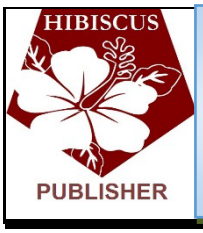
JOBIMB, 2023, Vol 11, No 2, 1-10
https://doi.org/10.54987/jobimb.v11i2.835



JOURNAL OF BIOCHEMISTRY, MICROBIOLOGY AND BIOTECHNOLOGY

Website: http://journal.hibiscuspublisher.com/index.php/JOBIMB/index

Application of Artificial Intelligence in Biochemistry Research: A Review

Mahmoud Suleiman Jada[1]*, Hyelda Stephen[1], Mohammed Maikudi Usman[2] and Abdullahi Usman Wurochekke[1]

[1]Department of Biochemistry, Faculty of Life Sciences, Modibbo Adama University, Yola, P.M.B. 2076, Yola Adamawa State, Nigeria.
[2]Department of Biotechnology, Faculty of Life Sciences, Modibbo Adama University, Yola, P.M.B. 2076, Yola Adamawa State, Nigeria.

*Corresponding author:
Dr. Mahmoud Suleiman Jada
Department of Biochemistry,
Faculty of Life Sciences,
Modibbo Adama University,
Yola, P.M.B. 2076,
Yola Adamawa State,
Nigeria.
Email: jadasm84@gmail.com

HISTORY

Received: 27th Sep 2023
Received in revised form: 21st Nov 2023
Accepted: 15th Dec 2023

KEYWORDS

Artificial Intelligence
Biochemistry Research
Data
Genomics
Protein folding

ABSTRACT

The exploration of Biochemistry research has been expanded by artificial intelligence (AI) and its ability to analyse immense and intricate datasets in a way that would be unattainable by human effort alone. This review delves into the most recent examples of AI breakthroughs that had a transformative impact on key aspects of biochemistry. AI has now led to the creation and improvement of drug molecules and the capability to predict which new proteins could be targeted for repurposing with current drugs. When it comes to protein structures, algorithms such as AlphaFold have made great strides in resolving the protein folding problem that has been a challenge for so long. Reliably identifying proteins and metabolites from spectral data is now possible with deep learning models. Meanwhile, AI can classify sequences and spot gene expression patterns in massive genomics and transcriptomics datasets with ease. The remarkable capabilities of AI to automate the analysis of medical images and natural language descriptions of patient symptoms have a promising potential for transforming disease diagnosis and treatment. Nevertheless, obstacles such as data availability, interpretability of AI models, ethical considerations, and generalization must be tackled as these technologies evolve. The collaboration between AI and biochemistry appears to be optimistic, with biochemical data powering the development of more robust AI systems that can extract new knowledge from vast datasets beyond human reach. Thus, this mutually beneficial relationship has the potential to vastly expedite discovery across molecular biology.

INTRODUCTION

Interacting with biological molecules that exist in living organisms, biochemistry research involves delving into the intricacies of structure and function while also exploring various chemical processes. Biomolecules such as proteins, nucleic acids, and lipids are meticulously scrutinized while metabolic pathways are mapped. Additionally, mechanisms of enzymatic reactions are explored extensively, among other areas [1]. Unsurprisingly, the field generates substantial and multidimensional datasets, including complex systems biology models, DNA and protein sequences, gene expression data, and metabolomics profiles [2]. Due to the increasingly challenging nature of analyzing these vast biochemical datasets through traditional computational methods, the integration of artificial intelligence offers great potential to revolutionize biochemistry research [3]. This is where artificial intelligence shows great promise to transform biochemistry research. Tasks, such as recognizing visuals, learning from previous input, and understanding spoken words required human intelligence until AI was developed. AI is the practice of creating computer systems that are capable of performing these tasks just as humans would [4].

Machine learning, a type of AI, enables computers to learn from data to make predictions and decisions without explicit programming [5]. The exceptional capacity of machine learning algorithms to rapidly analyse immense, multifaceted biochemical data has the potential to uncover novel insights previously out of reach [6]. For instance, deep learning neural networks have been

used to predict protein secondary structure from amino acid sequences with high accuracy [7] and identify enzyme commission numbers to infer metabolic pathways [8].
Looking forward, biochemistry and AI are poised to mutually advance each other's progress. As biochemistry research produces growing volumes of multidimensional data, increasingly sophisticated AI capabilities will be essential to extracting meaningful patterns and knowledge [9]. In turn, having vast biological datasets to train on will push AI techniques forward [10]. Together, the synergistic combination of biochemistry and AI promises exciting new frontiers in deciphering the complexity of living systems at a deeper level [11]. This review explores the state-of-the-art applications of AI propelling biochemistry research forward, as well as future directions in this rapidly developing field.

Literature search strategy and study design
An extensive literature search was conducted to identify relevant studies on the application of artificial intelligence (AI) in biochemistry research. The following search engines and databases were searched: Google Scholar, PubMed, Science Direct, Web of Science, Scopus, and IEEE Xplore. The search strategy included a combination of keywords related to AI and biochemistry. The reference lists of included articles were hand-searched to identify any additional relevant studies. Relevant studies on applications of artificial intelligence in biochemistry research published in English were included. Non peer-reviewed articles were excluded.

Data was extracted on study details, AI methods, biochemistry domain and application, and performance metrics. The findings were synthesized narratively by grouping articles based on the area of biochemistry research: drug discovery, protein structure prediction, genomics/transcriptomics, metabolomics/proteomics, disease diagnosis and treatment. A descriptive summary is provided for the types of AI methods applied, the specific tasks suited for AI, and the impact on biochemistry research.

Types of Artificial Intelligence

Machine learning and deep learning are two AI techniques that show significant potential for enhancing biochemistry research.

Machine learning
According to Dasgupta and Nath [12] machine learning is the idea of utilizing an algorithm to enhance its performance by learning from data. There are four main types of problems that machine learning can solve, namely prediction, clustering, classification, and dimensionality reduction, as described by Liu, Esan [13]. Machine learning techniques can be classified into four groups depending on their learning methods, namely supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning, according to Ayodele [14].

Supervised learning
Supervised learning is a type of machine learning where a model learns to map inputs to outputs based on labelled training data (**Fig. 1**) [15]. This method is task-driven since it is frequently employed when specified objectives must be satisfied from a collection of inputs [16]. Supervised learning tasks are divided into two major categories, namely classification tasks, which involves the segregation of data into specific categories by the model, and regression tasks, which involves fitting the data to a function [5].

For example, supervised learning can be used to predict the activity of a drug candidate based on its chemical structure or to classify a set of proteins into functional groups based on their sequence or structural features.
To ensure the accuracy of supervised learning models, it's important to have high-quality labelled data. This can be time-consuming and costly, and in some cases, it may not be feasible to obtain a sufficient amount of labelled data. In such cases, semi-supervised or unsupervised learning methods may be used instead [5].

