Figure 8.2)

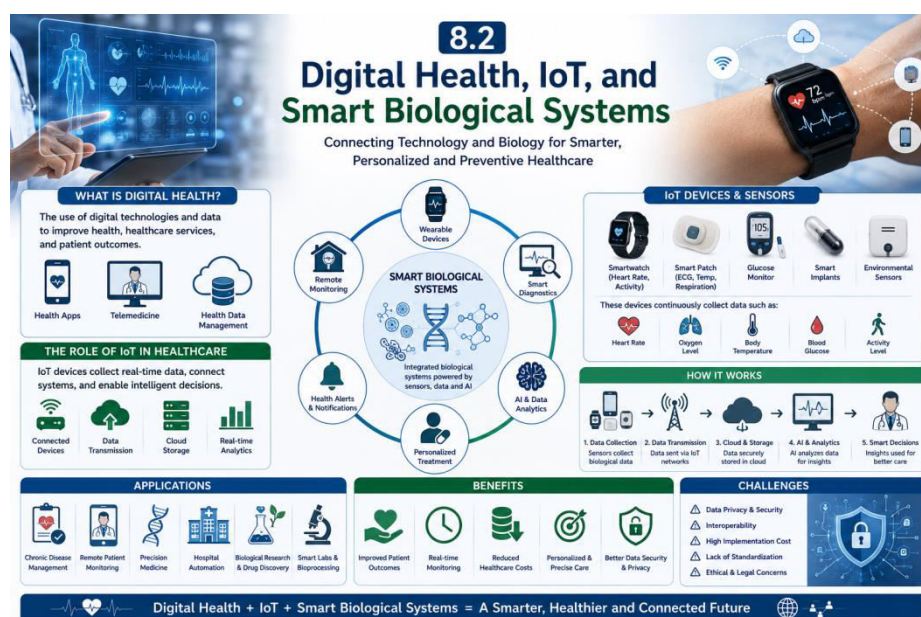


Figure 8.2: Digital Health, IoT, and Smart Biological Systems

8.2.1 Digital Health and Intelligent Healthcare

Digital health encompasses technologies such as electronic health records, telemedicine, mobile health applications, clinical decision-support systems, remote monitoring, digital therapeutics, and AI-enabled diagnostic tools. AI

adds an analytical layer to these systems by identifying patterns in large and complex datasets that may not be readily detectable through conventional approaches. Such systems can support disease prediction, risk stratification, medical imaging, personalized treatment, and clinical workflow optimization. The U.S. Food and Drug Administration (FDA) notes that AI/ML technologies can contribute to early disease detection, diagnosis, prognosis, risk assessment, and the identification of new patterns in physiology and disease progression (FDA, 2025a).

An important future direction is the movement from reactive healthcare toward predictive and preventive healthcare. Instead of waiting for symptoms to become clinically significant, AI systems can analyze longitudinal information to identify changes in an individual's health trajectory. For example, continuous monitoring of heart rate, activity, sleep, glucose, oxygen saturation, or other physiological parameters can help identify abnormalities at an earlier stage. However, these systems must be designed around data quality, interoperability, transparency, privacy, and clinical validation. WHO emphasizes that responsible AI adoption requires appropriate governance, ethical standards, regulation, and attention to equity (WHO, 2024).

8.2.2 IoT and Connected Biological Monitoring

The Internet of Things provides the technological infrastructure through which physical devices can communicate and exchange data. In healthcare and life sciences, IoT can connect wearable devices, implantable sensors, smart medical equipment, laboratory instruments, environmental monitors, and hospital information systems.

These devices generate continuous streams of data that can be transmitted to cloud or edge-computing platforms, where AI algorithms can process them in real time.

AI-based wearable sensors represent an important development in this area. Wearable technologies can collect physiological and behavioral information while individuals perform normal daily activities. Machine-learning models can subsequently convert raw sensor signals into meaningful health indicators, enabling applications in fitness monitoring, chronic disease management,

rehabilitation, elderly care, and early detection of physiological abnormalities (Shajari et al., 2023).

The future of IoT-enabled healthcare is likely to involve increasingly unobtrusive and intelligent sensors. Flexible electronics, biosensors, smart textiles, implantable devices, and miniaturized monitoring systems may allow biological information to be collected continuously with minimal disruption to patients. Such technologies can also contribute to remote healthcare, particularly for individuals who have limited access to specialized clinical facilities.

8.2.3 Smart Biological Systems

Smart biological systems represent a broader concept in which biological entities, sensors, computational models, and AI interact dynamically. These systems may operate at different scales, from molecular and cellular processes to tissues, organs, individuals, and populations. AI can integrate heterogeneous information, including genomic, transcriptomic, proteomic, imaging, clinical, environmental, and behavioral data, to generate more comprehensive representations of biological systems.

One particularly promising development is the emergence of digital twins in healthcare. A digital twin is a dynamic virtual representation connected to a physical individual, organ, disease, or biological system through continuous data exchange. Unlike a static digital model, a sophisticated health digital twin can incorporate real-time information and use AI, simulation, and predictive analytics to estimate future states and evaluate potential interventions (Katsoulakis et al., 2024).

Human-body digital twins could eventually support personalized treatment planning, surgical simulation, disease progression modelling, drug-response prediction, and individualized health management. Current research has already explored digital twins for areas such as oncology, cardiology, clinical trials, drug discovery, medical-device development, and personalized medicine. A systematic review published in 2024 found that digital-twin interventions showed beneficial outcomes across several areas of precision health, although the technology remains at an evolving stage (Shen et al., 2024).

8.2.4 Future Directions and Challenges

The integration of digital health, IoT, and smart biological systems will increasingly depend on interoperability between devices, healthcare institutions, databases, and AI platforms. Standardized data formats and secure data-sharing mechanisms will be essential for creating reliable interconnected ecosystems. At the same time, cybersecurity will become increasingly important because connected medical devices and biological datasets contain highly sensitive information.

Another major challenge is algorithmic bias and reliability. AI models trained on incomplete or unrepresentative datasets may produce inaccurate predictions for certain populations. Consequently, future systems will require continuous validation, transparent documentation, human oversight, and lifecycle monitoring. The FDA's recent guidance initiatives emphasize transparency, bias management, documentation, and lifecycle considerations for AI-enabled medical devices (FDA, 2025b).

Ultimately, the convergence of AI, IoT, digital health, and biological modelling is likely to shift life sciences toward continuously learning and increasingly personalized systems. The long-term objective is not simply to automate healthcare but to establish intelligent environments capable of sensing biological changes, interpreting complex information, predicting risks, and supporting timely interventions. If technological innovation is accompanied by strong governance, ethical safeguards, interoperability, and equitable access, smart biological systems could become a foundational component of next-generation precision and preventive healthcare.

