iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: II Month of publication: February 2024

DOI: https://doi.org/10.22214/ijraset.2024.58430

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

Decoding Tomorrow's Biotechnology: AI and Machine Learning Unveiled

Vijay P. Thikare[1], Prutha S. Pathak[2]
Trichy Institute of Biotechnology, Tamil Nadu

Abstract: This review paper explores the impact of AI and ML on biotechnology, focusing on drug discovery, protein prediction, automation, and ethics.AI and ML are transforming drug discovery by automating tasks, identifying new drug targets, and designing more effective drugs. AI-powered languages and functional genomics are being used to accelerate drug discovery processes. AI tools such as AlphaFold are being used to predict protein structures and functions. Biotechnology automation is being used to improve efficiency in laboratory settings. The use of AI and ML in biotechnology raises ethical and regulatory concerns. It is important to find the potential risks and benefits of this technology before it is widely adopted.

I. INTRODUCTION

In 1955, the term "Artificial Intelligence" first appeared in print. The power of Artificial Intelligence (AI) is making a significant impact in biotechnology, effectively addressing a diverse array of challenges. These problems include things like Drug Discovery, Drug Safety, functional and structural genomics, proteomics, metabolomics, pharmacology, pharmacogenetics, and pharmacogenomics Fig1. [1] AI is involved in all these areas and more, making a big difference in biotechnology. The pace of innovation in the biotechnology industry is quickening, and biotechnology companies are beginning to recognize the value that AI can bring to their entire business, in the form of accelerated R&D, analysis of humongous databases, effective decision-making, and cost-effectiveness. Biotechnology has long been at the forefront of scientific breakthroughs, from genetic engineering to personalized medicine. However, the complex nature of biological systems and the vast amount of data generated present significant challenges for traditional analytical approaches. This is where AI and ML step in, offering powerful tools to extract valuable insights and patterns from these complex datasets. Over the past decade, Artificial Intelligence (AI) has contributed substantially to the resolution of various medical problems, including cancer [2]. AI's development and application relies on digital technology, especially digital computers. Digital transformation refers to using digital tech to fundamentally change how entities like companies, research institutions, and universities operate [1]. In biotech, this means introducing new tech and methods to enhance research speed, accuracy, and product development. This transformation speeds up AI integration by offering big data access and automating tasks, boosting efficiency and precision.

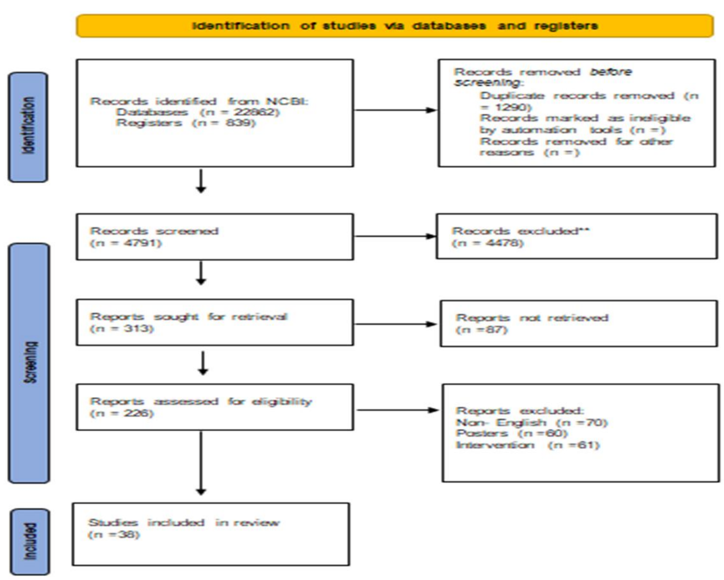


Fig 1. Prisma for AI and Machine Learning

In the world of biology, AI has changed a lot. Initially, AI used a method called "symbolic," where computers followed complex rules to do tasks. A famous example is chess, a game with simple rules but many possible moves. Computers could study all the options and choose the best move. One big success was Deep Blue, a computer made by IBM. It beat the top chess player in 1997. Before Deep Blue, computers couldn't do calculations as fast as humans. Today's smartphones are as fast as Deep Blue. This progress in technology is important in biotechnology, especially in areas like medicine and research [3].

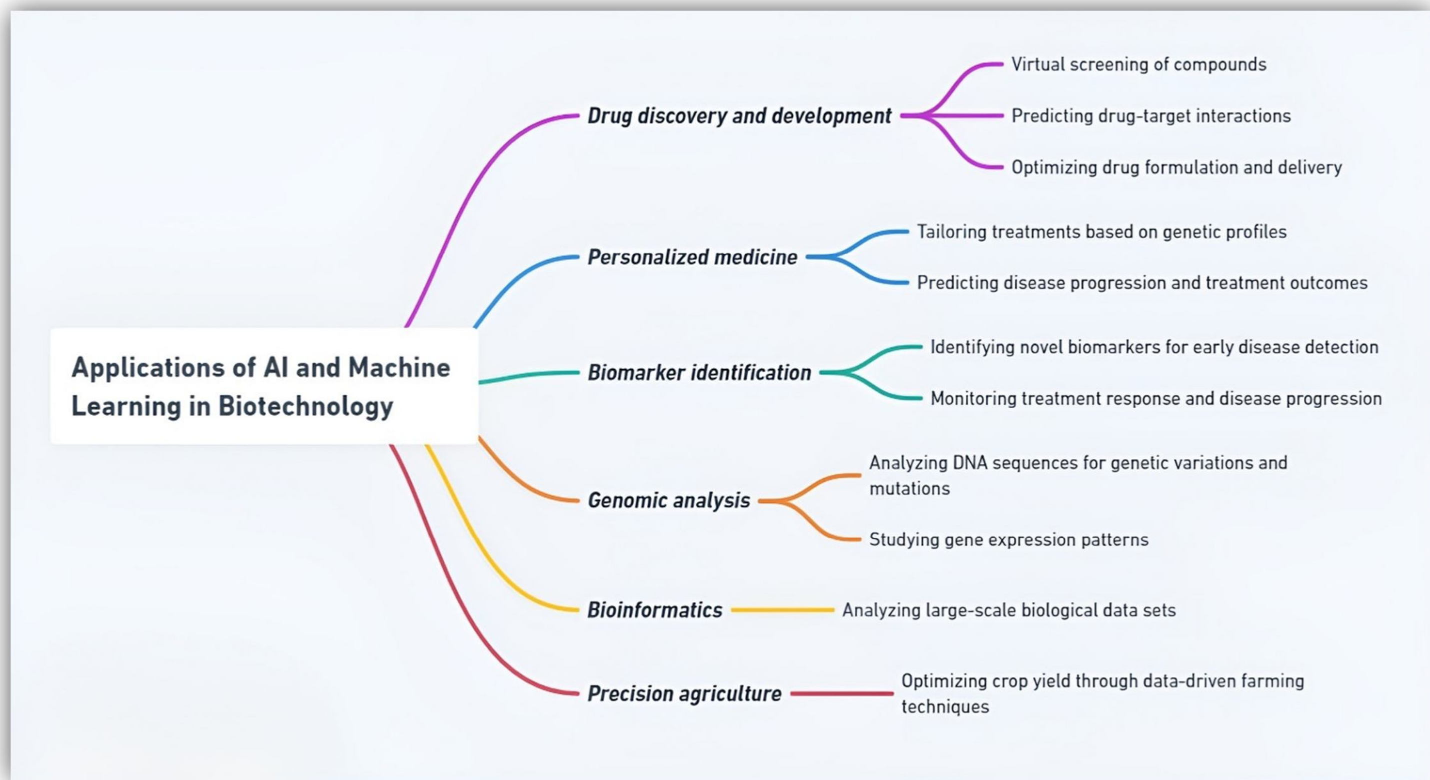


Fig.2. Application of AI and Machine Learning in Biotechnology

II. OVERVIEW OF AI AND ML

Artificial intelligence (AI) is currently a hot topic due to its widespread popularity (such as ChatGPT). When biotechnology and AI evolve together, unheard-of new potential solutions become available [1]. This can aid in solving numerous global issues and advance significant Sustainable Development Goals. Food security, health and well-being, clean water, clean energy, responsible consumption and production, combating climate change, life below the ocean, protecting, restoring, and promoting sustainable use of terrestrial ecosystems, sustainably managing forests, preventing desertification, halting and reversing land degradation, and halting biodiversity loss are some examples of current priorities. Today, AI is pervasive in the biological sciences [4].