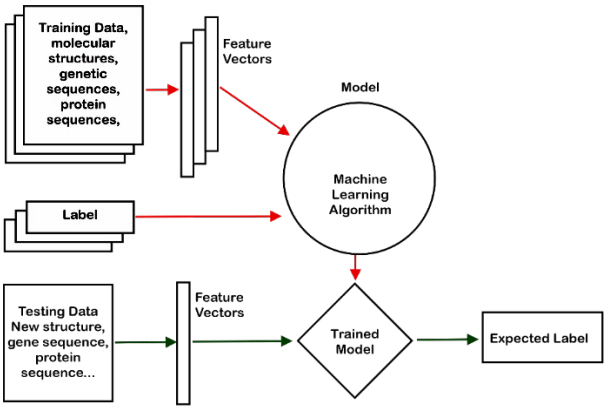


Fig 1. Supervised learning model. The main task is to construct an estimator able to predict the label of an object given by the set of features [17].

Unsupervised learning
Unsupervised learning is a type of machine learning that analyses unlabelled datasets (Fig 2.), without the need for human interference [15]. In biochemistry, unsupervised learning can be particularly useful for identifying patterns and relationships in large datasets, such as gene expression data or protein-protein interaction networks [18].
However, unsupervised learning is much harder than supervised learning because the computer must learn to perform tasks without explicit instructions [14]. There are two main approaches to unsupervised learning.

The first approach involves teaching the agent through a reward system, rather than explicit categorizations. This approach is particularly useful for decision-making problems, where the goal is to make a decision rather than to categorize the problem. In biochemistry research, this approach can be used for drug discovery, where the goal is to identify compounds that are likely to have a certain biological effect, without necessarily knowing the mechanism by which they work. The reward system can be learned from previous successes and failures [19].

The second approach is clustering, where similarities in the training data are identified to form groups, or clusters. The assumption is that the clusters will match reasonably well with an intuitive classification. In biochemistry research, clustering can be used to group proteins or genes based on their expression profiles or functional annotations, for example. Clustering can also be used for drug discovery, where compounds are grouped based on their chemical properties or biological activities [12].

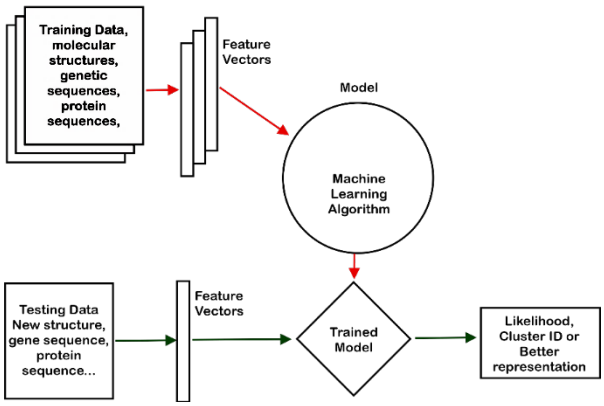


Fig 2. Unsupervised learning model. The unsupervised learning algorithms are searching the similarity between pieces of data in order to determinate if they can be categorized and create a group [17].

Semi-supervised learning

Semi-supervised learning is a method that combines both supervised and unsupervised learning by using both labelled and unlabelled data for training a model [15, 16]. In the field of biochemistry, labelled data may be scarce due to the high cost and time involved in labelling experimental data. Thus, semi-supervised learning can be an effective technique for training models in biochemistry research where only a limited number of labelled samples are available. Semi-supervised learning has been successfully applied in various areas of biochemistry, such as protein classification and prediction, gene expression analysis, and drug discovery [20].

Reinforcement learning

Through Reinforcement Learning (RL), a machine learning technique, an agent can acquire knowledge through experimentation and feedback from its encounters and activities in an interactive environment. This approach diverges from supervised learning as RL does not require labelled instances for training models. Rather, it depends on the agent's associations with its surroundings. The fundamental objective of RL is to develop the ability to make consecutive choices that exploit a long-lasting reward [5].

The RL problem was classified as a Markov Decision Process (MDP) in Puterman [21], where the agent interacts with the environment in a sequence of discrete time steps, and at each step, it receives feedback in the form of a reward signal. An RL problem typically consists of four components: an agent, an environment, a reward function, and a policy. The agent takes actions based on the current state of the environment and its policy, and the environment responds with a new state and a reward signal. **Fig.3** below illustrates the action-reward feedback loop of a generic RL model.

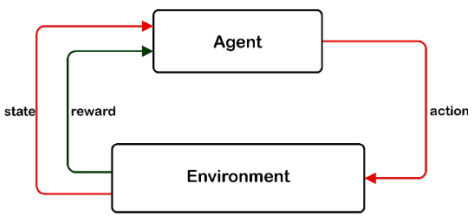


Fig. 3. Reinforcement learning, the goal is to find a suitable action model that would maximize the total cumulative reward of the agent [22].

RL can be broadly classified into two categories: model-based and model-free approaches. Model-based RL involves building model of the environment and using it to make decisions. The agent learns the optimal behaviour by performing certain actions and observing the results, which is made up of the next state and the instantaneous reward [23]. In contrast, model-free RL algorithms do not require model of the environment. They directly learn from experience by estimating the optimal policy or value function. Model-free algorithms such as Q-learning, Monte Carlo Control, SARSA (State–Action–Reward–State–Action), Deep Q Network, etc., are widely used in RL [24].

RL has several applications in biochemistry research, including drug discovery and protein structure prediction. RL has been used to design small molecules with desirable properties such as drug efficacy, solubility, and toxicity. It has also been used to optimize the parameters of molecular dynamics simulations and to predict protein-ligand binding affinities. In drug discovery, RL algorithms can be used to identify new compounds that are likely to be effective in treating diseases. In protein structure prediction, RL can be used to predict the tertiary structure of a protein from its primary sequence [18]. Overall, RL is a powerful tool for solving complex problems in biochemistry research.

Deep learning

Deep learning is a specialized subset of machine learning based on artificial neural networks composed of multiple layers. The layered architecture enables deep learning models to learn structured feature representations with hierarchy from raw input data. This ability to automatically extract meaningful patterns makes deep learning exceptionally adept at working with large, complex biochemical data [25].

Artificial Neural Networks (ANN)

Artificial neural networks are modelled after the structure and function of biological neural networks. They consist of layers of artificial neurons that are connected to each other through weighted connections. These connections allow information to flow through the network, with each neuron receiving input from other neurons, processing that input, and passing on its own output to other neurons in the network [26].