8.3 AI IN PANDEMIC PREDICTION AND GLOBAL HEALTH

The COVID-19 pandemic demonstrated that infectious disease outbreaks can develop rapidly and create consequences that extend far beyond the health sector. Early identification of emerging outbreaks is therefore a central component of global health security. Artificial intelligence (AI), particularly machine learning (ML), deep learning (DL), natural language processing (NLP), and predictive analytics, is increasingly being explored as a means of strengthening disease surveillance, forecasting transmission, identifying high-risk populations, and supporting public-health decision-making. Recent research indicates that the integration of diverse data sources with AI can

improve the development of early-warning systems for infectious diseases (Hu et al., 2025; Rodríguez et al., 2024). (Figure 8.3)

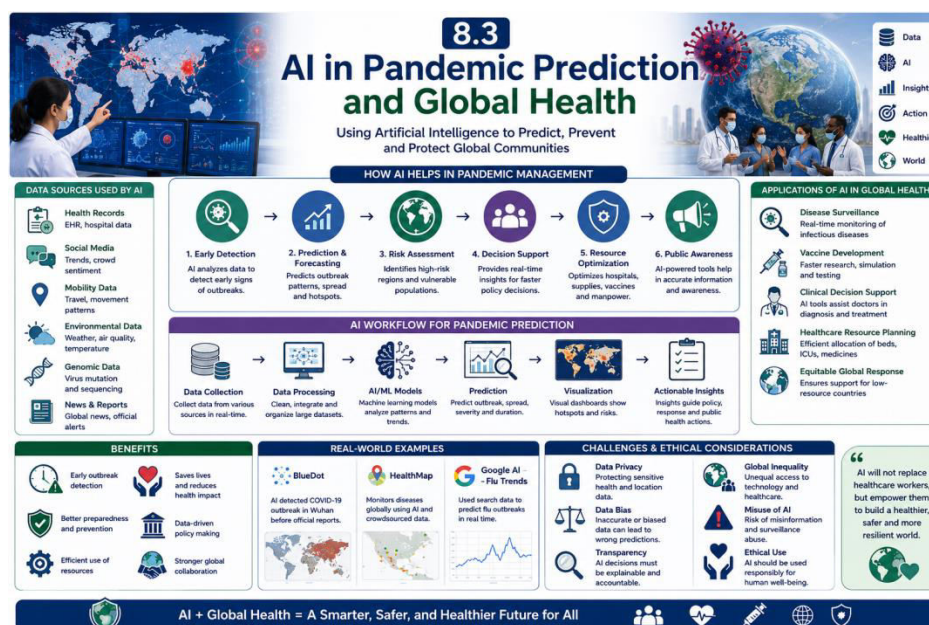


Figure 8.3: AI in Pandemic Prediction and Global Health

8.3.1 AI-Based Pandemic Prediction

Traditional epidemic forecasting generally relies on epidemiological models that describe disease transmission through parameters such as infection rates, recovery rates, population susceptibility, and contact patterns. These approaches remain highly valuable, but they can be difficult to adapt when information is incomplete or when an emerging pathogen behaves differently from previously observed diseases. AI-based models provide an additional approach because they can identify complex relationships within large and heterogeneous datasets.

Machine learning algorithms can process historical disease records, demographic information, environmental variables, mobility patterns, healthcare utilization, vaccination coverage, and other indicators to estimate the probability and potential scale of an outbreak. Deep-learning approaches can identify nonlinear patterns that may be difficult to detect using conventional statistical techniques. Research on epidemic forecasting

increasingly emphasizes data-centric approaches in which model performance depends not only on the sophistication of the algorithm but also on the quality, coverage, timeliness, and integration of the underlying data (Rodríguez et al., 2024).

AI can also support early-warning surveillance by continuously examining information from multiple sources. These may include laboratory reports, hospital admissions, syndromic surveillance, wastewater measurements, genomic surveillance, environmental information, news reports, search activity, and social media. Internet-based surveillance systems can provide signals that complement conventional public-health reporting and potentially identify unusual disease activity before complete clinical datasets become available (McClymont et al., 2024).

8.3.2 Role of NLP and Real-Time Data

Natural language processing has particular importance for pandemic intelligence because substantial amounts of health-related information are generated in unstructured textual formats. AI systems can examine news articles, health-agency announcements, scientific publications, online reports, and other textual sources to identify references to unusual disease events. Automated language analysis can classify reports according to disease, geographic location, severity, and temporal characteristics, helping public-health professionals prioritize potentially important signals.

However, information obtained from online platforms must be interpreted cautiously. Social media and internet data may contain misinformation, duplicate reports, geographical bias, and changes in user behavior that are unrelated to actual disease transmission. Consequently, AI-generated signals should normally be combined with laboratory, clinical, epidemiological, and environmental evidence rather than treated as definitive evidence of an outbreak. Reviews of infectious-disease early-warning systems emphasize the importance of integrating multiple sources of information rather than depending on a single indicator (Hu et al., 2025).

8.3.3 AI, Genomic Surveillance, and Variant Prediction

The increasing availability of pathogen genomic data creates another important application area for AI. Genomic surveillance enables researchers to

identify mutations and monitor the emergence and geographical distribution of pathogen variants. AI and computational methods can help analyze large genomic datasets, classify genetic patterns, and support assessment of potential changes in pathogen behavior.

During and after COVID-19, the combination of laboratory science, genomic surveillance, data analytics, and AI demonstrated the potential of technology-supported pandemic intelligence.

Future systems are likely to connect genomic information with epidemiological and clinical datasets to provide a more comprehensive understanding of emerging pathogens. Such integration could help researchers identify unusual genetic changes and prioritize variants for further laboratory and epidemiological investigation.

8.3.4 AI for Global Health Decision-Making

The value of pandemic prediction is ultimately determined by whether predictions can support timely action. AI systems can assist governments and health organizations in estimating potential healthcare demand, identifying regions at increased risk, planning medical supplies, allocating healthcare personnel, and evaluating possible intervention strategies. During COVID-19, AI applications were explored for forecasting, diagnosis, decision support, drug repurposing, and outbreak management (Alowais et al., 2024).

For example, an AI forecasting system could indicate that a particular region has a rising probability of increased infections over the following weeks. Public-health authorities could then use this information to increase testing capacity, strengthen hospital preparedness, improve communication, prepare essential medicines, or intensify surveillance. AI therefore has greater value when incorporated into a broader decision-making framework rather than functioning as an isolated prediction tool.

The World Health Organization (WHO) has emphasized that AI already has applications across disease surveillance, outbreak response, healthcare, and health-system management. At the same time, WHO stresses the importance of safety, equity, governance, and responsible implementation so that technological innovation does not increase existing health inequalities (WHO, 2024).

8.3.5 Challenges and Limitations

Despite its potential, AI cannot guarantee accurate prediction of the next pandemic. One major challenge is data quality. Disease surveillance data may be incomplete, delayed, inconsistent, or unavailable in low-resource settings. AI models trained primarily on data from high-income countries may therefore perform poorly when applied to different populations and healthcare systems.

Another challenge is model uncertainty. A highly accurate model under historical conditions may fail when a new pathogen emerges or when human behavior changes substantially. Pandemic conditions can produce unprecedented situations for which historical training data provide limited guidance. Researchers therefore increasingly emphasize robust evaluation, uncertainty estimation, external validation, and continuous updating of forecasting systems (Rodríguez et al., 2024).