Machine learning (ML) refers to the part of research on AI that seeks to provide knowledge to computers through data and observations without being explicitly programmed [5]. Machine learning has been utilized in biology for a number of decades, but its significance has continuously increased to the point that it is now utilized in almost all biological disciplines. Machine learning encompasses the practice of constructing predictive models from data and uncovering meaningful clusters within datasets. It strives to replicate human pattern recognition, but in an unbiased manner, through computational methods. This field becomes invaluable when dealing with extensive datasets or intricate data with numerous attributes, circumstances where human examination is impractical. Additionally, it serves the purpose of automating data analysis to create a consistent and efficient workflow [3]. Machine learning algorithms frequently employed in practice include linear and logistic regression, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and tree-based methods. These individual models can then be combined with one another using ensemble learning, a methodology which leverages the power of multiple weak classifiers to achieve optimal overall performance [6].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

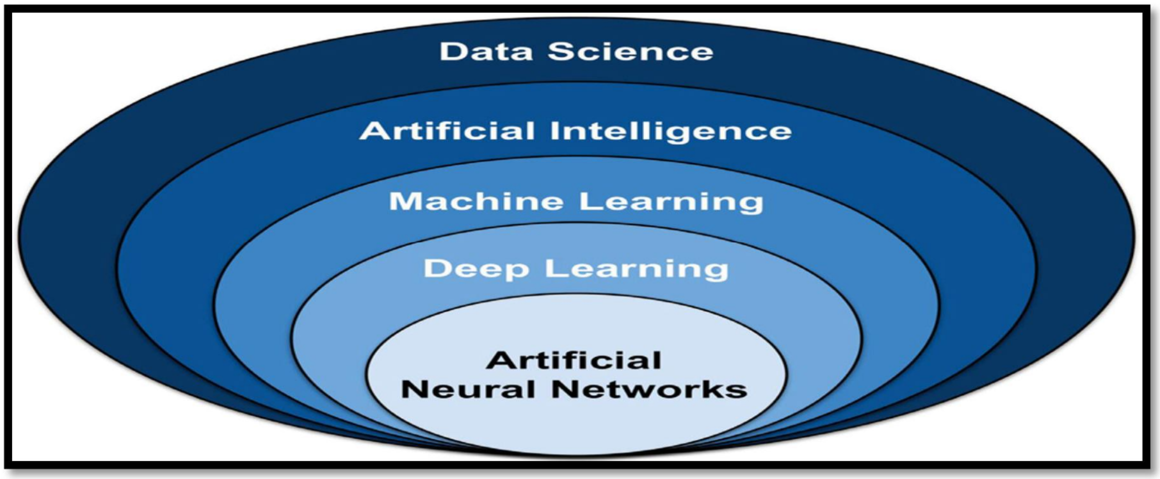


Fig 3. Umbrella of select data science techniques. Artificial intelligence (AI) falls within the realm of data science, and includes classical programming and machine learning (ML). ML contains many models and methods, including deep learning (DL) and artificial neural networks (ANN) [12].

A. Artificial Neural Network

An Artificial Neural Network (ANN) is a type of machine learning tool inspired by how our brains work. Each ANN has nodes that act like cells and talk to each other through connections, similar to how neurons communicate through axons and dendrites. Just like in real brains, when certain neurons often work together, their connections get stronger. In ANN, these connections between nodes are given weights based on how well they help achieve a specific goal. This is important for machine learning and deep learning [6 7 8]. Artificial Neural Network (ANN) that has many layers, often more than 6. This is like a complex version of Machine Learning (ML), and it's really good at studying data closely, using a lot of information, and understanding complicated ideas. Imagine each piece of information is like a note that one friend tells another friend, and they all talk to each other. The ANN's job is to be really smart about figuring out the right answers, kind of like a super-smart version of a game where you try to match things with a reference or example. Each node gets data from other nodes, and those nodes' outputs are weighted **[Fig.1].** It does this by adjusting how much each piece of information matters, depending on how wrong or right it was in its previous guesses [6 9].The ANN tries to get as many correct answers as possible by changing the weights on its nodes. It does this by looking at how wrong its answers are during each step and adjusting things as it goes forward [9 10].

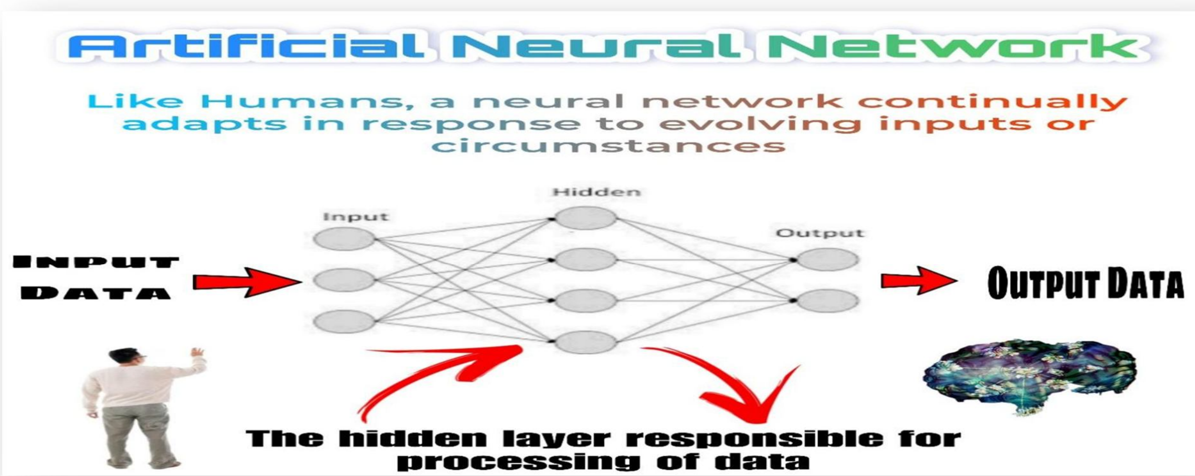


Fig.4. Artificial Neural Network.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

B. Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a type of artificial neural network that solves a problem in image understanding [6]. It's like a special version of the brain-inspired network. Instead of looking at one pixel at a time, a CNN looks at small sections of the image, like puzzle pieces. These pieces are passed on to another part of the network that understands shapes and patterns. This helps the network understand what's in the picture. The small sections help the network learn about specific things, and we call them "convolutional filters"[11 8].

CNNs have gained significant importance in the field of medical image analysis. Specifically, they are highly relevant for tasks like image segmentation, which involves accurately identifying and isolating organs within images – examples include lungs, brain, and bones. Furthermore, CNNs prove valuable in detecting abnormalities like tumors within these organs [6 10]. A notable drawback related to CNNs is demonstrated by adversarial examples, where CNNs mistakenly classify an entirely unrecognizable image as a recognizable object with a high level of confidence. An instance of this is identifying pixelated noise as an animal [12]. To address this concern, a type of CNNs based on unsupervised learning, known as generative adversarial networks (GANs), has been introduced. GANs continue to be an active area of research for enhancing data augmentation. Another recent breakthrough aimed at mitigating such challenges is the emergence of Capsule Networks (CapsNet). Unlike traditional CNNs that use deep layers, CapsNet employs capsules within its CNN layers [6].