The process of training an artificial neural network involves adjusting the weights of these connections in order to optimize the performance of the network on a given task. This process is similar to the way that biological neural networks are "trained" through experience and learning, a process called backpropagation, which allows the network to adjust its weights in response to feedback from the output layer [5]. Graph neural networks (GNNs) are a type of ANN that operates on graphs as inputs. GNNs have recently been used to learn representations of low-dimensional biomolecular networks [27, 28]. For example, Ahmed, Park [29] used two different GNN methods to develop a GNN that uses gene expression data and a network of genes that are expressed together. This network represents the relationship between the expression of gene pairs. GNN can also be used in prediction of the dynamic property of biochemical pathways [30].

Convolutional Neural Networks (CNN)

The traditional CNN model is a sophisticated and high-potential ANN variation that was created to handle growing degrees of complexity as well as data pre-treatment and compilation. It is based on how animal brains process visuals, specifically how the neurons in our visual cortex are organized [31]. In a CNN, the input data is first processed to extract important features, such as

edges and shapes. This is done using a series of filters; these filters work in stages (**Fig. 4**), each time focusing on different aspects of the data in a more detailed way [32]. These filters are similar to how enzymes in a biochemical pathway act to convert one molecule into another, from enzyme to enzyme, until the final product.

The output of these filters is then further processed to reduce the dimensionality of the data. This process of feature extraction and reduction continues through multiple layers of the network until the final output layer, which produces the desired classification or prediction [33]. CNNs are highly effective at processing data with and without images because of their ability to automatically learn and extract features from the input data without the need for manual feature engineering. This makes them highly flexible and adaptable to a wide range of applications, similar to how enzymes can work on different substrates and be used in various biochemical pathways [34].

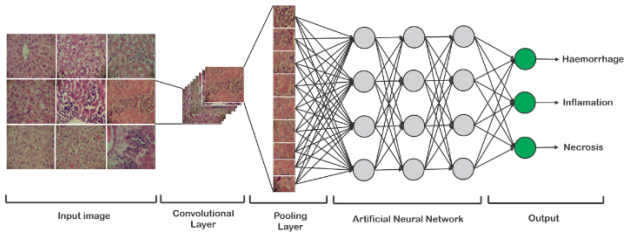
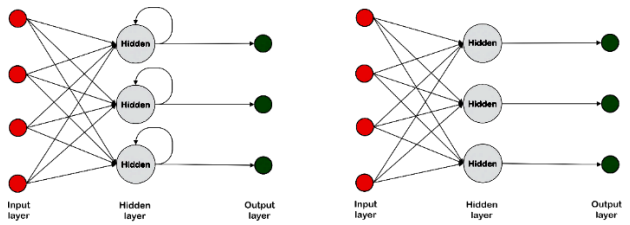


Fig. 4. Convolutional Neural Network demonstrating neural analysis of histopathology micrograph.

Recurrent Neural Networks (RNN)

RNNs are a type of neural network that are well-suited for sequence prediction tasks. They work by taking a sequence of inputs and predicting the next output in the sequence. RNNs are able to do this because they have a memory that allows them to remember the previous inputs as illustrated in **Fig. 5a**. This is in contrast to feed-forward neural networks (**Fig. 5b**), which cannot remember previous inputs and can only predict the next output based on the current input [35].



(a) Recurrent Neural Network (b) Feed-Forward Neural Network

Fig. 5. The comparison between Recurrent Neural Network (FFNN) (a) and Feed-Forward Neural Network (b). It demonstrates in FFNN there is only one direction for the data to move, whereas in RNN there is a loop [36].

One of the challenges of using RNNs is that they can be difficult to train. This is because they can easily forget information about previous inputs, which can lead to errors in the predictions. To address this challenge, researchers have developed a number of techniques, such as long short-term memory (LSTM) networks. LSTM networks are a type of RNN that are able to remember information about previous inputs for long periods of time. This makes them more accurate for sequence prediction tasks [37]. RNN is can be applied in predicting associations between biochemical markers and certain diseases [38].

Generative Adversarial Networks (GAN)

GANs are a deep learning approach that combines a generator and a discriminator neural network. The generator creates fake data, while the discriminator tries to distinguish between real and fake data [39]. The two networks compete with each other, with the generator trying to create more realistic fake data and the discriminator trying to become better at distinguishing between real and fake data. This competition helps both networks to improve over time [34]. A good example is ProteinGAN (**Fig. 6**), which has been shown to be able to generate functional protein sequences with a high degree of accuracy. It works by training two neural networks against each other. The first network, called the generator, is responsible for generating protein sequences. The second network, called the discriminator, is responsible for distinguishing between real and generated protein sequences [40].

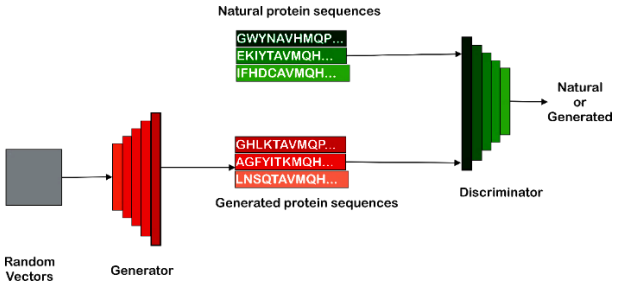


Fig. 6. Protein GAN training scheme. The Generator network creates a protein sequence from a random input vector, and the Discriminator network scores it by comparing it to real protein sequences. Since the generator has never really seen genuine enzyme sequences, it attempts to trick the discriminator by creating sequences that will ultimately resemble real ones [40].

GANs can be used to generate new molecules that have the potential to be new drugs. This can be done by generating molecules that are similar to known drugs, or by generating molecules with new properties that could be useful for treating diseases [20]. GANs can be used to improve the quality of biomedical images, such as by removing noise or enhancing contrast. This can be used to help researchers make better diagnoses and develop new treatments for diseases [41].

Natural Language Processing (NLP)

NLP encompasses various computational techniques that involve the representation, transformation, or utilization of text, speech, and other forms of data. This means that a broad range of tasks can be considered as NLP activities, including extracting relevant information from scientific literature, analysing and summarizing research papers, and generating hypotheses based on textual data. NLP has a variety of applications in analysing scientific literature [42]. One major application is the identification of trends in research, relevant research papers, and the summarization of their findings [43]. Additionally, NLP can be used to generate hypotheses based on scientific literature [42], which can be helpful for researchers who are looking to explore new areas of research. Also, NLP could be used to enhance clinical trial research, and facilitating inter-country/region and inter-institution collaborations [44]. Another way NLP can be used in science is by communicating scientific findings to the public [42].

Applications of AI in Biochemistry Research

Artificial intelligence is enabling breakthroughs across key domains in biochemistry including drug discovery, protein structure prediction, genomics, metabolomics, and proteomics. By assisting with critical tasks like drug design, sequence analysis, and spectral metabolite identification, AI holds enormous promise for accelerating discoveries in biochemistry.