Privacy and ethical concerns are also important. Pandemic surveillance may involve sensitive health, mobility, location, or demographic information. Appropriate governance mechanisms are required to ensure that data are collected and used proportionately, securely, and transparently. WHO's approach to AI for health emphasizes governance and standards intended to promote trustworthy and equitable adoption of AI technologies.

8.3.6 Future Directions

The future of AI-enabled pandemic prediction will likely involve multimodal and collaborative intelligence. Instead of relying on a single data source, future platforms may combine clinical records, genomic sequences, wastewater surveillance, climate data, mobility information, laboratory results, satellite observations, and online information. Advanced AI models may then generate continuously updated risk assessments for countries and regions.

Another important development will be the integration of AI with global collaborative surveillance networks. The WHO Hub for Pandemic and Epidemic Intelligence was established to strengthen the world's ability to detect, prepare for, and respond to health threats through data science, AI, laboratory detection, and collaborative surveillance. Future systems could increasingly connect national surveillance platforms so that emerging signals are identified and shared more rapidly across borders.

AI may therefore become an important component of a global early-warning ecosystem, but it should complement rather than replace epidemiologists, clinicians, laboratory scientists, public-health officials, and policymakers. The most effective approach will combine computational intelligence with human expertise, reliable surveillance infrastructure, transparent governance, and international cooperation.

WHO's research priorities for pandemic and epidemic intelligence similarly emphasize high-quality evidence and improved translation of surveillance research into informed public-health decisions (WHO, 2024).

Overall, AI has the potential to transform pandemic preparedness from a predominantly reactive model toward a more predictive and proactive system. Its greatest contribution may not be predicting exactly when or where the next pandemic will occur, but identifying weak signals earlier, estimating potential risks, improving preparedness, and helping decision-makers act before an outbreak becomes a global crisis.

8.4 FUTURE RESEARCH DIRECTIONS AND EMERGING INNOVATIONS



Figure 8.4: Future Research Directions and Emerging Innovations

Artificial intelligence (AI) is rapidly transforming life sciences from a collection of specialized computational applications into an increasingly integrated, data-driven scientific ecosystem. Future research is expected to move beyond conventional machine-learning models toward foundation models, multimodal AI, generative systems, autonomous research agents, digital twins, and AI-assisted experimental platforms. These developments have the potential to accelerate biological discovery, improve disease prediction, personalize treatment, and reduce the time and cost associated with drug and biotechnology development. At the same time, future research must address limitations related to data quality, interpretability, validation, privacy, bias, and responsible governance.

One of the most important future directions is the development of biological foundation models. Unlike conventional AI systems designed for a single task, foundation models can be pretrained on very large datasets and subsequently adapted to different biological applications. Such models can learn representations from DNA and RNA sequences, proteins, molecular structures, biomedical literature, clinical records, and other biological data. Their future development is likely to involve increasingly specialized models for genomics, proteomics, drug discovery, pathology, and clinical medicine. Foundation models are particularly promising because biological datasets frequently contain large amounts of unlabeled information, while high-quality annotation can be expensive and time-consuming (Li et al., 2024).

The next generation of models is therefore expected to combine self-supervised learning with biological knowledge and experimental evidence.

A second major research direction is multimodal AI. Biological phenomena are inherently multidimensional and cannot always be understood from a single data type. Future AI systems will increasingly integrate genomic sequences, transcriptomic profiles, proteomic information, medical images, electronic health records, physiological signals, and environmental data. Large multimodal models could allow researchers to examine relationships between molecular mechanisms and clinical outcomes within a unified computational framework. The World Health Organization (WHO, 2024) recognizes large multimodal models as an emerging technology with potential applications in healthcare, scientific research, public health, and drug development. However,

the reliability of multimodal systems will depend on the quality and compatibility of the underlying datasets.

AI-driven drug discovery and molecular design will remain another important area of innovation. Traditional drug development involves lengthy processes of target identification, compound screening, optimization, preclinical testing, and clinical evaluation. AI can support these stages by predicting molecular properties, identifying potential drug targets, generating candidate molecules, and estimating interactions between drugs and biological targets. AlphaFold 3 demonstrates the direction in which this field is moving: its unified deep-learning architecture can model complexes involving proteins, nucleic acids, small molecules, ions, and modified residues, expanding AI-assisted structural biology beyond protein structure prediction alone (Abramson et al., 2024). Future research may combine generative chemistry, molecular simulation, structural prediction, and experimental robotics to create iterative design–test–learn cycles.

Another emerging direction is the development of AI-generated biological hypotheses and autonomous scientific discovery. Rather than merely analyzing existing datasets, future AI systems may help formulate hypotheses, identify relationships that have not previously been recognized, recommend experiments, and interpret experimental results. This could create a new form of human–AI collaboration in which scientists provide research objectives while AI systems assist with literature analysis, experimental design, computational modelling, and prioritization of promising hypotheses. Such systems could substantially increase the speed of research; however, experimental verification will remain essential because computational predictions should be treated as testable hypotheses rather than definitive scientific conclusions (Richardson et al., 2024).

Digital twins and virtual biological models represent another promising research frontier. A digital twin is a computational representation of a biological system that can be continuously updated using relevant data. In life sciences, future digital twins could represent patients, organs, tumors, or cellular systems. Patient-level digital twins may eventually integrate medical history, imaging, genomics, laboratory measurements, and physiological signals to simulate disease progression or compare potential treatment

strategies. At the cellular level, AI-based models could help researchers investigate how genetic and environmental changes influence cellular behavior. Achieving this vision will require substantial advances in causal modelling, longitudinal data integration, mechanistic biology, and computational simulation.

Future research will also increasingly focus on AI-assisted precision medicine. Current clinical AI applications frequently concentrate on individual tasks, such as image classification or disease prediction. Future systems are expected to integrate multiple sources of patient information to support individualized risk assessment, diagnosis, treatment selection, and monitoring. The objective will be to move from population-level predictions toward models that can account for individual biological variation. Nevertheless, successful implementation will require prospective clinical validation, interoperability with healthcare systems, robust evaluation across demographic groups, and mechanisms for clinician oversight.

The integration of AI with laboratory automation and robotics may further change the research process. Automated laboratories can perform repetitive experiments, collect measurements, and feed experimental results back into computational models. Combined with AI, these systems could continuously select the next experiment based on previous outcomes. Such closed-loop systems may be particularly valuable in drug discovery, synthetic biology, protein engineering, and microbiology.

AI would therefore become not only a tool for analysing scientific data but also a component of an automated experimental discovery cycle.

Finally, responsible and trustworthy AI must remain a central research priority. As AI becomes more influential in healthcare and biological research, questions of privacy, fairness, transparency, accountability, explainability, and safety will become increasingly important. WHO guidance emphasizes that AI for health should be developed and deployed with appropriate ethical and governance safeguards (WHO, 2021; WHO, 2024). Future research should therefore develop standardized benchmarks, explainability techniques, bias-detection methods, privacy-preserving learning, reproducible evaluation

frameworks, and regulatory approaches that can keep pace with technological development.