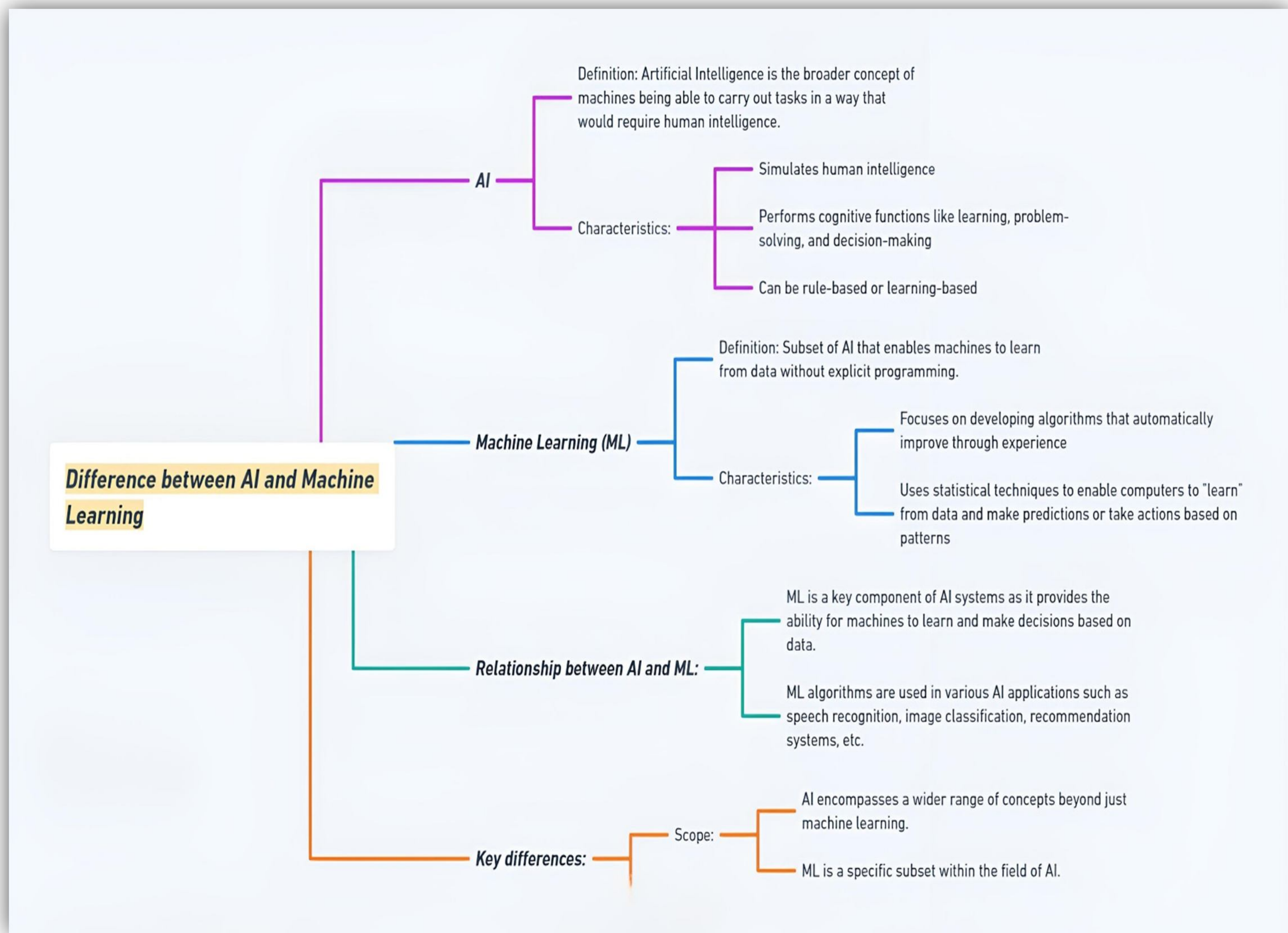


Fig.5. Difference between AI and Machine Learning

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

III. HISTORY OF AI AND MACHINE LEARNING IN BIOTECHNOLOGY

Table 1. Some important Discoveries in field of Biotechnology and significant role AI and Machine Learning is given in these table

Year	Discovery	References
1952	Arthur Samuel develops a program to play checkers, which is the first to ever learn the game independently [13]	(Samuel, 1959)
1959	Marvin Minsky and Seymour Papert publish "Perceptrons", a book that lays the foundation for neural networks.[14]	(Arbib, 1969)
1965	Joshua Lederberg and Esther Lederberg publish a paper on the use of computers for DNA sequencing [15]	(Lederberg and Lederberg, 1952)
1972	Edward Feigenbaum and Pamela McCorduck publish "The Fifth Generation Computer", a book that predicts the future of AI [16]	(Morgan et al., 1983)
1982	David Rumelhart, Geoffrey Hinton, and Ronald Williams publish a paper on the back propagation algorithm, which makes neural networks more powerful and efficient [17]	(Rumelhart et al., 1986)
1990	Craig Venter and colleagues sequence the first genome, that of the bacterium Haemophilus influenza [18]	(Fleischmann et al., 1995)
2000	Andrew Ng and Michael I. Jordan publish a paper on support vector machines, which are now widely used in machine learning [19]	(Sutskever et al., 2013)
2001	Craig Venter and colleagues sequence the first human genome [20].	(Venter et al., 2001)
2010-20	Deep Learning and Drug Discovery (2010s-2020s [21]	(Ching et al., 2018)
2012	Geoffrey Hinton and colleagues publish a paper on deep learning, which has revolutionized the field of machine learning [22].	(Hinton et al., 2012)
2020	The COVID-19 pandemic accelerates the development and adoption of AI in biotechnology, with applications such as drug discovery, vaccine development, and contact tracing [23]	(Mondal et al., 2021)
2023	Advancement in AI and ML in every field [1]	(Holzinger et al., 2023)

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

IV. ROLE OF AI AND ML IN DRUG DISCOVERY AND DEVELOPMENT

The pursuit of controlling and adapting to the ever-changing facets of life is a fundamental human goal, particularly in the realm of medicine and pharmaceuticals. These fields center on developing or uncovering chemical compounds and mixtures to alleviate physical and psychological suffering. Historically, drug production adhered to a regulatory framework that ensures the quality of final products through rigorous testing of raw materials, in-process materials, end-product characteristics, and batch-based operations under fixed process conditions [22].

However, the pharmaceutical and biopharmaceutical sectors have not been prolific sources of inventive technologies or machinery, lagging in the evolution of principles within chemical and mechanical engineering. The pharmaceutical industry urgently requires mechanical innovations to streamline the creation of safe medications for human consumption on a large scale [22].

The complexity of developing and manufacturing these medications and integrating them into mainstream therapeutic applications has posed significant challenges, primarily due to limitations in technological resources.AI-based approaches have demonstrated the ability to predict the toxicity of potential drug candidates, underscoring their capacity to enhance the efficiency and effectiveness of drug discovery. Nevertheless, the incorporation of AI in the development of new bioactive compounds presents challenges and limitations that must be addressed. Ethical considerations must be carefully weighed, and further research is essential to comprehensively grasp both the benefits and constraints of AI in this domain [22, 24].

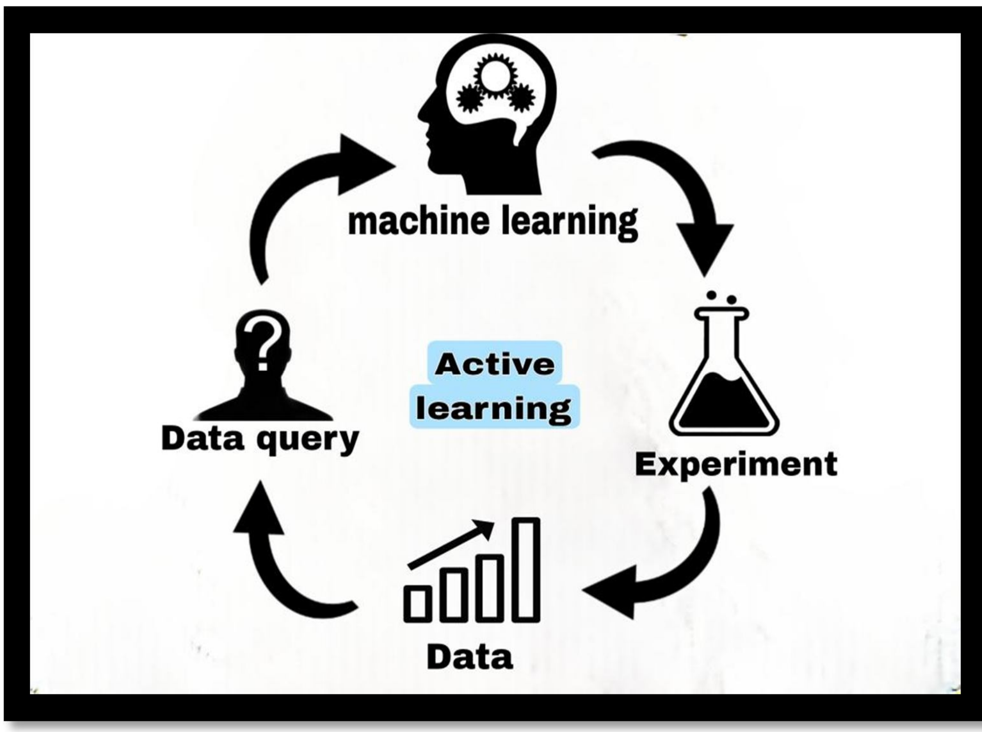


Fig.6 Drug discovery and development

A. AI-powered Language Models to Accelerate Target Identification

AI-powered language models play a pivotal role in advancing biomedical research. They enhance target identification and facilitate the discovery of relationships among chemicals, genes, targets, and diseases through automatic biomedical named entity recognition (BioNER). Innovative approaches, such as multiple-task learning with BioBERT, surpass conventional methods like BiLSTM, CRF, and MTM-CW in recognizing entities related to chemicals, diseases, and genes [25, 26]

Moreover, AI-powered models aid in biomedical association extraction from PubMed literature. Deep-GDAE, a hybrid transfer learning framework, integrates BiLSTM and CNN, achieving impressive results in gene-disease relationship extraction, with a high F-measure of 79.8%.Another valuable application is biomedical text summarization. By employing BERT models, researchers can generate synthetic abstracts from full PubMed articles, achieving state-of-the-art results. Additionally, incorporating domain-specific contextual embeddings from BioBERT can further enhance performance, offering a promising avenue for accelerating target identification in research papers [25, 22].