Drug discovery and development

The lengthy, complex, and costly process of discovering and developing new drugs is being transformed by the integration of artificial intelligence techniques across the pipeline (**Fig. 7**). Utilization of AI in drug discovery and development span from designing or optimizing drug candidate molecules using generative models, to repurposing existing therapeutics by predicting new protein targets, to improving clinical trial efficiency through patient selection and adverse event detection, to providing early predictions on drug safety and toxicity to reduce late-stage failures [34]. By enhancing molecular design, preclinical screening, clinical testing, and other aspects, the synergy between AI and the drug development process holds promise to increase the speed and reduce the costs associated with bringing new medicines to market [45].

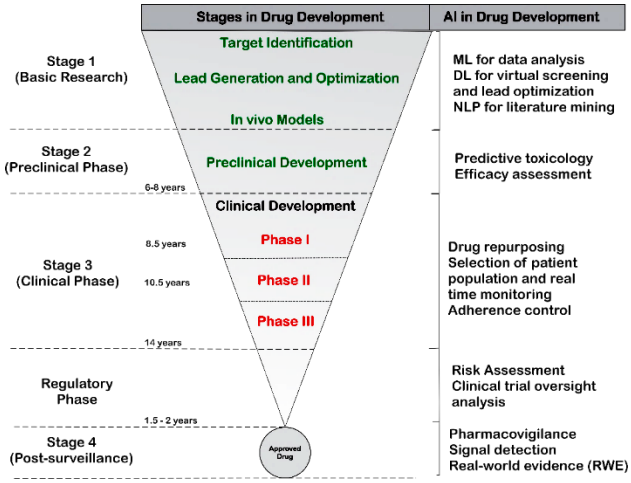


Fig 7. Drug development process showing the application of AI at each stage [45].

Drug discovery

Various AI models have been developed to design new drugs from scratch without using information from existing drugs [46]. Unfortunately, these de novo methods are not as popular as other drug design methods that are based on the structures of existing drugs. These de novo methods generate drugs that can be hard to make. One model is the variational autoencoder which has two neural networks: an encoder and a decoder [46]. The encoder translates a drug's Simplified Molecular Input Line Entry System (SMILES) code into a continuous vector. The decoder then translates that vector back into a SMILES code, which usually represents a similar drug. Researchers compared this model to an adversarial autoencoder [47].

Adversarial autoencoders have a model that can generate new chemical structures [48]. One study used this model to find new drugs that could target dopamine receptor type 2. Another used a generative adversarial network to find new anticancer drug candidates [49]. De novo drug design has also been done using RNNs. RNNs were initially employed for NLP, where they process sequential data. RNNs can create novel chemical structures since SMILES codes encode chemical structures as letter sequences. To learn how to create SMILES codes, RNNs are trained on vast datasets like ChEMBL or commercial drug databases [50]. New peptides have been produced using this method. The produced molecules have been biased to have particular characteristics using RL. Transfer learning has also been used to create medications with particular biological effects by transferring knowledge from one model to another. Different machine learning methods have been developed to investigate uncharted chemical territory and produce novel drug-like compounds [51].

Drug repurposing

We can find already available, FDA-approved medications that may interact with and limit the activities of proteins from new viruses like SARS-CoV-2 using AI and deep learning techniques like AlphaFold, which can predict 3D protein structures from 1D amino acid sequences [52]. Beck, Shin [53] developed a machine learning model called MT-DTI that can predict how strongly drugs and proteins may bind to each other based only on the chemical makeup of the drugs and amino acid sequences of the proteins. Using this model, they were able to identify several existing antiviral drugs, such as atazanavir, remdesivir, efavirenz, ritonavir, and dolutegravir, which MT-DTI predicted would inhibit a key protein called SARS-CoV-2 3C-like proteinase.

These drugs could potentially be repurposed as treatments for COVID-19 and further tested in clinical trials. However, it is important to note that while these computational predictions show promise, experimental validation through *in vitro* and *in vivo* studies is still necessary to confirm the actual efficacy and safety of these drugs against SARS-CoV-2 before advancing to human trials. MT-DTI allows researchers to discover potential new uses for existing drugs without needing the 3D structures of the drug targets [52].

Clinical development

AI algorithms can be used in clinical development to improve the likelihood of success in clinical trials, optimize trial design, detect adverse events, and improve the efficiency of data collection and analysis [54]. By analysing patient data and biomarkers, AI can identify patients more likely to respond to a particular therapy, reduce the number of patients needed for clinical trials, and improve patient safety. Automating data collection and analysis can also reduce errors and accelerate the clinical development process [55].

Toxicological studies

AI techniques like predictive toxicology modelling using machine learning (PTML) are being applied to gain insights into drug toxicity and accelerate drug development. PTML can help understand why drugs fail preclinical studies and redesign them to be safer, identify the most promising candidates earlier, predict how structural changes may impact toxicity, reduce animal testing, and streamline the overall drug development process [56].

Protein Structure Prediction

The protein folding problem

The notion of a folding “problem” first emerged around 1960, with the appearance of the first atomic-resolution protein structures [57]. Since then, three distinct issues have been identified with the protein folding problem: (a) Understanding the thermodynamic forces that stabilize a protein's native structure for a given amino acid sequence, including the equilibrium between interatomic interactions that defines a

protein's fold; (b) Developing computational methods to predict a protein's three-dimensional structure directly from its one-dimensional amino acid sequence; (c) Elucidating the kinetics of how proteins are able to fold rapidly into their native states, analysing the mechanisms and transition pathways involved [58]. Protein folding process has puzzled scientists for decades. One reason for this is the enormous number of possible 3D shapes that a protein chain can fold into before achieving the correct conformation. In 1969, molecular biologist Cyrus Levinthal estimated that a single protein could fold into 10^{300} conformations. Given the vast number of possibilities, it would seem mathematically infeasible for proteins to explore all potential conformations through random searching until they stumble upon the correct shape, yet, proteins consistently and efficiently adopt their functional structure in a fraction of a second, a phenomenon known as Levinthal's Paradox [59].

Analysing a protein's structure is crucial to understanding its function, but the current techniques require crystallized proteins. This approach is not ideal for hydrophobic membrane proteins that aggregate in water and are difficult to crystallize. This is where computational models come in. The development of AI-powered tools that can accurately predict and visualize protein structures could revolutionize the field by expanding our understanding of 3D protein structures and bridging the gap between 1D and 3D protein analysis [60].

Critical Assessment of Structure Prediction (CASP)

The Critical Assessment of Techniques for Protein Structure Prediction (CASP) is a biennial competition that aims to evaluate and improve the accuracy of computational methods for predicting the 3D structure of proteins. CASP was established in 1994 and has since become one of the most prominent and influential scientific competitions in the field of bioinformatics and structural biology [60].