Overall, the future of AI in life sciences is likely to be characterized by greater integration rather than isolated applications. Foundation models, multimodal systems, generative AI, molecular design platforms, digital twins, autonomous laboratories, and precision medicine may increasingly operate as interconnected components of a single research ecosystem. The most important challenge will not simply be developing more powerful AI models, but ensuring that these systems generate scientifically reliable, clinically meaningful, reproducible, and ethically responsible outcomes. The convergence of computational intelligence with experimental biology therefore has the potential to reshape how biological knowledge is generated and translated into practical solutions.

REFERENCES

- On digital twins in bioprocessing: Opportunities and limitations. (2025). *Process Biochemistry*, 156, 274–299. <https://doi.org/10.1016/j.procbio.2025.05.023>
- U.S. Department of Health & Human Services. (2026). The Security Rule. HHS.
- Vasey, B., Lippert, K. A. N., Khan, D. Z., et al. (2023). Intraoperative applications of artificial intelligence in robotic surgery: A scoping review of current development stages and levels of autonomy. *Annals of Surgery*, 278(6), 896–903. <https://doi.org/10.1097/SLA.0000000000005700>
- World Health Organization. (2021). Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. World Health Organization.
- Yang, Y., Zhang, H., Gichoya, J. W., et al. (2024). The limits of fair medical imaging AI in real-world generalization. *Nature Medicine*, 30, 2838–2848. <https://doi.org/10.1038/s41591-024-03113-4>
- Yi, L., Wang, W., Diao, Y., Yi, S., Shang, Y., Ren, D., Ge, K., Gu, Y., et al. (2024). Recent advances of artificial intelligence in quantitative analysis of food quality and safety indicators: A review. *TrAC Trends in Analytical Chemistry*, 180, 117944. <https://doi.org/10.1016/j.trac.2024.117944>
- Zhu, L., Spachos, P., Pensini, E., & Plataniotis, K. N. (2021). Deep learning and machine vision for food processing: A survey. *arXiv*. <https://arxiv.org/abs/2103.16106>
- Abioye, E. A., et al. (2024). The application of machine learning techniques for smart irrigation systems: A systematic literature review. *Smart Agricultural Technology*, 7, 100425. <https://doi.org/10.1016/j.atech.2024.100425>.
- Abramson, J., Adler, J., Dunger, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630, 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
- Abudayyeh, O. O., & Gootenberg, J. S. (2024). Programmable biology through artificial intelligence: From nucleic acids to proteins to cells. *Nature Methods*, 21, 1384–1386. <https://doi.org/10.1038/s41592-024-02338-y>
- Acosta, J. N., Falcone, G. J., Rajpurkar, P., & Topol, E. J. (2022). Multimodal biomedical AI. *Nature Medicine*, 28, 1773–1784. <https://doi.org/10.1038/s41591-022-01981-2>

-
-
- Alam, M. A., et al. (2024). Artificial intelligence as a tool in drug discovery and development. *World Journal of Experimental Medicine*. <https://doi.org/10.5493/wjem.v14.i4.99>.
 - Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., and Badreldin, H. A. (2024). Innovative applications of artificial intelligence during the COVID-19 pandemic. *Infectious Medicine*, 3(1), 100095. <https://doi.org/10.1016/j.imj.2024.100095>.
 - Applications of Artificial Intelligence in Biotech Drug Discovery and Product Development. (2025). Review of AI applications in small-molecule design, protein binders, antibody engineering, and biopharmaceutical product development.
 - Ardila, D., Kiraly, A. P., Bharadwaj, S., et al. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, 25(6), 954–961. <https://doi.org/10.1038/s41591-019-0447-x>
 - Ashley, E. A. (2016). Towards precision medicine. *Nature Reviews Genetics*, 17, 507–522. <https://doi.org/10.1038/nrg.2016.86>
 - Aversano, L., Bernardi, M. L., Cimitile, M., & Iammarino, M. (2024). A systematic review on artificial intelligence approaches for smart health devices. *Journal of Medical Systems*, 48, 1–21.
 - Bagga, M. (2024). Image-based detection and classification of plant diseases using deep learning: State-of-the-art review. *Urban Agriculture & Regional Food Systems*, 9(1), e20053. <https://doi.org/10.1002/uar.2.20053>.
 - Ben Abbes, A., Jarray, N., & Farah, I. R. (2024). Advances in remote sensing based soil moisture retrieval: Applications, techniques, scales and challenges for combining machine learning and physical models. *Artificial Intelligence Review*, 57, 224. <https://doi.org/10.1007/s10462-024-10734-1>.
 - Ben Abbes, A., Jarray, N., & Farah, I. R. (2024). Advances in remote sensing based soil moisture retrieval: Applications, techniques, scales and challenges for combining machine learning and physical models. *Artificial Intelligence Review*, 57, 224. <https://doi.org/10.1007/s10462-024-10734-1>
 - Cella, L., Lin, J., Goldenberg, M. G., & Hung, A. J. (2026). The future of robotic surgery in the age of artificial intelligence. *Nature Reviews Urology*. <https://doi.org/10.1038/s41585-026-01153-8>
-
-

-
-
- Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care-Addressing ethical challenges. *New England Journal of Medicine*, 378(11), 981–983. <https://doi.org/10.1056/NEJMp1714229>
 - Chee, S. W., et al. (2026). Hit identification in ultra large virtual screening: An integrative review and future challenges. *Drug Discovery Today*. <https://doi.org/10.1016/j.drudis.2026.104616>
 - Cheng, T., Li, Q., Zhou, Z., Wang, Y., & Bryant, S. H. (2012). Structure-based virtual screening for drug discovery: A problem-centric review. *The AAPS Journal*, 14(1), 133–141. <https://doi.org/10.1208/s12248-012-9322-0>
 - Deep learning based abiotic crop stress assessment for precision agriculture: A comprehensive review. (2025). *Journal of Environmental Management*, 125158.
 - Demissie, W. A., Sebastiani, L., & Rossetto, R. (2026). Integration of artificial intelligence and remote sensing for crop yield prediction and crop growth parameter estimation in Mediterranean agroecosystems: Methodologies, emerging technologies, research gaps, and future directions. *European Journal of Agronomy*, 173, 127894.
 - Deng, J., Yang, Z., Ojima, I., Samaras, D., & Wang, F. (2021). Artificial intelligence in drug discovery: Applications and techniques. *arXiv*. <https://arxiv.org/abs/2106.05386>.
 - Di Blasi, R., Gabrielli, J., Shabestary, K., et al. (2024). Understanding resource competition to achieve predictable synthetic gene expression in eukaryotes. *Nature Reviews Bioengineering*, 2, 721–732.
 - Ding, C., Ou Yang, O., & Ierapetritou, M. G. (2024). Towards digital twin for biopharmaceutical processes: Concept and progress. In *Biopharmaceutical Manufacturing* (pp. 179–211). Springer.
 - Du, Y., Jamasb, A. R., Guo, J., et al. (2024). Machine learning-aided generative molecular design. *Nature Machine Intelligence*, 6, 589–604. <https://doi.org/10.1038/s42256-024-00843-5>
 - Dunn, A. L., Burgess, S. J., & Garcia Martin, H. (2020). Machine learning for metabolic engineering: A review. *Metabolic Engineering*, 63, 1–15. <https://doi.org/10.1016/j.ymben.2020.10.005>.
 - Duong-Trung, N., Born, S., Kim, J. W., et al. (2022). When bioprocess engineering meets machine learning: A survey from the perspective of automated bioprocess development. *arXiv*.
-
-