Table 2. -Selected examples AI-based NLP applied in drug discovery [30]

NLP task	Data set	Algorithm	Notes
Biomedical term recognition	BC2GM; BC5CD; NCBI -Disease; JNLPBA	BioBERT	A Multi-Task approach for Biomedical Named Entity Recognition has been devised, leveraging BioBERT as the common backbone while employing distinct datasets in task-specific layer
Extraction of associations between genes and diseases.	DisGeNET: database of gene–disease associations	CNN and BiLSTM	The Deep-GDAE model, which has been proposed, combines the strengths of CNN and an attention-based BiLSTM to categorize associations between genes and diseases."
Text summarization of Biomedical	relationship between Gene–disease extraction PubMed	BERT and hierarchical clustering algorithm	A biomedical text summarization model, known as BERT-based-Sum, has been introduced. It utilizes BERT and a hierarchical clustering algorithm to generate concise summaries of biomedical content
Prediction of Drug Properties	There are one million SMILE codes representing compounds in the ChEMBL database	BiLSTM	A transfer learning framework called MolPMoFiT has been introduced for the prediction of various physical and biological endpoints, including but not limited to lipophilicity and blood-brain barrier penetration, for a given set of compounds

B. Challenges and Limitation

While AI holds significant promise in drug discovery, it brings forth several challenges and limitations that warrant consideration. Among these challenges, the availability of suitable data stands out as a primary concern. AI-driven approaches typically necessitate a substantial volume of data for effective training[24]. In numerous instances, the accessible data may be limited in quantity or marred by low quality and inconsistency, which can adversely impact result accuracy and reliability. Another formidable challenge arises from ethical considerations associated with AI applications, which may give rise to concerns regarding fairness and bias, as discussed in the subsequent section. For instance, if the data used to train machine learning algorithms exhibit bias or fail to represent the entire spectrum, the resulting predictions may be inaccurate or unjust. Ensuring ethical and equitable deployment of AI in the development of new therapeutic compounds emerges as a crucial imperative [27]

Several strategies and approaches can be deployed to surmount the hurdles confronted by AI in the realm of chemical medicine. One such approach is data augmentation, involving the generation of synthetic data to supplement existing datasets. This serves to augment the quantity and diversity of training data for ML algorithms, thereby enhancing result accuracy and dependability. Another strategy is the adoption of explainable AI (XAI) methods, aiming to furnish interpretable and transparent rationales for the predictions made by ML algorithms. This approach aids in addressing concerns related to bias and fairness in AI-driven methodologies while promoting a deeper comprehension of the underlying mechanisms and assumptions behind these predictions [24]. It is essential to recognize that current AI-based approaches cannot substitute traditional experimental methods and the expertise of human researchers. AI can provide predictions based on available data, but the subsequent validation and interpretation of results necessitate human intervention. Nevertheless, the synergy between AI and traditional experimental methods has the potential to enhance the drug discovery process. By harnessing the predictive capabilities of AI in conjunction with the expertise and experience of human researchers, it becomes possible to optimize drug discovery and expedite the development of new pharmaceuticals [24 27].

C. Functional Genomics in AI

In recent decades, scientific development has shifted from evaluating individual genes or a small set of genes to investigating thousands of genes at the same time. Various scientific disciplines contribute to the generation of extensive biological data, each named based on its primary focus.

These disciplines encompass the study of an organism or system's DNA information content (genomics), the reversible modifications that DNA can undergo (epigenetics), the RNA transcripts produced from a genome (transcriptomics), the collection and dynamics of RNA modifications (epitranscriptomics), the translational products derived from protein-coding transcripts (proteomics), and the metabolites found in a specific organism or system under various physiological and pathological conditions at a given time [28].

In 1981, the inaugural full genome sequence of a eukaryotic organelle, specifically the human mitochondrion spanning 16.6 kilobases in length, was successfully deciphered. This evolution marks a shift from examining discrete units of inheritance to investigating an organism's complete genome. The field of genomics, originally focused on deciphering DNA sequences (the specific order of nucleotides within a DNA fragment), has swiftly broadened its scope. It now encompasses a more functional dimension, delving into the study of gene and protein expression profiles as well as their respective functions [29].

Machine Learning (ML) has gained extensive traction across various "omics" disciplines, particularly in fields marked by the generation of substantial datasets and intricate processes influenced by the collaborative engagement of diverse factors. Key applications encompass the prediction of DNA regulatory segments, elucidation of cellular morphology and spatial arrangements, discernment of links between phenotypic traits and genotypic makeup, categorization of DNA methylation and histone modifications, revelation of biomarkers, detection of transcriptional enhancers, facilitation of cancer diagnosis, and exploration of evolutionary mechanisms [28 29]. **(Fig.5)**

In the contemporary landscape of biological research, marked by remarkable advancements in sequencing technologies and our deepening understanding of the intricate nature of DNA, this notion has been broadened to encompass the entire array of DNA sequences within a cell or organism. This expansion accounts for factors like the number of copies of the fundamental set of chromosomes, known as ploidy, and incorporates the genetic material derived from extra nuclear organelles like mitochondria [28].

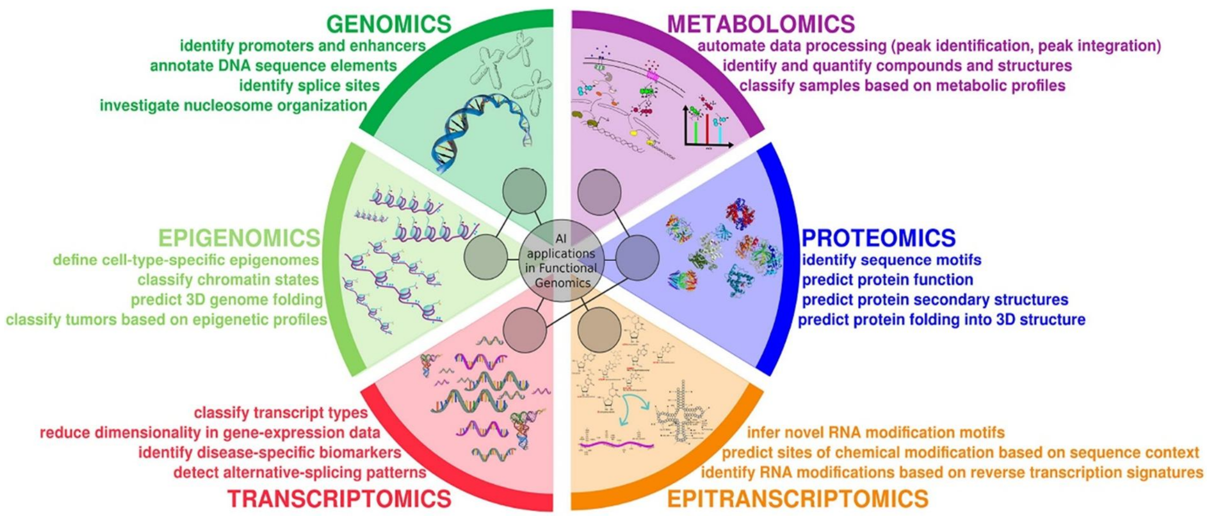


Fig.7. AI applications in functional genomics [28]

Functional genomics is the scientific field that examines, on a genome-wide scale, the interrelationships among the elements within a biological system, including genes, transcripts, proteins, metabolites, and more. It seeks to understand how these elements collaborate to generate specific phenotypic outcomes. The term "functional genomics" gained prominence within the scientific community during the emergence of the initial genome sequencing projects. These endeavours aimed to decipher the complete genome sequences of various organisms and to annotate functionally significant elements within them, such as protein-coding and non-coding genes, as well as DNA regulatory regions. A pivotal example of such an undertaking is the Human Genome Project (HGP), a global collaborative initiative initiated in 1990 and officially concluded in 2003 by the International Human Genome Sequencing Consortium. [30].