The idea behind CASP is to provide a platform for researchers to test and compare their computational methods for predicting protein structures against experimental data. During the competition, participants are provided with amino acid sequences for a set of proteins whose structures have not yet been experimentally determined. The participants use computational methods, such as homology modelling, molecular dynamics simulations, and fragment assembly, to predict the 3D structure of the protein. The predictions are then compared to the experimentally determined structures, which are released after the competition.

The primary measure used to evaluate the accuracy of computational models is the Global Distance Test (GDT), which ranges from 0 to 100 (**Fig 8**). It represents the percentage of amino acid residues that fall within a certain distance of the correct position, with the experimental structures serving as the "ground truth", with values above 80% denoting that local and global details are mostly modelled accurately and values below 20% denoting mostly random models. Moult has stated that achieving a GDT score of approximately 90 is comparable to experimental methods [61].

There are two approaches to developing computational methods for the protein folding problem. One is based on physical interactions and leverages our understanding of molecular driving forces. The other approach is rooted in evolutionary analysis and uses bioinformatics to study the evolutionary history of proteins [7].

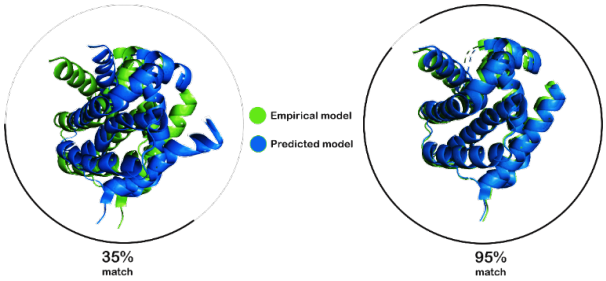


Fig. 8. The global distance test illustrated. The closer the predicted model is to the empirical model, the higher the GDT score.

AI in protein folding prediction

Historically, the GDT scores at CASP have typically been around 60 [61]. However, in 2020, Google's DeepMind unveiled AlphaFold 2.0, which managed to achieve an impressive average GDT score of 90 at CASP14 [7]. **Fig 9**. below is the CASP median free modelling category over the years showing the best model GDT score.

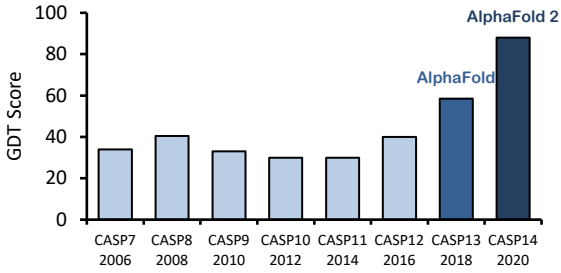


Fig 10. CASP median free modelling category over the years showing the best model GDT score [7].

In 2018, DeepMind achieved a high GDT score in CASP13 and a median score of approximately 70-75 using the first version of AlphaFold, which combined local physics and pattern recognition. An improvement in accuracy was observed after incorporating deep learning methods for contact prediction but often overestimated the impact of interactions between nearby residues [62]. DeepMind subsequently developed AlphaFold 2.0, a model that relies exclusively on pattern recognition and uses both physical and evolutionary constraints in its predictions [63].

The architecture is an attention-based neural network combined with a deep learning framework [7], that treats the prediction process like assembling a Lego set, where smaller sections of set, in this case, amino acids are connected before being joined together to form the built Lego set or the 3D protein structure. To train the network, DeepMind used a dataset consisting of 170,000 protein structures from the Protein Data Bank and 350,000 sequences from UniClust, with high-confidence predictions combined with PDB data to create a new dataset for further training [60].

AlphaFold uses an iterative process to refine its predictions, with an internal measure called pLDDT based on the Local Distance Difference Test (LDDT) to assess reliability by comparing local distances of atoms in computational models to experimentally determined structures. This system allows AlphaFold to improve its predictions and produce more accurate structures [7].

To improve its predictions, AlphaFold uses an iterative process and relies on an internal measure called the predicted Local Distances Difference Test (pLDDT), which assesses the reliability of its predictions by comparing the local distances of atoms in computational models to experimentally determined structures [64]. This metric assigns high scores to regions with high local accuracy, regardless of the accuracy of the entire predicted protein, allowing AlphaFold to refine its predictions and achieve more accuracy. In the recent CASP14 experiment, AlphaFold achieved a median score of 92.4 GDT, with an average error within the width of one atom [60].

RoseTTAFold, developed by a team led by David Baker, and was one of the top-performing programs in the latest CASP14, uses deep learning techniques to predict the structure of proteins. However, it employs a different approach to achieve this goal. Rather than relying solely on amino acid sequence data, RoseTTAFold combines sequence data with structural data, such as the positions of atoms within the protein. This allows the program to more accurately predict the folding of the protein. One of the unique features of RoseTTAFold is its ability to predict the structures of protein complexes. This is a significant advantage over AlphaFold, which is currently limited to predicting the structures of individual proteins. Additionally, RoseTTAFold can predict the structures of proteins containing cofactors or metals with an accuracy of 80%, which are not included in AlphaFold's predictions [65].

Genomics and Transcriptomics

Gene expression analysis

AI plays a crucial role in gene expression analysis, which involves studying the activity of genes in a given tissue or cell. By leveraging AI algorithms, researchers can analyse large-scale genomics data, such as RNA sequencing (RNA-seq) data, to understand gene expression patterns and identify differentially expressed genes [66]. AI-based approaches can help in identifying gene signatures associated with specific diseases or conditions such as cancer, allowing for better diagnosis and prognosis [67]. These methods can also assist in identifying potential therapeutic targets or biomarkers for various diseases.

Sequence alignment and classification

AI techniques are widely used in sequence alignment and classification tasks in genomics and transcriptomics [68]. Machine learning algorithms, including support vector machines (SVMs) and deep learning models, are employed to classify sequences based on their features, such as sequence motifs or structural properties [20].

Metabolomics and proteomics

Metabolomics and proteomics are being revolutionized by AI by enabling efficient and accurate identification of metabolites and proteins. AI algorithms like deep neural networks and support vector machines have been used to analyse mass spectrometry data and identify patterns, leading to breakthroughs in disease diagnosis and treatment.

Metabolite identification

AI has revolutionized the field of metabolomics by enabling efficient and accurate identification of metabolites in complex biological samples. Metabolite identification involves matching experimental data, such as mass spectrometry (MS) or nuclear magnetic resonance (NMR) spectra, with reference databases to determine the chemical identity of metabolites [69]. Deep neural networks, such as CNNs have been utilized for metabolite identification. This model analysed MS or NMR spectra and learned complex patterns and features to accurately identify metabolites [70]. Deep generative models, such as variational autoencoders (VAEs) and GANs have been used to generate synthetic metabolite spectra. These models learn the underlying distribution of metabolite spectra and can be used for metabolite identification by comparing experimental spectra with synthetic ones [71].