-
-
- Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Deep learning: New computational modelling techniques for genomics. *Nature Reviews Genetics*, 20, 389–403. <https://doi.org/10.1038/s41576-019-0122-6>.
 - Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Deep learning: New computational modelling techniques for genomics. *Nature Reviews Genetics*, 20, 389–403. <https://doi.org/10.1038/s41576-019-0122-6>
 - Esteva, A., Chou, K., Yeung, S., et al. (2021). Deep learning-enabled medical computer vision. *npj Digital Medicine*, 4, 5. <https://doi.org/10.1038/s41746-020-00376-2>
 - Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 115–118. <https://doi.org/10.1038/nature21056>
 - Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 115–118. <https://doi.org/10.1038/nature21056>.
 - European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).
 - European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
 - FDA-approved artificial intelligence and machine learning-enabled medical devices in general surgery: A review based on regulatory panel classification. (2026). *Surgery*, 190, 109846. <https://doi.org/10.1016/j.surg.2025.109846>
 - Food and Agriculture Organization of the United Nations (FAO). (2017). Can Artificial Intelligence help improve agricultural productivity? FAO e-Agriculture.
 - Food and Agriculture Organization of the United Nations (FAO). (2022). The State of Food and Agriculture 2022: Leveraging automation in agriculture for transforming agrifood systems. FAO.
 - Food and Agriculture Organization of the United Nations. (2021). Digital agriculture in action: Artificial intelligence for agriculture. FAO & ITU.
-
-

-
-
- Food and Agriculture Organization of the United Nations. (2026). Agricultural water management. FAO.
 - Food and Agriculture Organization of the United Nations. (2026). Irrigation management. FAO.
 - Food and Drug Administration (FDA). (2025). Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products: Draft guidance for industry and other interested parties. U.S. Department of Health and Human Services.
 - Food and Drug Administration (FDA). (2026). Artificial intelligence for drug development. U.S. Department of Health and Human Services.
 - Fu, Z., Yang, Y., Chung, M. K., & Cheng, F. (2026). Artificial intelligence to guide repurposing of drugs. *Annual Review of Medicine*, 77, 381–398. <https://doi.org/10.1146/annurev-med-050224-122802>.
 - Fushimi, K., Nakai, Y., Nishi, A., et al. (2025). Development of the autonomous lab system to support biotechnology research. *Scientific Reports*, 15, 6648. <https://doi.org/10.1038/s41598-025-89069-y>
 - Gangwal, A., & Lavecchia, A. (2024). Unleashing the power of generative AI in drug discovery. *Drug Discovery Today*, 103992. <https://doi.org/10.1016/j.drudis.2024.103992>.
 - Gangwal, A., Ansari, A., Ahmad, I., Azad, A. K., & Wan Sulaiman, W. M. A. (2024). Current strategies to address data scarcity in artificial intelligence-based drug discovery: A comprehensive review. *Computers in Biology and Medicine*, 179, 108734. <https://doi.org/10.1016/j.combiomed.2024.108734>.
 - Gao, H., Zhangzhong, L., Zheng, W., & Chen, G. (2023). How can agricultural water production be promoted? A review on machine learning for irrigation. *Journal of Cleaner Production*, 414, 137687. <https://doi.org/10.1016/j.jclepro.2023.137687>.
 - Gbashi, S., & Njobeh, P. B. (2024). Enhancing food integrity through artificial intelligence and machine learning: A comprehensive review. *Applied Sciences*, 14(8), 3421. <https://doi.org/10.3390/app14083421>
 - Ghislat, G., Hernandez-Hernandez, S., Piyawajanusorn, C., & Ballester, P. J. (2024). Data-centric challenges with the application and adoption of artificial intelligence for drug discovery. *Expert Opinion on Drug Discovery*, 19(11), 1297–1307. <https://doi.org/10.1080/17460441.2024.2403639>.
-
-

-
-
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
 - Gorgulla, C., et al. (2023). Recent developments in structure-based virtual screening approaches. *Current Opinion in Structural Biology*.
 - Goshisht, M. K. (2024). Machine learning and deep learning in synthetic biology: Key architectures, applications, and challenges. *ACS Omega*, 9, 9921–9945. <https://doi.org/10.1021/acsomega.3c05913>
 - Government of India. (2023). The Digital Personal Data Protection Act, 2023. India Code, Ministry of Law and Justice.
 - Gulshan, V., Peng, L., Coram, M., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
 - Gupta, R. S., Wood, C. E., Engstrom, T., Pole, J. D., & Shrapnel, S. (2025). Quantum machine learning for digital health? A systematic review. *npj Digital Medicine*, 8. <https://doi.org/10.1038/s41746-025-01597-z>
 - Han, H., Liu, Z., Li, J., et al. (2024). Challenges in remote sensing based climate and crop monitoring: Navigating the complexities using AI. *Journal of Cloud Computing*, 13, 34. <https://doi.org/10.1186/s13677-023-00583-8>
 - Hasin, Y., Seldin, M., & Lusi, A. (2017). Multi-omics approaches to disease. *Genome Biology*, 18, 83. <https://doi.org/10.1186/s13059-017-1215-1>
 - Hassan, M. A., et al. (2024). A review of machine learning techniques in agroclimatic studies. *Agriculture*, 14(3), 481. <https://doi.org/10.3390/agriculture14030481>
 - Hassan, M. M., Xu, Y., Sayada, J., Zareef, M., Shoaib, M., Chen, X., Li, H., & Chen, Q. (2025). Progress of machine learning-based biosensors for the monitoring of food safety: A review. *Biosensors and Bioelectronics*, 267, 116782. <https://doi.org/10.1016/j.bios.2024.116782>
 - Hermann, J., Spencer, J., Choo, K., Mezzacapo, A., Foulkes, W. M. C., Pfau, D., Carleo, G., & Noé, F. (2023). Ab initio quantum chemistry with neural-network wavefunctions. *Nature Reviews Chemistry*, 7, 692–709. <https://doi.org/10.1038/s41570-023-00516-8>
 - Hu, W.-H., Sun, H.-M., & Wei, Y.-Y. (2025). Global infectious disease early warning models: An updated review and lessons from the COVID-19
-
-