The emergence of the CRISPR/Cas9 system as a genome editing toolkit represented a profound paradigm shift, revolutionizing the precise manipulation of genomic elements. Originally observed within prokaryotic genomes as a component of their antiviral defence mechanism against bacteriophages, the CRISPR/Cas9 system was subsequently acknowledged as a groundbreaking genome editing tool that facilitates the precise addition, removal, and modification of genes with exceptional accuracy and specificity [31].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

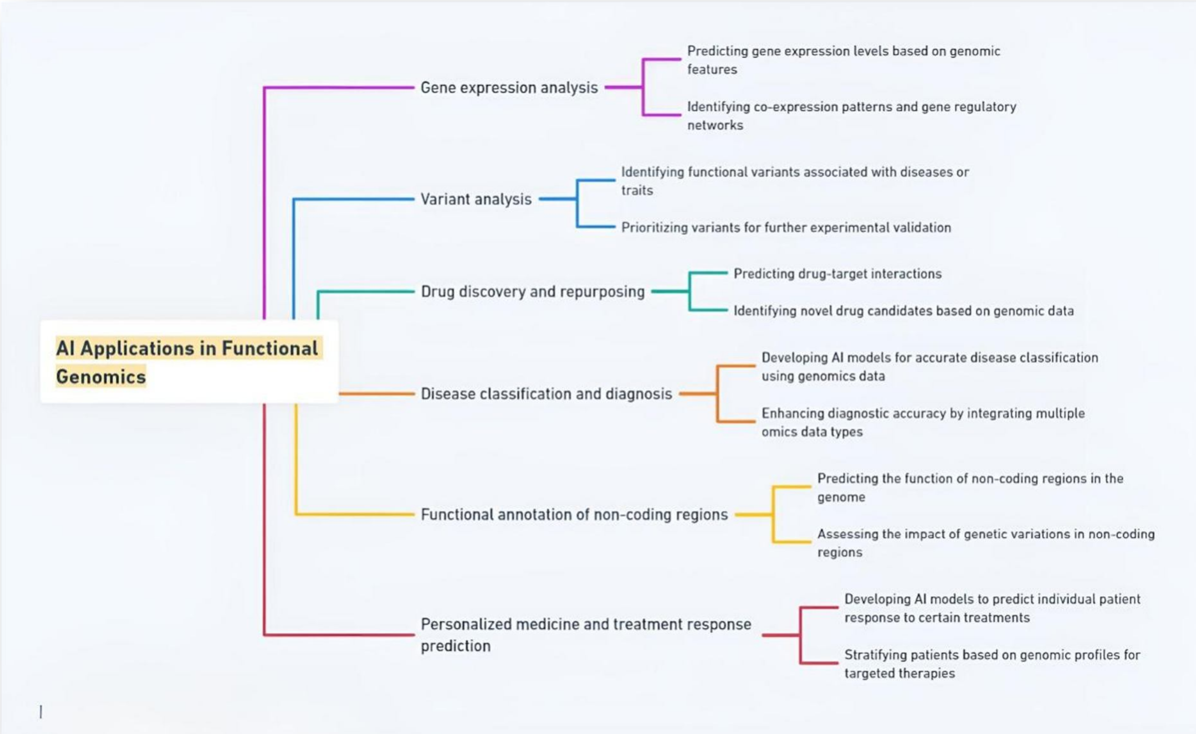


Fig.8. AI application in functional genomics

One of the most recent nuclease systems, rooted in the bacterial adaptive CRISPR-Cas9 immune mechanism, relies on a single guide RNA (sgRNA) composed of 100 nucleotides. This sgRNA guides the Cas9 endonuclease, forming the Cas9-sgRNA ribonucleoprotein (RNP) complex, to a specific genomic location. The binding of the sgRNA to the target DNA triggers Cas9 to cleave the chromosomal target, the swiftness, simplicity, and cost-effectiveness of sgRNA design and engineering, coupled with the high efficiency and precision of Cas9, have made it the favoured nuclease platform. As a result, it has significantly expanded the utility of genome editing, enabling precise genetic studies and advancing preclinical gene correction therapies. [31]

Table 3. Important AI tools and its Applications

Sr. No.	AI Tools	Year	Application
1.	Schrodinger’s Suite	1990	Utilizing AI and molecular modeling, this approach is employed for drug design, virtual screening, and the prediction of protein-ligand interactions.
2.	Acellera	2006	This method provides AI-enhanced simulations for in-depth exploration of both protein-ligand interactions and the dynamics of protein folding in the context of [specify the field, e.g., structural biology, computational chemistry].
3.	Numerate	2007	Leveraging AI-driven molecular design expedites drug discovery by creating and refining compound libraries.
4.	Atomwise	2012	Virtual screening of small molecules to identify potential drug candidates.

5.	IBM Watson	2010	Utilizing AI and NLP, it extracts insights from biomedical sources, assisting in target identification and drug repurposing.
6.	Recursion Pharmaceuticals	2012	Integrating AI with high-throughput biology uncovers new drug candidates and potential targets.
7.	Exscientia	2012	Combining AI, including machine learning and evolutionary algorithms, for small molecule design and drug development optimization.
8.	BenevolentA	2013	AI-powered hypothesis generation for drug targets, biomarkers, and potential therapies.
9.	Recursion Pharmaceuticals	2013	Integrates AI with high-throughput biology to discover new drug candidates and potential targets.
10.	cyclica	2013	Applies AI and computational biophysics in drug discovery for studying drug-target interactions and analyzing polypharmacology.
11.	Deepchem	2016	Utilizing AI and NLP, it extracts insights from biomedical sources, assisting in target identification and drug repurposing.
12.	Artoris	2016	Integrates AI and lab automation to optimize drug discovery workflows, spanning target validation to compound testing.
13.	Nebula Genomics	2017	Applies AI to assess genomic and clinical data, pinpointing potential biomarkers and therapeutic targets.
14.	Aiddison	2020	Virtual Drug discovery platform

V. ROLE OF AI AND ML IN PROTEIN STRUCTURE PREDICTION

Accurate protein structure descriptions are crucial for understanding biology. Despite advances in experimental methods like X-ray Crystallography, NMR spectroscopy, and Cryo-EM, there's a growing gap between protein sequences and known structures. While there were around 180,000 protein structures in the Protein Data Bank (PDB) as of March 2021, there were approximately 207 million protein sequences in Uniport/TrEMBL by the end of 2020. [28]

This highlights the significance of computational protein structure prediction. Deep Learning (DL) has become a dominant technology across scientific fields, including computer vision, natural language processing, speech recognition, and more. In protein structure prediction, DL has made significant strides due to algorithm advancements and increased computational power. The Critical Assessment of protein Structure Prediction (CASP) has been evaluating protein structure modeling since 1994. Notably, contact prediction gained importance around CASP12 in 2016. Deep learning-based methods like Raptor-X showed remarkable precision, especially for long-range predictions. Post-CASP12, Raptor-X and DNCON2 further improved their methods. CASP13 in 2018 saw advancements with distogram predictions, providing richer 3D modeling data. AlphaFold and AlQuraishi's end-to-end method inspired the development of other deep learning methods [32]

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

Following CASP13, Corte's group released an open-source version of Alpha Fold called ProSPr. The trRosetta method performed similarly to Alpha Fold and offered an open-source distogram prediction method. Various other methods focused on real-valued distance prediction, including DeepDist, RealDist, and GAN-based approaches. An open-source distance prediction framework called PDNET was also introduced around the same time [28].