Protein identification and quantification

AI has made significant contributions to protein identification and quantification in proteomics research. CNNs have been employed to accurately identify proteins from MS data. These models can handle large-scale datasets and provide highly accurate protein identification results [72]. Additionally, AI algorithms enable the quantification of proteins in complex mixtures. Machine learning techniques, such as support vector regression and random forest regression, have been used to estimate protein abundances based on spectral counts or peptide intensities [73]. These models learn patterns and correlations in the data, allowing for accurate protein quantification even in the presence of noise or missing values.

Pathway analysis

Machine learning performs and important role in pathway analysis, which aims to understand the interactions and relationships between metabolites, proteins, and other molecules within biological pathways [74]. By integrating metabolomics and proteomics data with existing pathway databases, Deep learning algorithms can predict and analyse metabolic and signalling pathways, helping to elucidate the underlying mechanisms of diseases and identify potential drug targets [75]. AI-based pathway analysis can uncover complex network interactions, identify key regulatory nodes, and provide insights into how specific disturbances impact cellular processes [74].

Disease Diagnosis and Treatment

Medical image analysis

AI can be used to analyse medical scans and detect anomalies. For example, AI systems can detect tumours, lesions and other abnormalities in CT scans, X-rays and MRI images [76]. AI can also analyse retinal scans to detect eye diseases [77].

Natural language diagnosis

AI systems can analyse written or spoken symptoms described by patients to determine the possible diagnosis or conditions. The systems compare the symptoms to a large database of diseases and disorders to provide possible matches [78]. Some companies are developing 'chatbots' to converse with patients and get additional details about symptoms to improve diagnosis [79].

Clinical decision support system

AI systems can analyse medical records, test results, symptoms and other data to provide decision support for physicians. The systems can detect possible conditions that match the patient's data and provide a list of recommended tests or treatments. Some AI systems can also analyse how physicians make decisions and detect potential biases or errors [80]. The systems aim to enhance human expertise, not replace physicians.

Challenges and Limitations of AI in Biochemistry Research

Although AI has shown great promises, there are still challenges and limitations to address, some of which include, availability of data, interpretability of results, ethical considerations and generalization and overfitting.

Availability of data
In order to fully harness the potential of AI in biology, it is important to develop technologies that can automatically collect biological data from various sources such as images, videos, and molecular profiles. However, the quality of the data collected is equally important, and data scientists must work with biologists to ensure that the data is accurate and reliable. This involves identifying and mitigating biases, understanding variations, and improving signal-to-noise ratios. To facilitate data sharing, tools should be developed that allow for transparent data sharing while also taking into consideration issues of security, privacy, and fairness.

High-quality reference datasets are also crucial for benchmarking AI applications in biology. The ImageNet dataset has served as a benchmark for AI methods in image processing, and similar reference datasets will be needed for AI applications in biology. Sharing data and developing reference datasets will allow researchers to form new hypotheses and build new theories, leading to further advancements in the field of AI and biochemistry [81].

Interpretability of results
AI models are often complex and difficult to understand, making it challenging to explain how they arrive at a particular decision. Interpretability tools and techniques are designed to help humans understand and interpret these models. For example, one approach is to visualize the model's decision-making process by generating heatmaps that highlight the regions of an image that contributed most to a particular classification. Another approach to interpretability is to generate feature importance scores that rank the importance of different input features in making a particular prediction. This can help identify which features are most relevant to the decision and provide insight into how the model works [82].

Generalization and overfitting
The goal of any machine learning model is to learn patterns in the training data that can be applied to new, unseen data. A model that has good generalization will be able to accurately predict outcomes on new data, even if the new data is different in some way from the training data. Good generalization is essential for creating models that are useful in real-world applications. However, overfitting, occurs when a model becomes too complex and begins to fit the noise in the training data, rather than the underlying patterns. This can result in a model that performs very well on the training data but poorly on new, unseen data. Overfitting can occur when a model is too complex for the amount of data available, or when the model is trained for too many iterations [83].

To avoid overfitting and promote generalization, machine learning practitioners use techniques such as cross-validation, regularization, and early stopping. These techniques help to prevent models from becoming too complex and overfitting the training data. Additionally, increasing the amount of training data can also help to improve generalization by giving the model more examples to learn from [83].

Ethical considerations
The use of AI in biochemistry research presents several ethical considerations. One of the main concerns is the potential for bias in the data used to train the AI models. Biases can arise from many sources, including the selection of data sets, the methods used to collect the data, and the algorithms used to analyse the data. If these biases are not addressed, they can lead to inaccurate predictions or perpetuate existing inequalities [84].

Another concern is the potential for AI to replace human expertise in biochemistry research. While AI can certainly help scientists process and analyse large amounts of data more quickly than humans can, it is not a substitute for the expertise and intuition of human researchers. In addition, the use of AI in research may also raise questions about the nature of scientific discovery and the role of human creativity in scientific breakthroughs [85].

Another ethical consideration with the use of AI in biochemistry research is the potential impact on data privacy and security. As more data is collected and analysed using AI, there is a risk that sensitive information about individuals, such as genetic information or health data, could be compromised. Scientists must be diligent in their efforts to protect data privacy and security while still making use of the vast amounts of data available [81].

Finally, there is a concern about the potential misuse of AI in biochemistry research. For example, AI could be used to develop new and more dangerous biological weapons, or it could be used to perpetuate existing inequalities in access to healthcare or genetic testing. As AI technology continues to advance, it will be essential for scientists, policymakers, and society as a whole to consider the potential risks and benefits and to develop ethical guidelines for the responsible use of AI in biochemistry research [86].

Addressing ethical issues related to AI must be considered a top priority in biochemistry research. It is crucial that developers and users receive proper training and education to be aware of these issues. Moreover, it is essential to ensure diversity in the workforce to prevent exclusion and ensure that everyone benefits from the advancements in AI technology. As we continue to use AI in biochemistry, we must remain vigilant to prevent any potential misuse of this technology, which could lead to harm to individuals or the environment. It is essential to take appropriate measures to evaluate and address these ethical concerns carefully [87].