-
-
- pandemic. *Infectious Disease Modelling*, 10(2), 410–422. <https://doi.org/10.1016/j.idm.2024.12.001>.
- Huang, J. D., Ramsey, E., Condell, J., & others. (2022). Applying artificial intelligence to wearable sensor data to diagnose and predict cardiovascular disease: A review. *Sensors*, 22.
 - Huang, K., Chandak, P., Wang, Q., et al. (2024). A foundation model for clinician-centered drug repurposing. *Nature Medicine*, 30, 3601–3613. <https://doi.org/10.1038/s41591-024-03233-x>.
 - Indian Council of Medical Research (ICMR). (2023). Ethical guidelines for application of artificial intelligence in biomedical research and healthcare. Government of India.
 - Intelligent Pharmacy. (2024). AI's role in revolutionizing personalized medicine by reshaping pharmacogenomics and drug therapy. *Intelligent Pharmacy*, 2(5), 643–650. <https://doi.org/10.1016/j.ipha.2024.08.005>
 - Isoko, K., Cordiner, J. L., Kis, Z., & Moghadam, P. Z. (2024). Bioprocessing 4.0: A pragmatic review and future perspectives. *Digital Discovery*, 3, 1662–1681. <https://doi.org/10.1039/D4DD00127C>
 - Javed, M. A., & Murad, M. A. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*, 10(24), e40836. <https://doi.org/10.1016/j.heliyon.2024.e40836>
 - Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., et al. (2019). Predicting splicing from primary sequence with deep learning. *Cell*, 176(3), 535–548.e24. <https://doi.org/10.1016/j.cell.2018.12.015>.
 - Jiménez-Luna, J., Grisoni, F., Schneider, G., & others. (2021). Deep learning in virtual screening: Recent applications and developments. *International Journal of Molecular Sciences*, 22(9), 4435.
 - Jin, W., et al. (2023). Machine learning for metabolic pathway optimization: A review. *Frontiers in Bioengineering and Biotechnology*.
 - Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
 - Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
-
-

-
-
- Katsoulakis, E., Wang, Q., Wu, H., et al. (2024). Digital twins for health: A scoping review. *npj Digital Medicine*, 7, 77. <https://doi.org/10.1038/s41746-024-01073-0>.
 - Khan, M. A., et al. (2024). Artificial intelligence applications to genomic data in cancer research: A review of recent trends and emerging areas. *Discover Analytics*.
 - Kučera, T. (2016). Virtual screening in drug design-Overview of most frequent techniques. *Medicinal and Medical Sciences Letters*, 85(2), 75–79. <https://doi.org/10.31482/mmsl.2016.014>
 - Larkin, P. M. K., et al. (2021). Emerging applications of machine learning in food safety. *Annual Review of Food Science and Technology*, 12, 513–538. <https://doi.org/10.1146/annurev-food-071720-024112>
 - Lawson, C. E., Martí, J. M., Radivojevic, T., et al. (2021). Machine learning for metabolic engineering: A review. *Metabolic Engineering*, 63, 34–52. <https://doi.org/10.1016/j.jymben.2020.10.005>
 - LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
 - Leong, S. X., Griesbach, C. E., Zhang, R., et al. (2025). Steering towards safe self-driving laboratories. *Nature Reviews Chemistry*, 9, 707–722.
 - Li, Q., Hu, Z., Wang, Y., Li, L., Fan, Y., King, I., Song, L., & Li, Y. (2024). Progress and opportunities of foundation models in bioinformatics. *arXiv preprint arXiv:2402.04286*.
 - Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6), 321–332. <https://doi.org/10.1038/nrg3920>
 - Litjens, G., Kooi, T., Bejnordi, B. E., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
 - Mahmud, M. I., & Banerjee, T. (2026). Artificial intelligence in genomics: A comprehensive survey of methods, resources, challenges, and prospects. *Briefings in Bioinformatics*, 27(3), bbag229. <https://doi.org/10.1093/bib/bbag229>
 - McClymont, H., Lambert, S. B., Barr, I., & Vardoulakis, S. (2024). Internet-based surveillance systems and infectious diseases prediction: An
-
-

-
-
- updated review of the last 10 years and lessons from the COVID-19 pandemic. *Journal of Epidemiology and Global Health*, 14, 645–657.
- Meskó, B., & Görög, M. (2020). A short guide for medical professionals in the era of artificial intelligence. *NPJ Digital Medicine*, 3, 126.
 - Min, S., Lee, B., & Yoon, S. (2017). Deep learning in bioinformatics. *Briefings in Bioinformatics*, 18(5), 851–869. <https://doi.org/10.1093/bib/bbw068>
 - Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 1419. <https://doi.org/10.3389/fpls.2016.01419>.
 - Mondal, P. P., Galodha, A., Verma, V. K., et al. (2023). Review on machine learning-based bioprocess optimization, monitoring, and control systems. *Bioresource Technology*, 370, 128523. <https://doi.org/10.1016/j.biortech.2022.128523>
 - Morley, J., Machado, C. C. V., Burr, C., Cows, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine*, 260, 113172. <https://doi.org/10.1016/j.socscimed.2020.113172>
 - Moshawrab, M., Adda, M., & Bouzouane, A. (2023). Smart wearables for the detection of cardiovascular diseases: A systematic literature review. *Sensors*, 23(2), 828. <https://doi.org/10.3390/s23020828>
 - National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce.
 - National Institute of Standards and Technology. (2023). NIST AI Risk Management Framework. U.S. Department of Commerce.
 - Ngugi, H. N., Ezugwu, A. E., Akinyelu, A. A., et al. (2024). Revolutionizing crop disease detection with computational deep learning: A comprehensive review. *Environmental Monitoring and Assessment*, 196, 302. <https://doi.org/10.1007/s10661-024-12454-z>.
 - Notin, P., Rollins, N., Gal, Y., et al. (2024). Machine learning for functional protein design. *Nature Biotechnology*, 42, 216–228. <https://doi.org/10.1038/s41587-024-02127-0>
-
-

-
-
- Notin, P., Rollins, N., Gal, Y., Sander, C., & Marks, D. S. (2024). Machine learning for functional protein design. *Nature Biotechnology*, 42, 216–228. <https://doi.org/10.1038/s41587-024-02127-0>.
 - Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
 - OECD. (2015). *Health Data Governance: Privacy, Monitoring and Research*. OECD Publishing. <https://doi.org/10.1787/9789264244566-en>
 - Olawade, D. B., Kade, A., Egbon, E., Usman, S. O., Fapohunda, O., Ijiwade, J., & Ogbonna, C. E. (2025). Bioinformatics and artificial intelligence in genomic data analysis: Current advances and future directions. *Molecular Genetics and Genomics*, 300, 111. <https://doi.org/10.1007/s00438-025-02314-x>
 - Pacal, I., Kunduracioglu, I., Alma, M. H., et al. (2024). A systematic review of deep learning techniques for plant diseases. *Artificial Intelligence Review*, 57, 304. <https://doi.org/10.1007/s10462-024-10944-7>.
 - Parra-López, C., Ben Abdallah, S., Garcia-Garcia, G., et al. (2025). Digital technologies for water use and management in agriculture: Recent applications and future outlook. *Agricultural Water Management*.
 - Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14, 417–419. <https://doi.org/10.1038/nmeth.4197>.
 - Peng, J., et al. (2026). Machine learning methods for small data and upstream bioprocessing applications: A comprehensive review. *Biotechnology Advances*, 87, 108749. <https://doi.org/10.1016/j.biotechadv.2025.108749>.
 - Peterson, J. F., Feng, Q., Malin, B. A., & Wei, W.-Q. (2024). Leveraging generative AI to prioritize drug repurposing candidates for Alzheimer's disease with real-world clinical validation. *npj Digital Medicine*, 7, 46. <https://doi.org/10.1038/s41746-024-01038-3>.
 - Pinzi, L., & Rastelli, G. (2019). Molecular docking: Shifting paradigms in drug discovery. *International Journal of Molecular Sciences*, 20(18), 4339.
 - Pirmohamed, M. (2023). Pharmacogenomics: Current status and future perspectives. *Nature Reviews Genetics*, 24(6), 350–362. <https://doi.org/10.1038/s41576-022-00572-8>
-
-