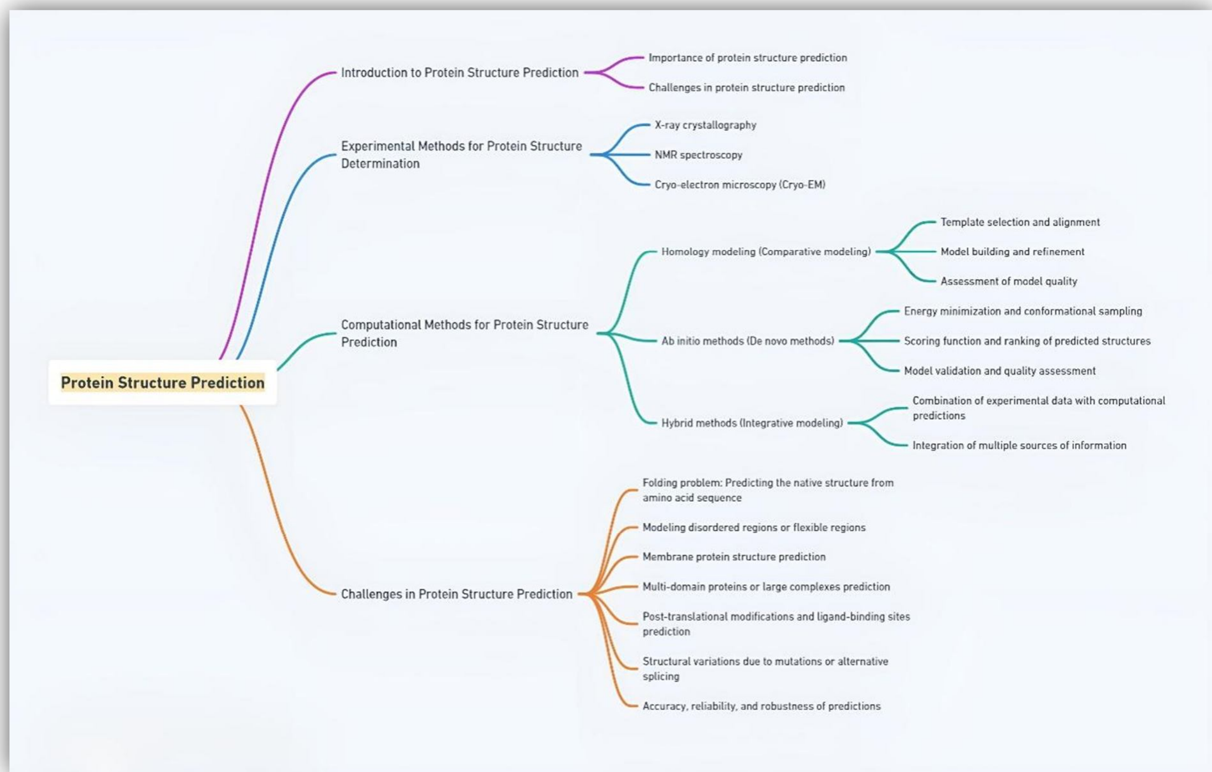


Fig. 9 Flowchart explaining the protein structure prediction

A. Alpha Fold Network for Prediction

Alpha Fold significantly enhances the precision of protein structure prediction through the incorporation of innovative neural network architectures and training methodologies rooted in the evolutionary, physical, and geometric constraints inherent in protein structures. Notably, this work introduces a novel architectural framework for the simultaneous integration of multiple sequence alignments (MSAs) and pairwise features. It also introduces a fresh output representation coupled with an associated loss function, which collectively facilitate the achievement of accurate end-to-end structure prediction. [33]

B. Working of Alpha Fold Network

The Alpha Fold network conducts direct predictions of the 3D coordinates of all heavy atoms within a given protein by leveraging primary amino acid sequences and aligned sequences from homologous proteins as input data. The methodological intricacies pertaining to the inputs, encompassing database utilization, construction of multiple sequence alignments (MSAs), and template utilization, are detailed elsewhere. Notably, the core architecture and training procedures are provided in Supplementary Methods.

The Alpha Fold network operates through two primary stages. First, the network's core, referred to as the "trunk," processes the inputs through iterative layers of an innovative neural network module, denoted as the "Evoformer." This process yields two essential outputs: an Nseq × Nres array, which encapsulates the processed MSA, and an Nres × Nres array, representing residue pairs. The MSA representation is initially seeded with the raw MSA data, albeit with specific considerations for handling exceptionally deep MSAs, as elaborated in Supplementary Methods. The Evoformer blocks encompass both attention-based and non-attention-based components.

It is demonstrated in the "Interpreting the neural network" section that the Evoformer blocks give rise to a concrete structural hypothesis early in their processing and continually refine it. Notably, the Evoformer block introduces novel mechanisms for interchanging information within the MSA and pair representations, facilitating direct inference of spatial and evolutionary relationships [34]

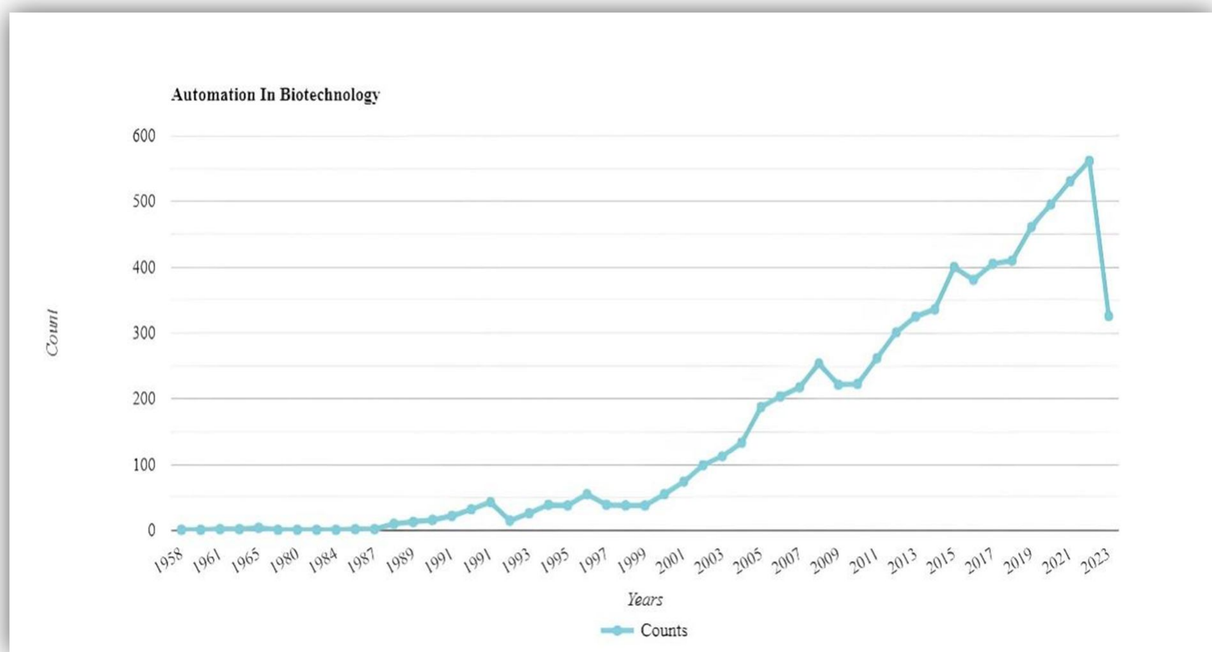
Following the trunk, the network includes a "structure module" that introduces an explicit 3D structure, incorporating rotation and translation parameters for each protein residue in the form of global rigid body frames [33]. These representations are initialized in a rudimentary state, with all rotations set to identity and positions set to the origin. However, they swiftly evolve and refine, ultimately yielding a highly accurate protein structure replete with atomic precision. Key innovations within this segment of the network involve the breaking of chain structures to enable simultaneous local refinement of all structural components. Additionally, a novel equivariant transformer is introduced to allow implicit reasoning about unrepresented side-chain atoms, complemented by a loss term that accords significant importance to the orientational fidelity of residues.

Both within the structure module and across the entire network, the principle of iterative refinement is consistently reinforced. This involves the recurrent application of the final loss function to network outputs, followed by their recursive feeding back into the same network modules. This iterative refinement process, termed "recycling," akin to concepts employed in computer vision, significantly enhances predictive accuracy with only marginal additional training time [34]

VI. BIOTECHNOLOGY AUTOMATION

Automation has consistently embodied technological advancements across various fields, including industrial biotechnology. The key feature of incorporating automation into production processes lies in the replacement of manual tasks with mechanized tools. These automated tools facilitate superior process control and enable rapid optimization with heightened accuracy. They offer several advantages, such as accelerating the pace of data generation while substantially reducing inconsistencies stemming from human errors. [35].

Automation enhances safety by eliminating human distractions, fatigue, and exposure to hazardous substances like carboxylic acids. It also reduces contamination risks in fermentation environments. On an industrial scale, automation simplifies regulatory compliance, improving quality assurance and reducing supply chain risks. [36].Biotechnology demands efficient production and innovation due to growing demand. Increased automation is essential, but it's lower compared to traditional sectors like automobiles. This is because biotech requires flexibility, individualized production, and sterility, along with strict safety regulations. Automation lowers production costs, boosts throughput, and ensures safety, particularly in high-wage countries. It also expedites the transition from research to clinical applications. [37].



Conventional biological laboratories typically have equipment at automation level five, with static workstations designed for specific tasks such as centrifugation, PCR thermo cycling, or spectrophotometry. In biotechnology, some systems reach automation level seven, like fully automated facilities producing cell cultures such as StemCellFactory and StemCellDiscovery. These systems utilize robotic arms on linear axes for autonomous movement between bench top stations, performing reagent dispensing and handling. [35]

To further increase automation, labs must interconnect existing equipment physically or digitally or acquire new devices capable of complex tasks. This presents a technical challenge due to integrating a variety of complex operations and device interfaces. Additionally, the presence of multiple equipment suppliers and limited standardization in software, lab ware, and consumables complicates this issue. [37 35]

A. Benefits of Laboratory Automation

Automation improves reproducibility through three key mechanisms: reducing human-induced variability, increasing data generation speed, and decreasing contamination risk. Human-induced variability is common in research, arising from differences in how tasks are performed by individuals. Automation mitigates some of this variability by executing repetitive tasks consistently. It also accelerates data generation and allows for testing a wider range of variables, enhancing the chances of reproducibility. Additionally, automation minimizes contamination risks associated with human involvement and environmental exposure during manual handling. Efficiency in manufacturing means achieving a high production rate with fewer resources. Automation boosts production rates and reduces resource usage, leading to increased profits [37 35].