CONCLUSION

The application of AI in biochemistry research has been transformed by the way complex biological problems are approached. Vast amounts of data with incredible speed and accuracy have been analysed. New avenues of exploration and discovery that were once unimaginable have been opened up. From drug discovery to protein structure prediction, genomics to metabolomics, the impact of AI on biochemistry research has been undeniable. However, the adoption of AI in biochemistry research has not been without its challenges. Data quality and curation, algorithmic transparency, and ethical considerations are all important factors that need to be addressed as move forward. Nonetheless, the opportunities presented by AI are too significant to be ignored, and the future of biochemistry research undoubtedly lies in the integration of these powerful computational tools. As the possibilities of AI in biochemistry research continue to be explored, it is important to remember that the ultimate goal is to improve human health and well-being. By leveraging the power of AI to understand the complex processes that underlie life, more effective therapies and interventions can be developed to tackle some of the world's most pressing health challenges and further the understanding of biochemical processes.

REFERENCES

1. Nelson DL, Cox MM. Lehninger Principles of Biochemistry 8th Edition. Macmillan Learning; 2021.
2. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol. 2012; 8(2):e1002375.
3. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet. 2015; 16(6):321-32.
4. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Sci. 2015; 349(6245):255-60.
5. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Comput Sci. 2021; 2(3).
6. Min S, Lee B, Yoon S. Deep learning in bioinformatics. Brief Bioinformatics. 2017; 18(5):851-69.
7. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596(7873):583-9.
8. Matsuta Y, Ito M, Tohsato Y. ECOH: An Enzyme Commission number predictor using mutual information and a support vector machine. Bioinformatics. 2012; 29(3):365-72.
9. Baum ZJ, Yu X, Ayala PY, Zhao Y, Watkins SP, Zhou Q. Artificial intelligence in chemistry: current trends and future directions. J Chem Inf Model. 2021; 61(7):3197-212.
10. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018; 15(141).
11. AlQuraishi M. AlphaFold at CASP13. Bioinformatics. 2019; 35(22):4862-5.
12. Dasgupta A, Nath A. Classification of machine learning algorithms. Int J Innov Res Adv Eng. 2016; 3(3):6-11.
13. Liu Y, Esan OC, Pan Z, An L. Machine learning for advanced energy materials. Energ AI. 2021; 3:100049.
14. Ayodele TO. Types of machine learning algorithms. 2010; 3:19-48.
15. Han J. M. Kamber in J. Pei, Data mining: Concepts and techniques: Concepts and techniques, 3. izd. Amsterdam, Netherlands: Elsevier; 2011.
16. Sarker IH, Kayes A, Badsha S, Alqahtani H, Watters P, Ng A. Cybersecurity data science: an overview from machine learning perspective. J Big Data. 2020; 7:1-29.
17. Nasteski V. An overview of the supervised machine learning methods. Horizons Int Sci J. 2017; 4:51-62.
18. Patel L, Shukla T, Huang X, Ussery DW, Wang S. Machine learning methods in drug discovery. Molecules. 2020; 25(22):5277.
19. Ramadass S, Abraham A. Unsupervised control paradigm for performance evaluation. Int J Comput Appl. 2012; 975:8887.
20. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Mol Divers. 2021; 25:1315-60.
21. Puterman ML. Markov decision processes: discrete stochastic dynamic programming: John Wiley & Sons; 2014.
22. Xu Y, Wang C, Liang J, Yue K, Li W, Zheng S, et al. Deep Reinforcement Learning Based Decision Making for Complex Jamming Waveforms. Entropy. 2022; 24(10):1441.
23. Polydoros AS, Nalpantidis L. Survey of model-based reinforcement learning: Applications on robotics. J Intell Robot Syst. 2017; 86(2):153-73.
24. Çalışır S, Pehlivanoğlu MK, editors. Model-free reinforcement learning algorithms: A survey. 2019 27th Signal Processing and Communications Applications Conference (SIU); 2019: IEEE.
25. Bengio Y, Lecun Y, Hinton G. Deep learning for AI. Commun ACM. 2021; 64(7):58-65.
26. Nichols JA, Herbert Chan HW, Baker MA. Machine learning: applications of artificial intelligence to imaging and diagnosis. Biophys Rev. 2019; 11:111-8.
27. Hamilton WL. Graph representation learning. Synth Lect Artif Intell Mach Learn. 2020; 14(3):1-159.
28. Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY. A comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst. 2020; 32(1):4-24.
29. Ahmed KT, Park S, Jiang Q, Yeu Y, Hwang T, Zhang W. Network-based drug sensitivity prediction. BMC Med Genomics. 2020; 13(11):1-10.
30. Bove P, Micheli A, Milazzo P, Podda M, editors. Prediction of Dynamical Properties of Biochemical Pathways with Graph Neural Networks. Bioinformatics; 2020.
31. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018; 9:611-29.
32. Chauhan R, Ghanshala KK, Joshi R, editors. Convolutional neural network (CNN) for image detection and recognition. 2018 first international conference on secure cyber computing and communication (ICSCCC); 2018: IEEE.
33. Mohamed C, Nsiri B, Abdelmajid S, Brahim B, editors. Deep convolutional networks for image segmentation: application to optic disc detection. 2020 International Conference on Electrical and Information Technologies (ICEIT); 2020: IEEE.
34. Askr H, Elgeldawi E, Aboul Ella H, Elshaier YA, Gomaa MM, Hassanien AE. Deep learning in drug discovery: an integrative review and future challenges. Artif Intell Rev. 2022:1-63.
35. Zia T, Zahid U. Long short-term memory recurrent neural network architectures for Urdu acoustic modeling. Int J Speech Technol. 2019; 22:21-30.
36. Eliasy A, Przychodzen J. The role of AI in capital structure to enhance corporate funding strategies. Array. 2020; 6:100017.
37. Hewamalage H, Bergmeir C, Bandara K. Recurrent neural networks for time series forecasting: Current status and future directions. Int J Forecast. 2021; 37(1):388-427.
38. Vickram A, Kamini AR, Das R, Pathy MR, Parameswari R, Archana K, et al. Validation of artificial neural network models for predicting biochemical markers associated with male infertility. Syst Biol Reprod Med. 2016; 62(4):258-65.
39. Aggarwal A, Mittal M, Battineni G. Generative adversarial network: An overview of theory and applications. Int J Inf Manag Data Insights. 2021; 1(1):100004.
40. Repecka D, Jauniskis V, Karpus L, Rembeza E, Rokaitis I, Zrimec J, et al. Expanding functional protein sequence spaces using generative adversarial networks. Nat Mach Intell. 2021; 3(4):324-33.
41. Kazeminia S, Baur C, Kuijper A, van Ginneken B, Navab N, Albarqouni S, et al. GANs for medical image analysis. Artif Intell Med. 2020; 109:101938.
42. Khurana D, Koli A, Khatter K, Singh S. Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl. 2023; 82(3):3713-44.
43. Thessen AE, Cui H, Mozzherin D. Applications of Natural Language Processing in Biodiversity Science. Adv Bioinformatics. 2012; 2012:391574.
44. Chen X, Xie H, Cheng G, Poon LKM, Leng M, Wang FL. Trends and Features of the Applications of Natural Language Processing Techniques for Clinical Trials Text Analysis. Appl Sci. 2020; 10(6):2157.
45. Gallego V, Naveiro R, Roca C, Ríos Insua D, Campillo NE. AI in drug development: a multidisciplinary perspective. Mol Divers. 2021; 25:1461-79.
46. Lin E, Lin C-H, Lane H-Y. Relevant applications of generative adversarial networks in drug design and discovery: molecular de novo design, dimensionality reduction, and de novo peptide and protein design. Molecules. 2020; 25(14):3250.
47. Grisoni F, Moret M, Lingwood R, Schneider G. Bidirectional molecule generation with recurrent neural networks. J Chem Inf Model. 2020; 60(3):1175-83.
48. Pogány P, Arad N, Genway S, Pickett SD. De novo molecule design by translating from reduced graphs to SMILES. J Chem Inf Model. 2018; 59(3):1136-46.
49. Kell DB, Samanta S, Swainston N. Deep learning and generative methods in cheminformatics and chemical biology: navigating small molecule space intelligently. Biochem J. 2020; 477(23):4559-80.
50. Prykhodko O, Johansson SV, Kotsias P-C, Arús-Pous J, Bjerrum EJ, Engkvist O, et al. A de novo molecular generation method using latent vector based generative adversarial network. J Cheminform. 2019; 11(1):1-13.
51. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow Jr RA, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020; 19(5):353-64.

52. Kolluri S, Lin J, Liu R, Zhang Y, Zhang W. Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: a Review. AAPS J. 2022; 24(1):19.
53. Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput Struct Biotechnol J. 2020; 18:784-90.
54. Cheng F, Kovács IA, Barabási A-L. Network-based prediction of drug combinations. Nat Commun. 2019; 10(1):1197.
55. Dey S, Luo H, Fokoue A, Hu J, Zhang P. Predicting adverse drug reactions through interpretable deep learning framework. BMC Bioinformatics. 2018; 19(21):1-13.
56. Parvathaneni M, Awol AK, Kumari M, Lan K, Lingam M. Application of Artificial Intelligence and Machine Learning in Drug Discovery and Development. J Drug Deliv Ther. 2023; 13(1):151-8.
57. Kendrew JC. The three-dimensional structure of a protein molecule. Sci Am. 1961; 205(6):96-111.
58. Dill KA, Ozkan SB, Shell MS, Weikl TR. The Protein Folding Problem. Annu Rev Biophys. 2008; 37(1):289-316.
59. Service RF. 'The game has changed.'AI triumphs at protein folding. American Association for the Advancement of Science; 2020.
60. Al-Janabi A. Has DeepMind's AlphaFold solved the protein folding problem? : Future Science; 2022.
61. Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. High-accuracy protein structure prediction in CASP14. Proteins. 2021; 89(12):1687-99.
62. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020; 577(7792):706-10.
63. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022; 50(D1):D439-D44.
64. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics. 2013; 29(21):2722-8.
65. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Sci. 2021; 373(6557):871-6.
66. Pandey D, Onkara Perumal P. A scoping review on deep learning for next-generation RNA-Seq. data analysis. Funct Integr Genomics. 2023; 23(2):134.
67. Mathema VB, Sen P, Lamichhane S, Orešič M, Khoomrung S. Deep learning facilitates multi-data type analysis and predictive biomarker discovery in cancer precision medicine. Comput Struct Biotechnol. 2023.
68. Schmidt B, Hildebrandt A. Deep learning in next-generation sequencing. Drug Discov Today. 2021; 26(1):173-80.
69. Corsaro C, Vasi S, Neri F, Mezzasalma AM, Neri G, Fazio E. NMR in Metabolomics: From Conventional Statistics to Machine Learning and Neural Network Approaches. Appl Sci. 2022; 12(6):2824.
70. Wei W, Liao Y, Wang Y, Wang S, Du W, Lu H, et al. Deep Learning-Based Method for Compound Identification in NMR Spectra of Mixtures. Molecules. 2022; 27(12).
71. Hennessy A, Clarke K, Lewis M. Generative Adversarial Network Synthesis of Hyperspectral Vegetation Data. Remote Sens. 2021; 13(12):2243.
72. Cox J. Prediction of peptide mass spectral libraries with machine learning. Nat Biotechnol. 2023; 41(1):33-43.
73. Chen C, Hou J, Tanner JJ, Cheng J. Bioinformatics methods for mass spectrometry-based proteomics data analysis. Int J Mol Sci. 2020; 21(8):2873.
74. Helmy M, Smith D, Selvarajoo K. Systems biology approaches integrated with artificial intelligence for optimized metabolic engineering. Metab Eng Commun. 2020; 11:e00149.
75. You Y, Lai X, Pan Y, Zheng H, Vera J, Liu S, et al. Artificial intelligence in cancer target identification and drug discovery. Signal Transduct Target Ther. 2022; 7(1):156.
76. Kaur C, Garg U. Artificial intelligence techniques for cancer detection in medical image processing: A review. Mater Today Proc. 2021.
77. Keenan TD, Clemons TE, Domalpally A, Elman MJ, Havilio M, Agrón E, et al. Retinal specialist versus artificial intelligence detection of retinal fluid from OCT: age-related eye disease study 2: 10-year follow-on study. Ophthalmology. 2021; 128(1):100-9.
78. Belkacem AN, Ouhbi S, Lakas A, Benkhelifa E, Chen C. End-to-end AI-based point-of-care diagnosis system for classifying respiratory illnesses and early detection of COVID-19: a theoretical framework. Front Med. 2021; 8:585578.
79. Zhang H, Zheng J. The Application Analysis of Medical Chatbots and Virtual Assistant. Front Soc Sci Technol. 2021; 3:11-6.
80. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearb Med Inform. 2019; 28(01):128-34.
81. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. Am J Med. 2019; 132(7):795-801.
82. John-Mathews J-M. Some critical and ethical perspectives on the empirical turn of AI interpretability. Technol Forecast Soc Change. 2022; 174:121209.
83. Ying X, editor An overview of overfitting and its solutions. Journal of physics: Conference series; 2019: IOP Publishing.
84. Lo Piano S. Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. Hum Soc Sci Commun. 2020; 7(1):1-7.
85. Chesterman S. Artificial intelligence and the problem of autonomy. Notre Dame J Emerg Tech. 2020; 1:210.
86. Lentzos F. AI and Biological Weapons. Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm: Springer; 2022. p. 91-100.
87. Hassoun S, Jefferson F, Shi X, Stucky B, Wang J, Rosa Jr E. Artificial intelligence for biology. Integr Comp Biol. 2021; 61(6):2267-75.