-
-
- Poplin, R., Chang, P.-C., Alexander, D., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36, 983–987. <https://doi.org/10.1038/nbt.4235>.
 - Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36, 983–987. <https://doi.org/10.1038/nbt.4235>
 - Powell, D. (2024). Walk, talk, think, see and feel: Harnessing the power of digital biomarkers in healthcare. *npj Digital Medicine*, 7, 45.
 - Precision agriculture in the age of AI: A systematic review of machine learning methods for crop disease detection. (2025). *Smart Agricultural Technology*, 12, 101491.
 - Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, 28, 31–38. <https://doi.org/10.1038/s41591-021-01614-0>
 - Rapp, J. T., Bremer, B. J., & Romero, P. A. (2024). Self-driving laboratories to autonomously navigate the protein fitness landscape. *Nature Chemical Engineering*, 1, 97–107. <https://doi.org/10.1038/s44286-023-00002-4>
 - Ream, M., & Prather, K. L. J. (2024). Engineered autonomous dynamic regulation of metabolic flux. *Nature Reviews Bioengineering*, 2, 233–243.
 - Reardon, B., Culhane, A. C., & Van Allen, E. M. (2026). Convergence of machine learning and genomics for precision oncology. *Nature Reviews Cancer*, 26, 217–229.
 - Remote sensing and artificial intelligence for crop pest and disease monitoring: A systematic review of sensing technologies, AI approaches, and barriers to field adoption. (2026). *Smart Agricultural Technology*, 14, 102279.
 - Richardson, J. S., Read, R. J., et al. (2024). AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nature Methods*, 21, 110–116.
 - Rodríguez, A., Kamarthi, H., Agarwal, P., Ho, J., Patel, M., Sapre, S., & Prakash, B. A. (2024). Machine learning for data-centric epidemic forecasting. *Nature Machine Intelligence*, 6, 1122–1131.
-
-

-
-
- Sadee, W., Wang, D., Hartmann, K., et al. (2023). Pharmacogenomics: Driving personalized medicine. *Pharmacological Reviews*, 75(4), 789–814. <https://doi.org/10.1124/pharmrev.122.000810>
 - Sadybekov, A. V., & Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature*, 616, 673–685. <https://doi.org/10.1038/s41586-023-05905-z>
 - Saeidi, H., et al. (2022). A systematic review of robotic surgery: From supervised paradigms to fully autonomous robotic approaches. *International Journal of Medical Robotics and Computer Assisted Surgery*.
 - Schirone, M., D'Ambrosio, G., & Paparella, A. (2026). Artificial intelligence in foodborne pathogen detection from sensing to food safety systems: A systematic review. *Foods*, 15(14), 2562. <https://doi.org/10.3390/foods15142562>
 - Schoenherr, J. R., et al. (2024). Ethical considerations in the design and conduct of clinical trials of artificial intelligence. *JAMA Network Open*.
 - Shajari, S., Kuruvinashetti, K., Komeili, A., & Sundararaj, U. (2023). The emergence of AI-based wearable sensors for digital health technology: A review. *Sensors*, 23(23), 9498. <https://doi.org/10.3390/s23239498>.
 - Shen, M.-d., Chen, S.-b., & Ding, X.-d. (2024). The effectiveness of digital twins in promoting precision health across the entire population: A systematic review. *npj Digital Medicine*, 7, 145. <https://doi.org/10.1038/s41746-024-01146-0>.
 - Shukla, K., Nagpal, S., & Vyas, A. (2026). Opportunities of digital twin in controlling and monitoring microbial processes in food fermentation. *Systems Microbiology and Biomanufacturing*, 6(3), 56. <https://doi.org/10.1007/s43393-026-00454-8>
 - Silva, M. P. da, Correa, S. P. L. P., Schaefer, M. A. R., Reis, J. C. S., Nunes, I. M., dos Santos, J. A., & Oliveira, H. N. (2025). Advancing agricultural remote sensing: A comprehensive review of deep supervised and self-supervised learning for crop monitoring. *Pattern Recognition Letters*, 133, 104434.
 - Smaldone, A. M., Shee, Y., Kyro, G. W., Xu, C., Vu, N. P., Dutta, R., Farag, M. H., Galda, A., Kumar, S., Kyoseva, E., & Batista, V. S. (2025). Quantum machine learning in drug discovery: Applications in academia and pharmaceutical industries. *Chemical Reviews*, 125(12), 5436–5460. <https://doi.org/10.1021/acs.chemrev.4c00678>
-
-

-
-
- Su, J., Xin, C., Shang, A., Wu, S., Xie, Z., Xiong, R., Xu, X., Zhang, C., Chen, G., Chan, Y.-T., Tang, G., Wang, N., Xu, Y., & Feng, Y. (2026). Artificial intelligence in drug discovery and development: Current landscape, challenges, and future perspectives. *Critical Reviews in Oncology/Hematology*. <https://doi.org/10.1016/j.critrevonc.2026.105461>.
 - Suay-García, B., & Falcó, A. (2026). Topological deep learning for drug-target interaction, virtual screening, and docking scoring: A practical, benchmark-driven review. *Briefings in Bioinformatics*, 27(4), bbag370. <https://doi.org/10.1093/bib/bbag370>
 - Sultana, H., et al. (2026). AI in genomics: From variant calling to multi-omics integration. *BioEssays*. <https://doi.org/10.1002/bies.70160>.
 - Sultana, H., et al. (2026). AI in genomics: From variant calling to multi-omics integration. *BioEssays*. <https://doi.org/10.1002/bies.70160>
 - Tabassi, E. (2023). Artificial intelligence risk management framework (AI RMF 1.0). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
 - Taha, M. O., & Daoud, S. (2026). Innovative integration of molecular docking and machine learning for drug discovery: From virtual screening to nanomolar inhibitors. *Chemical Communications*. <https://doi.org/10.1039/D5CC06025G>
 - Tang, C., Yi, W., Occhipinti, E., et al. (2024). A roadmap for the development of human body digital twins. *Nature Reviews Electrical Engineering*, 1, 199–207.
 - Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25, 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
 - Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25, 44–56. <https://doi.org/10.1038/s41591-018-0300-7>.
 - Trends in machine and deep learning techniques for plant disease identification: A systematic review. (2024). *Agriculture*, 14(12), 2188.
 - Tropsha, A., Isayev, O., Varnek, A., Schneider, G., & Cherkasov, A. (2024). Integrating QSAR modelling and deep learning in drug discovery: The emergence of deep QSAR. *Nature Reviews Drug Discovery*, 23, 141–155. <https://doi.org/10.1038/s41573-023-00832-0>
-
-

-
-
- Tuia, D., Schindler, K., Demir, B., et al. (2023). Artificial intelligence to advance Earth observation: A review of models, recent trends, and pathways forward. *IEEE Geoscience and Remote Sensing Magazine*.
 - U.S. Food and Drug Administration & Clinical Trials Transformation Initiative. (2025). Leveraging artificial intelligence in drug and biological product development: An FDA and Clinical Trial Transformation Initiative workshop report. *Clinical and Translational Science*. <https://doi.org/10.1056/aipc2500801>.
 - U.S. Food and Drug Administration (FDA). (2026). Artificial intelligence-enabled medical devices. U.S. Department of Health and Human Services.
 - U.S. Food and Drug Administration. (2021). Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device Action Plan. U.S. Department of Health and Human Services.
 - U.S. Food and Drug Administration. (2025a). Artificial Intelligence Program: Research on AI/ML-based medical devices.
 - U.S. Food and Drug Administration. (2025b). FDA issues comprehensive draft guidance for developers of artificial intelligence-enabled medical devices.
 - UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. United Nations Educational, Scientific and Cultural Organization.
 - Wang, J., Zhang, S., Lizaga, I., et al. (2024). UAS-based remote sensing for agricultural monitoring: Current status and perspectives. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2024.109501>
 - Wearables research for continuous monitoring of patient outcomes: A scoping review. (2025).
 - Wei, H., Xu, W., Kang, B., et al. (2024). Irrigation with artificial intelligence: Problems, premises, promises. *Human-Centric Intelligent Systems*, 4, 187–205.
 - Whalen, S., Schreiber, J., Noble, W. S., & Pollard, K. S. (2022). Navigating the pitfalls of applying machine learning in genomics. *Nature Reviews Genetics*, 23, 169–181. <https://doi.org/10.1038/s41576-021-00434-9>.
 - World Health Organization (WHO). (2021). Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization.
-
-

-
-
- World Health Organization. (2021). Ethics and Governance of Artificial Intelligence for Health: WHO Guidance-Executive Summary. World Health Organization.
 - World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. <https://www.who.int/publications/i/item/9789240029200>.
 - World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization.
 - World Health Organization. (2021). Global strategy on digital health 2020–2025. Geneva: WHO.
 - World Health Organization. (2024). Artificial Intelligence for Health: Supporting countries to deploy responsible AI technologies to accelerate equitable health for all. Geneva: WHO.
 - World Health Organization. (2024). Artificial Intelligence for Health: Supporting countries to deploy responsible AI technologies to accelerate equitable health for all.
 - World Health Organization. (2024). Ethics and governance of artificial intelligence for health: Large multi-modal models. Geneva: World Health Organization.
 - World Health Organization. (2024). Guidance for human genome data collection, access, use and sharing. World Health Organization.
 - World Health Organization. (2024). Harnessing artificial intelligence for health. World Health Organization.
 - World Health Organization. (2024). Research prioritization for pandemic and epidemic intelligence: Technical brief. World Health Organization.
 - World Health Organization. (2025). Global strategy on digital health 2020–2027. Geneva: WHO.
 - World Health Organization. (2025a). Global Strategy on Digital Health 2020–2027.
 - World Health Organization. (2025b). The World Health Assembly endorses the extension of the Global Strategy on Digital Health to 2027 and approves the next phase for 2028–2033.
 - World Health Organization. (2026). Genomics. World Health Organization.
-
-

-
-
- Zhang, et al. (2024). Integration of remote sensing and machine learning for precision agriculture: A comprehensive perspective on applications. *Agronomy*, 14(9), 1975. <https://doi.org/10.3390/agronomy14091975>.
 - Zhang, K., Yang, X., Wang, Y., et al. (2025). Artificial intelligence in drug development. *Nature Medicine*, 31, 45–59. <https://doi.org/10.1038/s41591-024-03434-4>.
 - Zhao, H. (2026). Artificial intelligence tools for enzyme engineering and metabolic engineering. *Current Opinion in Biotechnology*, 100, 103522. <https://doi.org/10.1016/j.copbio.2026.103522>
 - Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., & Hoffman, M. M. (2018). Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, 50, 71–91. <https://doi.org/10.1016/j.inffus.2018.09.012>
 - Zou, J., Huss, M., Abid, A., et al. (2019). A primer on deep learning in genomics. *Nature Genetics*, 51, 12–18. <https://doi.org/10.1038/s41588-018-0295-5>.
 - Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2019). A primer on deep learning in genomics. *Nature Genetics*, 51, 12–18. <https://doi.org/10.1038/s41588-018-0295-5>
 - Zou, X., et al. (2026). Docking-based virtual screening: Past, present, and future. *Biophysical Journal*. <https://doi.org/10.1016/j.bpj.2026.04.011>

ABOUT THE AUTHORS



Dr. Prabhavati

Assistant Professor

Department of Zoology, K.L.E. Society's S. Nijalingappa College, Rajajinagar, Bengaluru, Karnataka, India



Dr. Soumya M. Hegde

Assistant Professor

Department of Botany, KLE Society's S. Nijalingappa College, Rajajinagar, Bengaluru, Karnataka, India



Dr. Jagadevi Shivaputrappa

Guest Lecturer

Department of Botany, Gulbarga University, Kalaburagi, Karnataka, India



Dr. Mahesh Vilas Tanwade

Assistant Professor

Department of Postgraduate Studies and Research in Zoology, Sharnbasva University, Kalaburagi, Karnataka, India

ABOUT THE BOOK

The AI Revolution in Life Sciences: Technologies, Applications, and Future Directions explores the transformative role of Artificial Intelligence (AI) in life-science research, healthcare, biotechnology, agriculture, and biological engineering. The book provides a concise and accessible overview of AI technologies, applications, opportunities, and challenges shaping this rapidly evolving field.

It introduces key concepts such as machine learning, deep learning, natural language processing, computer vision, predictive analytics, and generative AI, highlighting their ability to process complex datasets, identify patterns, improve prediction, and support scientific decision-making.

The book examines important applications of AI in drug discovery, disease diagnosis, precision medicine, genomics, biotechnology, bioengineering, and precision agriculture. These applications demonstrate how AI can accelerate research, improve healthcare outcomes, enhance biological analysis, and support more efficient and sustainable systems.

It also addresses responsible AI adoption through discussions of data privacy, cybersecurity, algorithmic bias, transparency, explainability, ethics, and regulatory challenges. Looking ahead, the book explores emerging areas such as generative AI, autonomous laboratories, digital biology, AI-driven biological discovery, and AI-quantum computing integration.

Designed for students, researchers, academicians, healthcare professionals, biotechnology specialists, and life-science practitioners, the book provides a foundation for understanding both the current capabilities and future possibilities of AI. Ultimately, the book presents AI as a powerful enabling force capable of reshaping scientific discovery, healthcare, biotechnology, agriculture, and sustainable development, while emphasising the ethical and practical responsibilities that accompany its adoption.



India | UAE | Nigeria | Uzbekistan | Montenegro | Iraq | Egypt | Thailand | Uganda | Philippines |
Indonesia

Empyreal Publishing House || www.empyrealpublishinghouse.com || info@empyrealpublishinghouse.com