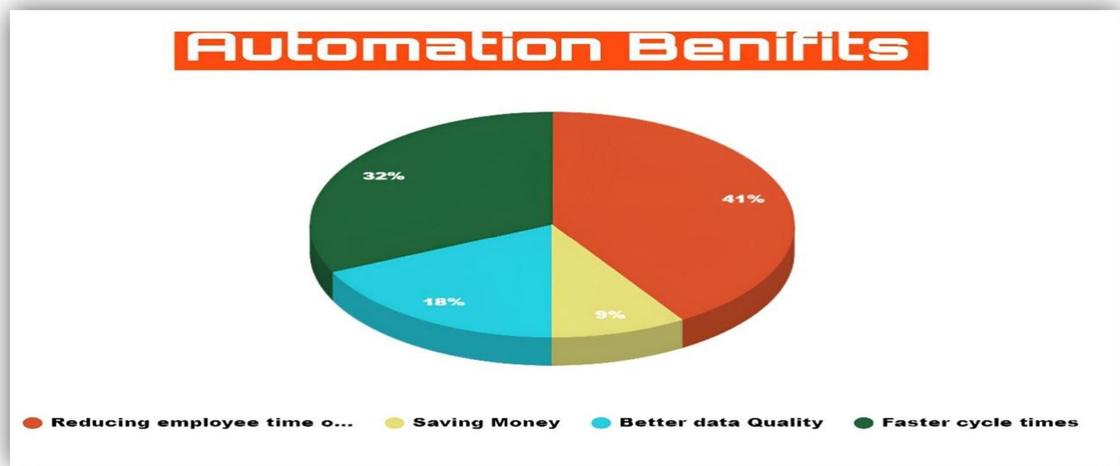


Fig 11. Automation Benefits

In research labs, automation enhances researchers' efficiency by producing more experiments without needing constant human intervention [35]. It saves researchers time, allowing them to focus on other tasks and parallel experiments. Automation also supports quick adjustments and the "fail fast, fail often" approach in pharmaceutical development. Precise reagent dispensing reduces material usage, further improving efficiency. Automation plays a vital role in applied research labs developing therapies. Integrating automation early improves product quality, speeds up lab-scale to manufacturing transition, and accelerates commercialization for clinical use. [35]

B. Limitation of Automation

While laboratory automation offers numerous benefits, there are notable limitations to consider. Incorrect implementation can lead to reduced efficiency and error propagation. Variability between automated systems and potential obsolescence pose challenges. Automation may also hinder innovation by limiting protocol alterations and could impact the workforce, particularly those focused on repetitive tasks. Additionally, both vendors and researchers should avoid exaggerating automation's capabilities, as it still requires careful operation, maintenance, and human input for experimental design and analysis. Acknowledging these limitations is essential for effective automation adoption and trust-building between commercial vendors and academic institutions. [35]

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

VII. ETHICAL AND REGULATORY CONSIDERATION

The implementation of AI-driven explanation and fairness declarations within practical applications presents a multifaceted challenge. It necessitates a comprehensive examination that encompasses diverse factors, including the typology of explanations, the extent of fairness integration, and the contextual scenarios wherein AI-informed decision-making is employed. To optimize user confidence and foster the perception of equity in AI-influenced decision-making processes, it becomes imperative to tailor the specification of explanation formats and the degree of fairness articulation within the user interface of intelligent applications [38]

Numerous pressing research questions in the AI field await exploration. Firstly, we must ensure that AI systems are developed and employed in ways aligning with ethical and societal standards, while safeguarding fundamental human rights and values. Second, it's imperative to ensure AI systems remain impartial, preventing the perpetuation or worsening of existing biases and discrimination. Furthermore, establishing transparency and explainability in AI systems is crucial to foster trust among users and stakeholders. Equally important is strengthening the security of AI systems to proactively mitigate risks to individuals and organizations. In addition, creating an inclusive environment for the development and operation of AI systems that welcomes diverse perspectives is vital.

Moreover, addressing the ethical and societal implications of emerging technologies like artificial general intelligence, machine learning, and autonomous systems is a complex challenge. Formulating and implementing effective policies, regulatory frameworks, and governance structures specific to AI is another key area of inquiry. To tackle these issues, promoting dialogue and collaboration among researchers, policymakers, industry stakeholders, civil society representatives, and other relevant actors is essential. Lastly, raising awareness and understanding of AI ethics, fairness, and trust among both the general public and those engaged in AI system design, development, and utilization is of utmost importance. [1]

User trust over fairness and explanations graph

Fig12. User confidence regarding fairness and explaining a) user confidence when fairness condition are implemented b) user confidence concerning types of explanation

VIII. EMERGING TRENDS AND FUTURE DIRECTION

The future directions of AI and ML in biotechnology are highly promising, with these transformative technologies poised to revolutionize various aspects of the biotech industry. These advancements leverage the capabilities of AI and ML algorithms to process vast datasets, recognize intricate patterns, and provide predictions that surpass human capacity for manual analysis. This technological synergy is propelling biotechnology research into new realms, particularly in drug discovery, personalized medicine, and beyond.

In the realm of drug discovery, AI and ML are on the brink of a breakthrough. These technologies facilitate the identification of novel drug targets, the creation of innovative pharmaceutical compounds, and the forecasting of drug interactions within the human body. This expedites drug development, significantly enhancing its efficiency and effectiveness.[11]

Another compelling frontier for AI and ML in biotechnology lies in personalized medicine. By scrutinizing patient data, AI and ML can pinpoint biomarkers crucial for predicting individual responses to diverse treatments. This personalization ensures that patients receive treatments tailored to their unique needs, optimizing therapeutic outcomes.[8]

Moreover, AI and ML are extending their influence into various other biotechnological domains, including gene editing. These technologies contribute to the design of more precise and efficient gene-editing tools, revolutionizing genetic research. In agricultural biotechnology, AI and ML are instrumental in developing crop varieties resilient to pests and diseases, promising to enhance global food security. Furthermore, in the realm of biomaterials, AI and ML-driven design processes are yielding innovative biomaterials with superior properties, suitable for use in medical devices and an array of applications.

In summary, AI and ML are ushering in an exciting era for biotechnology, driving progress in drug discovery, personalized medicine, gene editing, agricultural biotechnology, and biomaterials. These technologies hold the potential to reshape the industry, offering more efficient and tailored solutions to longstanding challenges.

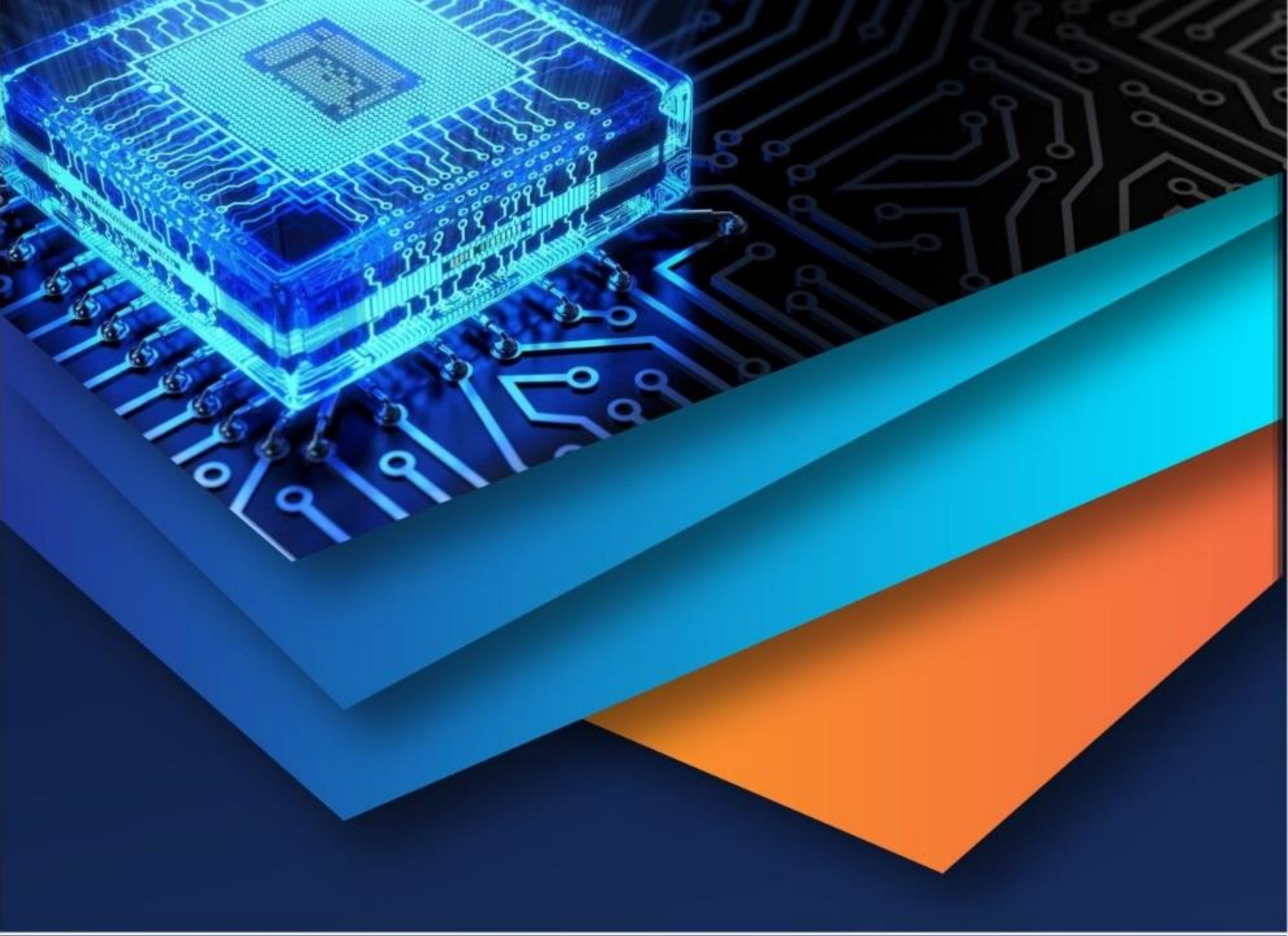
IX. CONCLUSION

The journey of AI and ML in life sciences has been marked by remarkable milestones, from early attempts at expert systems to the current sophisticated deep learning models. These technologies have revolutionized functional genomics by enabling the analysis of vast genomic datasets, deciphering genetic variations, and identifying potential therapeutic targets. In drug discovery and development, AI and ML have accelerated the identification of novel drug candidates, reduced development timelines, and enhanced precision medicine approaches. The automation of biotechnological processes has been streamlined through AI-driven robotics intelligent systems and reproducibility. Moreover, the prediction of protein structures has been greatly enhanced, opening doors to a deeper understanding of biological mechanisms and facilitating the design of novel proteins with diverse applications. The potential for personalized medicine and the discovery of new biomarkers will continue to push the boundaries of what is possible in the life sciences. Collaboration between interdisciplinary experts, including biologists, data scientists, and computational engineers, will be pivotal in harnessing the full potential of AI and ML in biotechnology. The future holds immense promise, with these technologies poised to drive groundbreaking discoveries, improve patient outcomes, and address some of the most pressing global health challenges

REFERENCES

[1] Holzinger, A., Keiblinger, K., Holub, P., Zatloukal, K. & Müller, H. AI for life: Trends in artificial intelligence for biotechnology. N. Biotechnol. **74**, 16–24 (2023).
[2] Xiao, Q. et al. High-throughput proteomics and AI for cancer biomarker discovery. Adv. Drug Deliv. Rev. **176**, 113844 (2021).
[3] Hassoun, S. et al. Artificial Intelligence for Biology. Integr. Comp. Biol. **61**, 2267–2275 (2021).
[4] Tang, A. et al. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. Can. Assoc. Radiol. J. **69**, 120–135 (2018).
[5] Greener, J. G., Kandathil, S. M., Moffat, L. & Jones, D. T. A guide to machine learning for biologists. Nat. Rev. Mol. Cell Biol. **23**, 40–55 (2022).
[6] Al'Aref, S. J. et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. Eur. Heart J. **40**, 1975–1986 (2019).
[7] Studies, C. PHYSIOLOGICAL PSYCHOLOGYl. (1950).
[8] Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F. & Peter Campbell, J. Introduction to machine learning, neural networks, and deep learning. Transl. Vis. Sci. Technol. **9**, 1–12 (2020).
[9] McBee, M. P. et al. Deep Learning in Radiology. Acad. Radiol. **25**, 1472–1480 (2018).
[10] Currie, G., Hawk, K. E., Rohren, E., Vial, A. & Klein, R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. J. Med. Imaging Radiat. Sci. **50**, 477–487 (2019).
[11] Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature **521**, 436–444 (2015).
[12] Nguyen, A., Yosinski, J. & Clune, J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. **07-12-June**, 427–436 (2015).
[13] Samuel, A. L. Eight-move opening utilizing generalization learning. (See Appendix B, Game G-43.1 Some Studies in Machine Learning Using the Game of Checkers. IBM J. 210–229 (1959).

[14] Engelhart, M. D. & Moughamian, H. Book Reviews : Book Reviews. Educ. Psychol. Meas. **28**, 619–620 (1969).
[15] LEDERBERG, J. & LEDERBERG, E. M. Replica plating and indirect selection of bacterial mutants. J. Bacteriol. **63**, 399–406 (1952).
[16] Morgan, M. G. & Wynne, B. Rationality and Ritual: The Windscale Inquiry and Nuclear Decisions in Britain. J. Policy Anal. Manag. **3**, 156 (1983).
[17] Rumelhart, D. E. & Hintont, G. E. Learning Representations by Back-Propagating Errors. Cogn. Model. 3–6 (2019) doi:10.7551/mitpress/1888.003.0013.
[18] Fleischmann, R. D. et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science (80-.). **269**, 496–512 (1995).
[19] Sutskever, I., Martens, J., Dahl, G. & Hinton, G. momentum， Nesterov accelerate， On the importance of initialization and momentum in deep learning. J. Mach. Learn. Res. **28**, 1139–1147 (2013).
[20] Craig Venter, J. et al. The sequence of the human genome. Science (80-.). **291**, 1304–1351 (2001).
[21] Ching, T. et al. Opportunities and obstacles for deep learning in biology and medicine. Journal of the Royal Society Interface vol. 15 (2018).
[22] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. 1–18 (2012).
[23] Mondal, M. R. H., Bharati, S. & Podder, P. Diagnosis of COVID-19 Using Machine Learning and Deep Learning: A Review. Curr. Med. Imaging Former. Curr. Med. Imaging Rev. **17**, 1403–1418 (2021).
[24] Blanco-González, A. et al. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. Pharmaceuticals **16**, 891 (2023).
[25] Liu, Z. et al. AI-based language models powering drug discovery and development. Drug Discov. Today **26**, 2593–2607 (2021).
[26] Giorgi, J. M. & Bader, G. D. Towards reliable named entity recognition in the biomedical domain. Bioinformatics **36**, 280–286 (2020).
[27] Tsuji, S. et al. Artificial intelligence-based computational framework for drug-target prioritization and inference of novel repositionable drugs for Alzheimer's disease. Alzheimer's Res. Ther. **13**, 1–15 (2021).
[28] Caudai, C. et al. AI applications in functional genomics. Comput. Struct. Biotechnol. J. **19**, 5762–5790 (2021).
[29] Lesk, A. CHAPTER 1. Introduction to genomics. Introd. To Genomics **823**, 3–20 (2017).
[30] Abdellah, Z. et al. Finishing the euchromatic sequence of the human genome. Nature **431**, 931–945 (2004).
[31] Bak, R. O., Gomez-Ospina, N. & Porteus, M. H. Gene Editing on Center Stage. Trends Genet. **34**, 600–611 (2018).
[32] Pakhrin, S. C., Shrestha, B., Adhikari, B. & Kc, D. B. Deep learning-based advances in protein structure prediction. Int. J. Mol. Sci. **22**, (2021).
[33] Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature **596**, 583–589 (2021).
[34] Imaging, N. & Angeles, L. Auto-context and Its Application to High-level Vision Tasks and 3D Brain Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 1744–1757 (2010).
[35] Holland, I. & Davies, J. A. Automation in the Life Science Research Laboratory. Front. Bioeng. Biotechnol. **8**, 1–18 (2020).
[36] Wainaina, S. & Taherzadeh, M. J. Automation and artificial intelligence in filamentous fungi-based bioprocesses: A review. Bioresour. Technol. **369**, 128421 (2023).
[37] Solutions, E. & Developments, O. Existing Solutions and Ongoing Developments. (2021).
[38] Angerschmid, A., Zhou, J., Theuermann, K., Chen, F. & Holzinger, A. Fairness and Explanation in AI-Informed Decision Making. Mach. Learn. Knowl. Extr. **4**, 556–579 (2022).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY