

Article

Not peer-reviewed version

---

# Role of Artificial Intelligence in Modern Drug Discovery: A Literature Overview

---

[Aisha Farooq](#) and [Rizwan Ayaz](#) \*

Posted Date: 23 March 2026

doi: 10.20944/preprints202603.1558.v1

Keywords: drug discovery; artificial intelligence; machine learning; molecular property prediction; AlphaFold



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

# Role of Artificial Intelligence in Modern Drug Discovery: A Literature Overview

Aisha Farooq<sup>1</sup> and Rizwan Ayaz<sup>2,\*</sup>

<sup>1</sup> Ziauddin University, Karachi, Pakistan

<sup>2</sup> Taylor's University, Subang Jaya, Malaysia

\* Correspondence: rizayazuddin@gmail.com

## Abstract

Bringing a new drug to today's market is extremely costly and time-consuming. To accelerate this process, the pharmaceutical industry has turned to Artificial Intelligence (AI). In addition to AI, machine learning (ML), deep learning (DL), and big data analytics have emerged as transformative tools that have improved the efficacy and accuracy. AI helps in the analysis of molecular structures and in the evaluation of in vivo and in vitro characteristics without putting human or animal lives at risk. AI applications in drug development include peptide synthesis, ligand-based virtual screening, toxicity prediction, drug repositioning, pharmacophore modeling, quantitative structure–activity relationships, polypharmacology, and drug release monitoring. Although AI involvement and application in drug discovery are still in their early stages, it has the potential to revolutionize this field entirely. In recent years, AI has enabled researchers to solve complex problems such as designing drugs with low toxicity, identifying targets for difficult diseases, and developing drugs with improved efficacy. However, AI implementation in drug discovery still faces notable gaps and limitations, including data quality issues, poor interpretability of AI-generated models, ethical considerations, and the need to validate AI-generated predictions through experimental studies. The findings reveal that AI, Machine learning, and Deep learning have advanced multiple stages of drug discovery by improving target identification, toxicity prediction, and molecular design, and have the potential to increase the success rate of drug discovery. In this paper, we discuss conventional methods of drug development and then examine how artificial intelligence and deep learning are changing the process of drug discovery.

**Keywords:** drug discovery; artificial intelligence; machine learning; molecular property prediction; AlphaFold

---

## 1. Introduction

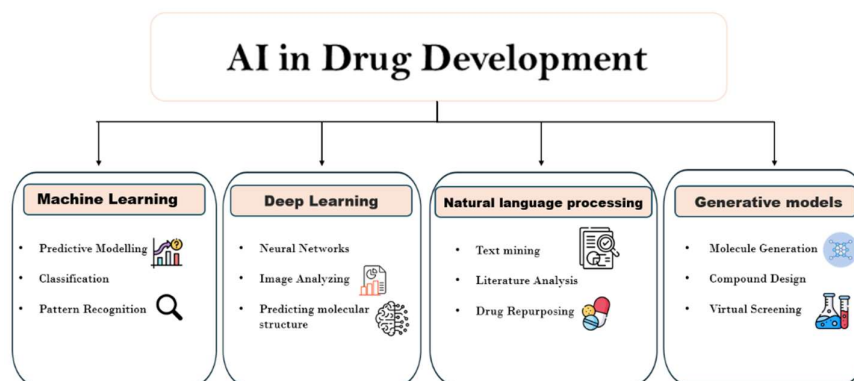
The drug discovery process is a complex and time-consuming process. The procedure begins with target identification, where researchers first identify the biological molecule that plays the main role in a disease, and then target validation is done to confirm that modifying this target will produce a therapeutic effect [1]. After the target is confirmed, researchers screen a large number of chemical compounds to identify molecules that can interact with the target [2]. Further refinement of promising compounds known as hits, which are converted to lead compounds through chemical modification and laboratory testing.

Preclinical studies are then carried out to evaluate their safety and biological activity in laboratory models. If the results are positive, the drug candidate proceeds to the next stage, clinical trials, where it is tested in humans to evaluate its safety, dosing, and efficacy. Depending on the results of clinical trials, the drug may receive regulatory approval and become available for medical use. Traditional methods are slow, costly, and yield results that are of low accuracy [3].

The development of a single drug may take up to a decade, and the cost goes up to \$2.6 billion per drug [4]. Drug development has a 99.6% failure rate. For example, for Alzheimer's disease(AD), 244 compounds entered the clinical development stage, but only one drug, memantine, received

regulatory approval. This shows that traditional drug discovery is inefficient and has a high clinical trial failure rate.

Although these methods were successful in the past, their success rate has always been limited due to their dependence on trial-and-error methods. The term Artificial Intelligence (AI) was coined by John McCarthy at a conference at Dartmouth College in 1956. AI gained popularity in the 1980s with the development of systems that summarized rules derived from expert knowledge to help non-specialists make decisions. Examples include XCON, designed by Carnegie Mellon University, and MYCIN, designed by Stanford University [5]

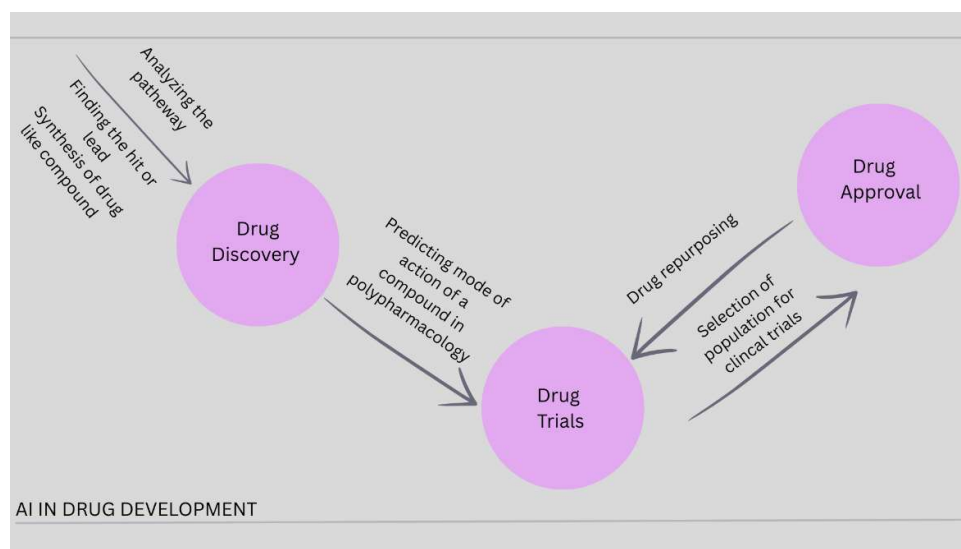


**Figure 1.** AI in Drug Development Overview.

Artificial intelligence (AI) in recent years has played a significant role in the betterment of the pharmaceutical industry. Conventional data storage methods are becoming obsolete as data volumes grow rapidly. This big data can be used for deeper research and improved development through data mining [6].

Figure 2. Represents key artificial intelligence technologies utilized in drug development in modern drug development. Machine learning (ML) and natural language processing (NLP) can help speed up the process, which results in more efficient and accurate analysis of a large amount of data. There have been studies on the use of GPU computing and deep learning models in accelerating the drug discovery process [7].

Advancements in technologies such as next-generation sequencing, medical imaging, and electronic health records have resulted in vast amounts of biomedical data that need sophisticated tools for interpretation. This huge amount of data cannot be handled through traditional statistical methods. AI-based models can analyze genomic sequences to identify disease-related genes, mutations, and any biological pathways linked with various disorders. For example, deep learning algorithms have been used to predict protein structures and interactions [8]. Its applications in medical imaging and disease diagnosis have also been highly appreciable. Deep learning techniques, specifically convolutional neural networks (CNNs), are capable of analyzing complex imaging data such as MRI, CT scans, and histopathological images. This can also assist clinicians in the detection of diseases such as cancer and neurological conditions with improved accuracy and efficiency [9]. AI systems and specialized software are used to analyze medical data, make decisions, and perform tasks that were previously carried out by healthcare professionals.



**Figure 2.** AI in Drug Development pipeline.

Figure 2. maps out three main stages of drug development from drug discovery through trials to approval. At various steps of drug discovery, AI helps us save time and increase efficiency.

These include its use in real-time cell sorting and classification, calculation of compound properties using quantum mechanics (QM), development of new compounds, computational organic synthesis, and many others. During the preclinical drug discovery, the main objective is to identify at least one clinical candidate molecule that shows biological activity against a disease-related target while having safe and drug-like properties for human testing, but most of the time these compounds fail due to problems such as toxicity, poor pharmacokinetics, or insufficient potency [10]. There are huge data sets that are generated from genomic, proteomics, clinical trials, and electronic health records that need to be interpreted accurately, and this can be done through advanced analytical tools. This is where Artificial Intelligence and big data technologies provide us with a powerful tool to analyze these complex datasets and to extract data that will support drug development. A project known as EHR4CR was developed to make use of hospital EHR data for medical research purposes [11]. Artificial intelligence is being used for analysis of epigenetic data, which is essential for understanding cancer mechanisms and for developing targeted treatments [12]

Artificial intelligence is also being used to make advances in surgical technologies. Example Da Vinci Surgical System developed by Intuitive Surgical. This was approved by the U.S. Food and Drug Administration in 2000 and has been used since in hospitals worldwide. This is commonly used for procedures such as prostatectomy and various gynecological surgeries [13]. Tools like NLP (Natural Language Processing) and OCR (Optical Character Recognition) can scan online databases, trial registries, and social media to match patients with suitable clinical trials. AI increases efficiency and reach. There has also been rapid growth in medical literature, which makes it difficult for healthcare professionals to stay updated on the latest research. A physician needs to spend nearly 29 hours per day to remain informed about the current medical findings. This is where AI helps, for example, IBM Watson, an AI-based system designed to analyze and interpret vast amounts of medical literature and clinical data [14].

The active participation of doctors and healthcare professionals can increase AI's potential to improve patient care and advance the pharmaceutical industry. This article presents a comprehensive study of the latest technologies, recent research advances, and the future impact of AI on the pharmaceutical industry. Both the opportunities and limitations are examined to provide balanced information regarding its role in shaping the future of drug development.

In this research our aim is to discuss how Artificial Intelligence (AI) is improving the drug discovery and development process. Since traditional methods are slow, expensive, and result in

frequent failures. This study explains how researchers are turning to AI tools to process vast amounts of medical data more effectively.

The subsequent sections review the key areas of AI-driven drug discovery from biological data and omics analysis to drug design tools and existing challenges.

Core contributions of the study:

- This research provides an overview of the limitations of conventional drug discovery methods.
- In this research we discuss how AI-driven techniques, including machine learning, deep learning and natural language processing, allow researchers to rapidly analyze large biomedical datasets.
- We also examines how these technologies improve target identification, virtual screening and refine lead compounds.
- We conduct a computational case study in this research using machine learning to predict molecular solubility of selected drug molecules to show and analyze the practical application of AI techniques in evaluating physicochemical properties of potential drug candidates.
- Our obtained experimental results provide quantitative evidence of model performance, where the Gradient Boosting model achieved the best prediction accuracy (RMSE = 0.787 and  $R^2 = 0.869$ ).
- We performed a case study analysis also where we show how AI-based prediction models can assist in screening and evaluating drug molecules such as Aspirin, Ibuprofen, Paracetamol, Metformin, and Remdesivir for supporting early-stage decision-making in drug discovery.

## 2. Literature Review

The limitations of traditional drug discovery methods have led researchers to explore computational approaches to enhance efficiency and accuracy. In recent years, artificial intelligence (AI) has significantly accelerated various stages of drug development. Machine learning is a fundamental branch of Artificial Intelligence that allows computers to learn patterns from data and generate predictions without the need for explicit programming.

Natural language processing (NLP), a specialized branch of AI, enables computers to understand, process, and interpret human languages in a meaningful way. In drug discovery, NLP is used to extract valuable information from biomedical literature, scientific research, clinical trial reports, and extensive medical databases.

AI technologies can analyze extensive biomedical datasets, recognize patterns, and predict molecular properties, guiding researchers in identifying promising therapeutic compounds. Figure 3 presents a comprehensive overview of how Artificial Intelligence operates in drug discovery pipelines.

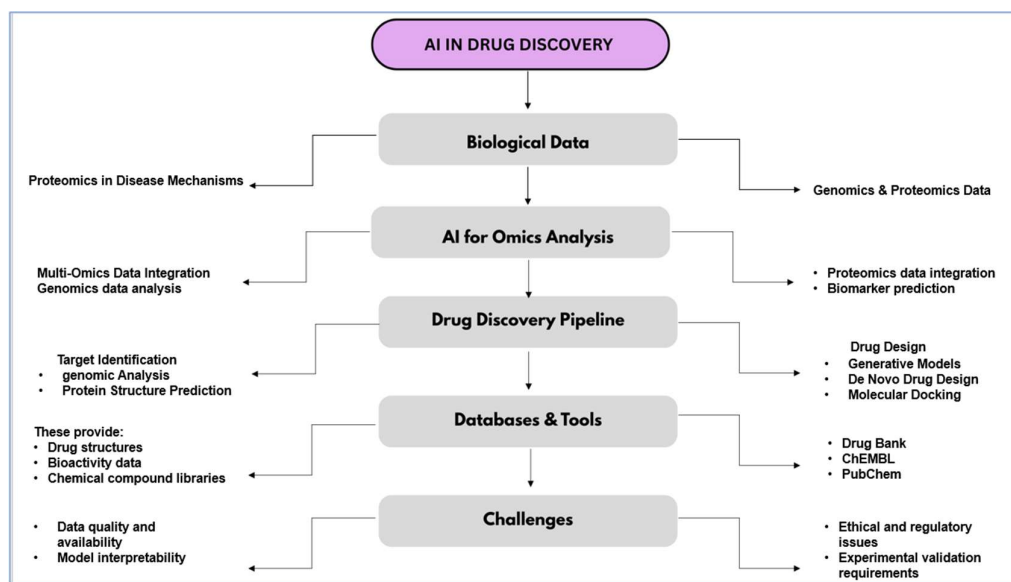


Figure 3. Key Components of AI Technologies in Drug Discovery and Development.

The foundational step in drug discovery is the identification of biological targets, as it determines the molecular components that result in disease development. Advances in genomic technologies have enabled researchers to analyze genetic variations and mutations associated with diseases. Computational and in-silico methods are increasingly used to analyze genomics datasets and shortlist candidate targets for further investigation [15]. Numerous AI applications have been introduced into the drug discovery pipeline, ranging from target identification to drug repurposing, as summarized in Table 1.

Table 1. AI Application Across Drug Development.

Drug Discovery Stage	AI Application	Description	References
Target Identification	Genomic and proteomic data analysis	AI models analyze large-scale genomic and proteomic datasets to identify disease-associated genes, proteins, and potential therapeutic targets.	[34,35]
Target Validation	Biological pathway analysis	Machine learning algorithms validate targets by predicting interactions taking place between genes, proteins, and diseases.	[36]
Hit Identification	Virtual screening	AI screens large chemical libraries this saves time and cost.	[37]
Lead Optimization	Molecular property prediction	Deep learning models predict molecular properties that enhance drug efficacy and selectivity.	[38]
Drug Design	De novo molecule generation	Generative AI designs novel drug molecules with desired pharmacological properties.	[39]
Toxicity Prediction	ADMET prediction	AI evaluates absorption, distribution, metabolism, excretion, and toxicity profiles to improve safety.	[40]
Clinical Trials	Patient recruitment and trial optimization	AI analyzes patient data and electronic health records to recruit suitable participants.	[41]
Drug Repurposing	Identification of new therapeutic uses	AI analyzes biomedical literature and drug databases.	[42]

### 2.1. *The Role of Proteomics in Disease Mechanisms and Drug Discovery.*

Proteomics has emerged as an important tool for understanding disease mechanisms by studying protein expression, interactions, and functions within biological systems. It allows researchers to identify protein targets of small molecules and map out how drugs interact with biological pathways, which provides them with critical insights regarding drug development [16]. Large-scale genomic and proteomic studies have been highly effective in the identification of biomarkers and potential therapeutic targets linked with complex diseases. These studies have combined multiple data sources, which reveal novel molecular interactions, regulatory networks, and signaling pathways that would otherwise go undetected [17]. Likewise, both genome-wide association studies (GWAS) and proteomic data have elucidated proteins associated with cardiovascular diseases, providing insights into biological pathways that are involved in disease progression [18]. Pan-cancer proteogenomic studies have further increased our understanding of tumor biology by linking genomic alteration to protein expression profiles, supporting the discovery of novel cancer therapeutic targets [19].

### 2.2. *Artificial Intelligence in Multi-Omics Data Analysis*

Artificial intelligence (AI) and Machine learning (ML) methods are increasingly being used for the analysis of large multi-omics datasets and uncovering complex biological relationships. These approaches improve the process of drug discovery by enabling deeper exploration of complex biological systems [20].

Experimental validation is a critical step in modern scientific research as it confirms if computational models, theoretical predictions, or analytical methods actually represent the real-world outcomes. Many recent studies combine computational techniques with laboratory experimentations to improve their reliability and reduce development costs. For instance, computational screening approaches have been used to predict pharmaceutical co-crystals, where predicted candidates are tested later on to confirm that they are stable and formed as expected [21]. In biomedical engineering, numerical models that describe fluid flow behavior in an organ-on-a-chip system require experimental evaluation to confirm that stimulated microfluidic behavior reflects its real biological condition [22].

Pharmaceutical and biomedical research, network pharmacology predictions are often confirmed through laboratory experiments to confirm therapeutic mechanisms, as seen in research exploring traditional herbal formulations for colorectal cancer treatment [23].

In drug discovery computation approaches are mainly used, such as molecular docking and virtual screening, to identify potential therapeutic compounds paired with computational validation to confirm the stability and interaction of predicted molecules [24]. Moreover, computational techniques used in medical prediction models often depend on dataset splitting and cross-validation to confirm their performance and improve prediction reliability [25].

### 2.3. *De Novo Drug Design Using Artificial Intelligence*

De novo design generates entirely new molecular structures with desired pharmacological properties. Generative models based on AI include neural networks (RNNS), Transformer architectures, and deep learning frameworks, which can analyze large chemical datasets and identify structural patterns that allow them to formulate new candidate molecules [26].

Such computational approaches allow researchers to design molecules that meet the key criteria, such as potency, toxicity, and synthetic feasibility [27]. These also enable the discovery of new molecular scaffolds that may lead to new bioactive compounds [28].

The success of artificial intelligence (AI) in drug discovery mainly depends on the availability of large and organized chemical and biological datasets.

#### 2.4. Databases in AI-Driven Drug Discovery

One of the most commonly used resources is ChEMBL, a database that contains bioactive molecules with experimentally measured biological activity data that strengthens drug discovery research [29]. Another widely used database is PubChem, which serves as a large public repository of chemical compounds, molecular structures, and bioassay data extensively applied in computational screening and chemical evaluation [30].

In the same way, Drug Bank integrates detailed chemical, pharmacological, and pharmaceutical information about approved and experimental drugs, making it a prominent resource for drug repurposing and computational modeling studies [31].

Comparative studies of molecular databases such as PubChem, ChEMBL, and Drug Bank have demonstrated their importance and significance in providing structured chemical libraries for virtual screening and computational drug designing [32].

Drug repurposing, also referred to as drug repositioning, is the identification of new therapeutic uses for existing drugs. This focuses on drugs that are already approved or have already cleared early safety testing. This reduces development risks and shortens the timeline for therapeutic applications.

AI enables rapid screening of large chemical libraries, allowing researchers to identify promising compounds, helps predict molecular properties, toxicity and drug-target interactions with high accuracy, supports drug repurposing and generative AI techniques facilitate the design of novel molecules through de novo drug design.

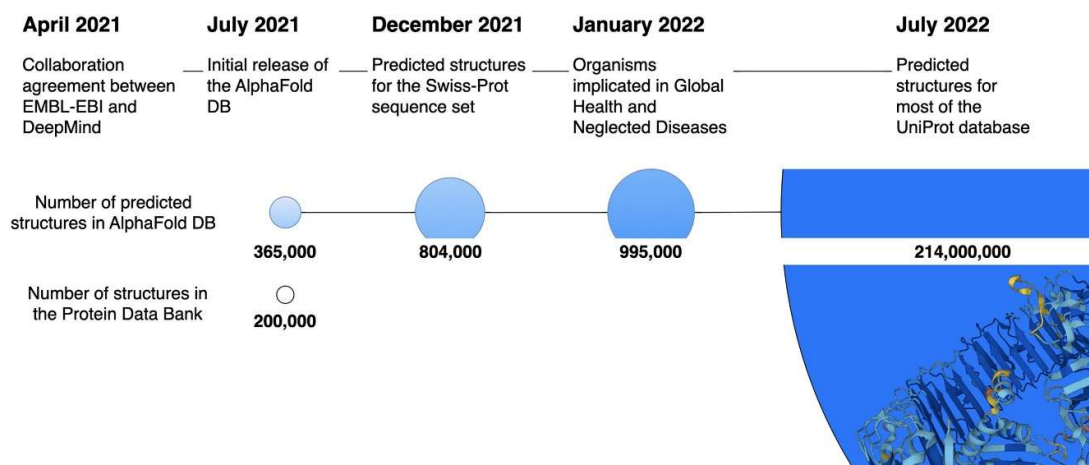
#### 2.5. Challenges of AI-Driven Drug Discovery

Despite the numerous advantages of artificial intelligence, a range of limitations exist that challenges its widespread implementation. One of the most encountered challenges in AI-based drug discovery is the lack of availability and quality of data used to train models. Machine learning and deep learning systems are only as reliable as the data they are built on; if the training datasets are incomplete, then the reliability and predictive accuracy of AI systems are negatively affected. In biomedical research, this problem is more serious because biological and clinical datasets are mostly incomplete, inconsistent or biased. Further challenge is the problem of publications bias. Studies showing positive results are often published, while those with negative results are frequently omitted from the literature. Consequently, the resulting models may produce inaccurate outcomes [33]

### 3. Case Study

One of the biggest challenges in molecular biology has long been figuring out the 3D shape of proteins often referred to as the “protein folding problem” Traditional experimental techniques such as X-ray, crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (Cryo-EM) have been widely used to solve these structures but they are time consuming, expensive and technically demanding.

The genetic codes were well documented, but what these proteins actually looked like in 3D remained largely unknown. This is where Alpha Fold stepped in and changed everything. AlphaFold works by applying deep learning algorithms to predict these 3D protein structures directly from amino acid sequences. What once took years can now be done in hours or even minutes on a computer.



**Figure 4.** AlphaFold's rapid rise in protein structure prediction [43].

As shown in Figure 4, the journey of Alpha fold from a specialized research tool to a broadly accessible scientific platform. It also highlights the remarkable capacity of AI to generate structural data in comparison to traditional laboratory methods. AlphaFold is able to generate hundreds of millions of predictions within a short time frame.

The partnership between DeepMind and EMBL-EBL was a game changer; it made Alpha fold's predictions freely available, which meant that even researchers from smaller institutions and developing countries had the same access to cutting edge protein structural data as those working in well-funded laboratories in the world. The drug discovery process is dependent on knowing the detailed structure of proteins.

AlphaFold has the ability to fill in the gaps as it provides structural data for proteins that were previously uncharacterized. Scientists can now utilize computational tools to test and refine potential drug candidates virtually, thereby reducing both the time and financial investment required in the early stages of drug development.

AlphaFold's key output is called "Predicted Aligned Error" or PAE, which measures the expected error in the predicted portions of different parts of a protein. It is an error score that tells how certain Alpha Fold is about where different parts of the protein sit in relation to one another [43].

Despite these achievements, AlphaFold has several limitations like any other scientific tool, its predictions must be carefully interpreted. One of the most significant limitations of AlphaFold relates to the kind of structures it is designed to predict. AlphaFold generally predicts static protein structure, which is an impressive achievement[44-46], but it only captures a fraction of what proteins actually do. Proteins are not rigid and unchanging; they are dynamic and flexible molecules that change shape depending on their surrounding environment. This is called dynamic conformational changes; these are deeply meaningful events that define how a protein behaves, interacts, and fulfills its purpose in the body. These movements play a key role in ligand binding, enzyme catalysis, and protein-protein interactions.

Another limitation is that it faces difficulty in predicting how protein complexes and multi-protein assemblies work together. Most biological processes rely on interactions taking place between multiple proteins, and accurately modeling these assemblies remains a challenge.

AlphaFold-Multimer was developed to address this problem, but its accuracy is not always consistent, and the prediction depends on the complexity of the system[47-49]. It also does not fully incorporate post-translational modifications such as phosphorylation or glycosylation, which influence the shape and function of the protein.

Whatever the shortcomings may be, there is no denying that the advantages of AlphaFold outweigh its current limitations. Its capacity to rapidly generate high-quality structural prediction

has accelerated the progress across multiple disciplines like enzymology, molecular biology, and pharmaceutical sciences.

## 4. Methodology

In this research we adopt a computational experimental approach, and we combine this with the above discussed case study analysis to establish and demonstrate how machine learning techniques can assist in predicting molecular properties relevant to drug discovery[51-53]. The methodology consists of four main stages: dataset preparation, feature extraction, model training and evaluation, and case study analysis[54-56].

### 4.1. Dataset Preparation

The experiment was conducted using the ESOL (Delaney) dataset, which contains molecular structures represented as SMILES strings along with experimentally measured solubility values expressed as log solubility in mols per litre. The dataset is widely used in cheminformatics research for evaluating molecular property prediction models.

From the dataset, two relevant fields were selected: the SMILES representation of molecules and the measured log solubility values. Entries containing missing values were removed to ensure the integrity of the training data.

### 4.2. Feature Extraction

Molecular descriptors were extracted from the SMILES strings using the RDKit cheminformatics library. These descriptors represent physicochemical properties of molecules that influence their behavior in biological environments. The following descriptors were used as input features for the machine learning models:

- Molecular Weight (MolWt)
- Octanol–Water Partition Coefficient (LogP)
- Topological Polar Surface Area (TPSA)
- Number of Hydrogen Bond Donors
- Number of Hydrogen Bond Acceptors
- Number of Rotatable Bonds
- Ring Count

These descriptors provide structural and chemical information about molecules and are commonly used in computational drug discovery tasks.

### 4.3. Model Development and Evaluation

To predict molecular solubility, three regression-based machine learning models were implemented:

- Random Forest Regressor
- Gradient Boosting Regressor
- Support Vector Regressor (SVR)

We divided the dataset into training and testing sets using an 80:20 split to evaluate model performance. We used the molecular descriptors to train each model as input features and the measured solubility values as the target variable.

Model performance was assessed using two evaluation metrics:

- Root Mean Square Error (RMSE): This metric measures the average magnitude of prediction errors

- Coefficient of Determination ( $R^2$  score): This metric indicates the proportion of variance in the target variable explained by the model

The model with the lowest RMSE and highest  $R^2$  score was selected as the best-performing model.

#### 4.4. Case Study Analysis

The practical application of the trained model was explained using a case study approach. This was done using several commonly known drugs: Aspirin, Ibuprofen, Paracetamol, Metformin, and Remdesivir. The molecular structures of these drugs were retrieved from the PubChem database using the PubChemPy library.

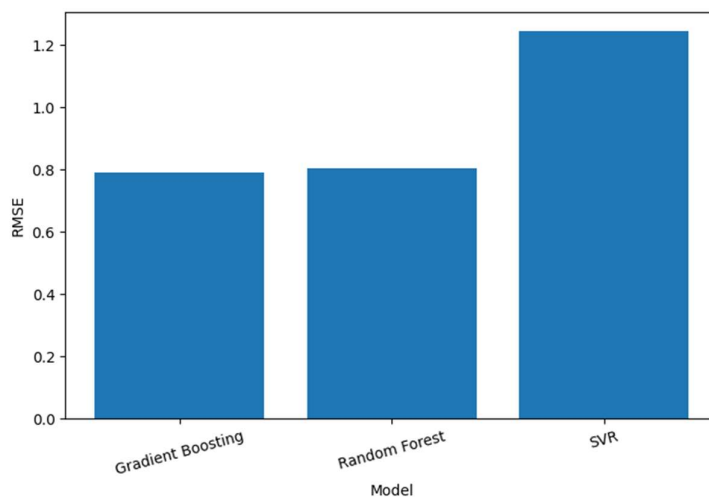
For each drug molecule, molecular descriptors were calculated using RDKit in the same manner as the training dataset. We used the train ML models to predict the solubility values for these drugs (section 5 the results are presented). After visualizing the predicted results, we discussed in section 6, how machine learning models can support early-stage screening of drug molecules.

## 5. Results

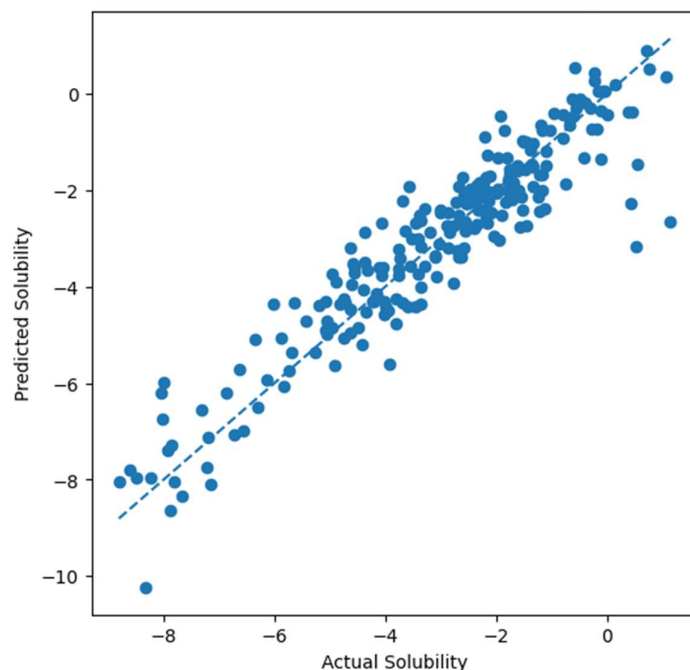
We evaluated all three machine learning models to determine their effectiveness in predicting molecular solubility. In table 2 the performance of the models based on RMSE and  $R^2$  score is shown. Figure 5 the comparison of the models based on Root Mean Square Error (RMSE) is shown. In figure 6 the relationship between the actual and predicted solubility values for the Gradient Boosting model.

**Table 2.** The Models Evaluation Result.

Model	RMSE	$R^2$ Score
Gradient Boosting	0.787	0.869
Random Forest	0.804	0.863
SVR	1.243	0.673



**Figure 5.** Model Comparison by RMSE.



**Figure 6.** Actual vs Predicted Gradient Boosting Model.

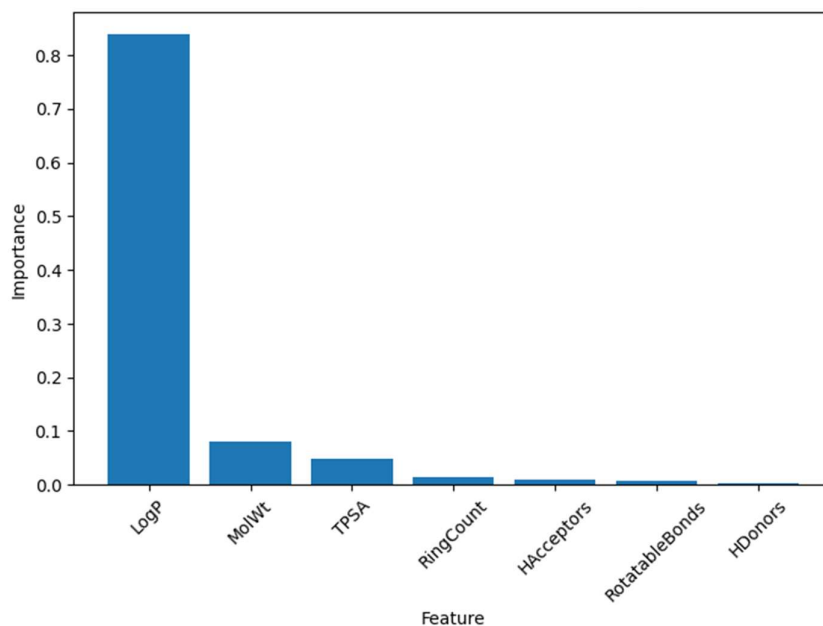
Among the evaluated models, Gradient Boosting Regressor achieved the best performance, producing the lowest RMSE value of 0.787 and the highest  $R^2$  score of 0.869. The results clearly show and indicate that the model is able to explain approximately 86.9% of the variance in molecular solubility, demonstrating strong predictive capability. The Random Forest model also performed well with similar performance metrics. The Support Vector Regressor however, performed comparatively lower in terms of accuracy as seen from the RMSE and  $R^2$  score.

### 5.1. Feature Importance Analysis

We performed feature importance analysis which helped us understand which features the model considers the most influential when making predictions. The feature importance results of gradient boosting algorithm are shown in table 3. Figure 6 shows feature importance visualization bar chart.

**Table 3.** Feature Importance Results of Gradient Boosting.

Feature	Importance Score
LogP	0.837791
MolWt	0.080135
TPSA	0.049242
RingCount	0.014979
HAcceptors	0.009607
RotatableBonds	0.006056
HDonors	0.002190



**Figure 6.** Feature Importance Gradient Boosting.

As seen from the table and graph LogP is the most influential feature. It contributes 83.8% of the predictive power of the model. LogP represents the hydrophobicity of a molecule and is known to strongly influence solubility and drug absorption characteristics. Other descriptors such as molecular weight (MolWt) and topological polar surface area (TPSA) also contribute to the prediction but to a lesser extent.

### 5.2. Case Study Results

We applied the GB model to predict the solubility of selected drugs retrieved from the PubChem database. The predicted solubility values are presented in table 4 and figure 7.

**Table 4.** Predicted Log Solubility values (GB Model only).

Drug	Predicted Log Solubility
Aspirin	-2.02
Ibuprofen	-3.40
Paracetamol	-1.86
Metformin	-0.99
Remdesivir	-4.76

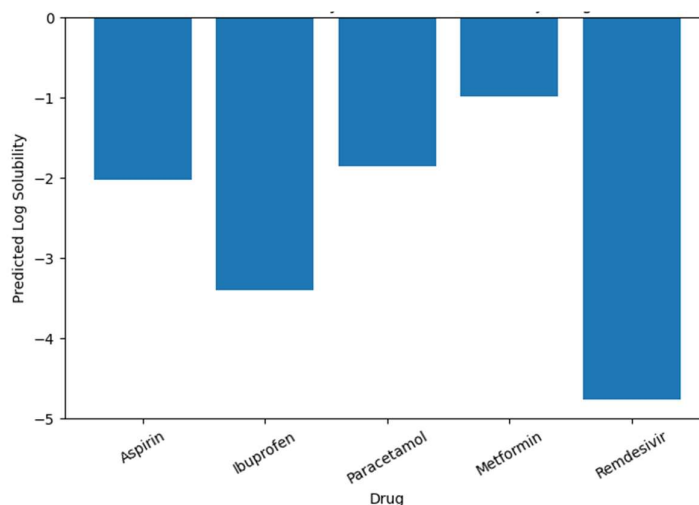


Figure 7. Predicted Solubility of Selected Case Study Drugs.

## 6. Discussion

The obtained results in this research support the growing importance of artificial intelligence in modern drug discovery. The experimental findings show that machine learning models are capable of learning relationships between molecular descriptors and physicochemical properties such as solubility. The results indicate that LogP contributes the majority of the predictive power of the model. This observation is consistent with existing pharmaceutical research, which identifies hydrophobicity as a key factor influencing molecular solubility, membrane permeability, and drug absorption. Other descriptors, including molecular weight and topological polar surface area (TPSA), also contribute to the prediction process, reflecting the importance of molecular size and polarity in determining solubility behavior. The predicted solubility values show clear differences among the analyzed compounds. For instance, Remdesivir exhibits the lowest predicted solubility, while Metformin shows comparatively higher solubility.

Modern drug discovery involves processing vast amounts of molecular information derived from genomics, proteomics, and chemical databases. Traditional experimental approaches alone are often insufficient to handle this level of complexity. Artificial intelligence enables researchers to identify patterns, predict molecular properties, and prioritize promising compounds more efficiently.

## 7. Conclusion

This study concludes that Artificial Intelligence has emerged as a transformative tool in the field of drug discovery. AI accelerates drug design through generative models, de novo molecule creation, and molecular docking. AI has proven capable of analyzing huge biomedical datasets, predicting how molecules behave, finding potential drug targets, and speeding up the drug development process.

Present research trends highlight the incorporation of deep learning, generative AI, and machine learning algorithms into the drug discovery pipeline. Researchers are progressively using neural networks, reinforcement learning, and transformer-based models to form new drug molecules and improve their pharmacological techniques.

Even with impressive advancements, several challenges remain in the application of AI in drug discovery. A majority of AI models remain dependent on large, high-quality datasets, which are not always available for all diseases or molecular targets.

A persistent limitation is the interpretability of AI models, since these complex deep learning systems tend to function as black boxes, which makes it difficult to understand the reasoning behind their predictions. Beyond this, the shift from computational predictions to experimental validation

remains a fundamental challenge that can limit the practical application of AI-generated drug candidates.

This research ultimately affirms that AI-driven approaches have the potential to significantly increase the success rate of drug development, reducing both time and cost, paving the way for a new era of effective and precisely targeted medical treatment.

## References

1. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18(6), 463–477 <https://www.nature.com/articles/s41573-019-0024-5>.
2. Schneider, P., Walters, W. P., Plowright, A. T., et al. (2020). Rethinking drug design in the artificial intelligence era. *Nature Reviews Drug Discovery*, 19(5), 353–364 <https://www.nature.com/articles/s41573-019-0050-3>.
3. Pushpakom, S., Iorio, F., Eyers, P. A., et al. (2019). Drug repurposing: progress, challenges, and recommendations. *Nature Reviews Drug Discovery*, 18(1), 41–58 <https://www.nature.com/articles/nrd.2018.168>.
4. Ray, S. K., Pawlikowski, K., & Sirisena, H. (2008, October). A fast MAC-layer handover for an IEEE 802.16 e-based WMAN. In *International Conference on Access Networks* (pp. 102-117). Berlin, Heidelberg: Springer Berlin Heidelberg.
5. Mak, K. K., & Pichika, M. R. (2019). Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*, 24(3), 773–780 <https://doi.org/10.1016/j.drudis.2018.11.014>.
6. Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., ... & Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The innovation*, 2(4) [Volume 2, Issue 4](https://doi.org/10.1016/j.inno.2021.07.002)100179 November 28, 2021.
7. Samaras, V., Daskapan, S., Ahmad, R., & Ray, S. K. (2014, November). An enterprise security architecture for accessing SaaS cloud services with BYOD. In *2014 Australasian Telecommunication Networks and Applications Conference (ATNAC)* (pp. 129-134). IEEE.
8. Solanki, P., Baldaniya, D., Jogani, D., Chaudhary, B., Shah, M., & Kshirsagar, A. (2022). Artificial intelligence: New age of transformation in petroleum upstream. *Petroleum Research*, 7(1), 106-114. <https://doi.org/10.1016/j.ptlrs.2021.07.002>
9. Pandey, M., Fernandez, M., Gentile, F., Isayev, O., Tropsha, A., Stern, A. C., & Cherkasov, A. (2022). The transformational role of GPU computing and deep learning in drug discovery. *Nature Machine Intelligence*, 4(3), 211-221. <https://doi.org/10.1038/s42256-022-00463-x>
10. Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589 <https://www.nature.com/articles/s41586-021-03819-2>.
11. Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29.
12. Dey, K., Ray, S., Bhattacharyya, P. K., Gangopadhyay, A., Bhasin, K. K., & Verma, R. D. (1985). Salicyladehyde 4-methoxybenzoylhydrazone and diacetylbis (4-methoxybenzoylhydrazone) as ligands for tin, lead and zirconium. *J. Indian Chem. Soc. (India)*, 62(11).
13. Mohs, R. C., & Greig, N. H. (2017). Drug discovery and development: Role of basic biological research. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3(4), 651-657. <https://doi.org/10.1016/j.trci.2017.10.005>
14. De Moor, G., Sundgren, M., Kalra, D., Schmidt, A., Dugas, M., Claerhout, B., ... & Coorevits, P. (2015). Using electronic health records for clinical research: the case of the EHR4CR project. *Journal of biomedical informatics*, 53, 162-173.
15. Anwesa Chaudhuri, A. C., & Sanjib Ray, S. R. (2015). Antiproliferative activity of phytochemicals present in aerial parts aqueous extract of *Ampelocissus latifolia* (Roxb.) Planch. on apical meristem cells.
16. Ray, S. K., Sirisena, H., & Deka, D. (2013, October). LTE-Advanced handover: An orientation matching-based fast and reliable approach. In *38th annual IEEE conference on local computer networks* (pp. 280-283). IEEE.

17. Wilson, S., & Filipp, F. V. (2018). A network of epigenomic and transcriptional cooperation encompassing an epigenomic master regulator in cancer. *Npj Systems Biology and Applications*, 4(1), 24.
18. Muzafar, S., & Jhanjhi, N. Z. (2020). Success stories of ICT implementation in Saudi Arabia. In *Employing Recent Technologies for Improved Digital Governance* (pp. 151-163). IGI Global Scientific Publishing.
19. Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, 69, S36-S40.
20. Curioni-Fontecedro, A. (2017). A new era of oncology through artificial intelligence. *ESMO open*, 2(2).
21. Jabeen, T., Jabeen, I., Ashraf, H., Ullah, A., Jhanjhi, N. Z., Ghoniem, R. M., & Ray, S. K. (2023). Smart wireless sensor technology for healthcare monitoring system using cognitive radio networks. *Sensors*, 23(13), 6104.
22. Ivanisevic, T., & Sewduth, R. N. (2023). Multi-omics integration for the design of novel therapies and the identification of novel biomarkers. *Proteomes*, 11(1), 3.
23. Pun, F. W., Ozerov, I. V., & Zhavoronkov, A. (2023). AI-powered therapeutic target discovery. *Trends in Pharmacological Sciences*, 44(6), 409-423.
24. Rasooly, D., Peloso, G. M., Pereira, A. C., et al. (2023). Genome-wide association analysis and Mendelian randomization proteomics identify drug targets for heart failure
25. Zhang, X., Wu, F., Yang, N., Zhan, X., Liao, J., & Mai, S. (2022). In silico methods for identification of potential therapeutic targets.
26. Kaur, N., Verma, S., Jhanjhi, N. Z., Singh, S., Ghoniem, R. M., & Ray, S. K. (2023). Enhanced QoS-aware routing protocol for delay sensitive data in Wireless Body Area Networks. *IEEE Access*, 11, 106000-106012.
27. Zhao, J. H., Stacey, D., Eriksson, N., et al. (2023). Genetics of circulating inflammatory proteins identifies drivers of immune-mediated disease risk and therapeutic targets.
28. Zou, M., Zhou, H., Gu, L., Zhang, J., & Fang, L. (2024). Therapeutic target identification and drug discovery driven by chemical proteomics. *Biology*
29. Molajafari, F., Li, T., Abbaschaleshtori, M., et al. (2024). Computational screening for prediction of co-crystals: Method comparison and experimental validation. *Royal Society of Chemistry*.
30. Carvalho, V., Gonçalves, I. M., Rodrigues, N., et al. (2024). Numerical evaluation and experimental validation of fluid flow behavior within an organ-on-a-chip model. *Elsevier*.
31. Shang, L., Wang, Y., Li, J., et al. (2023). Mechanism of Sijunzi Decoction in the treatment of colorectal cancer based on network pharmacology and experimental validation. *Elsevier*.
32. Vasconcelos, D., Chaves, B., Albuquerque, A., & Andrade, L. (2022). Development of new potential inhibitors of  $\beta 1$  integrins through in silico methods—screening and computational validation. *Life*.
33. Muzammal, S. M., Murugesan, R. K., Jhanjhi, N. Z., & Jung, L. T. (2020, October). SMTrust: Proposing trust-based secure routing protocol for RPL attacks for IoT applications. In *2020 International Conference on Computational Intelligence (ICCI)* (pp. 305-310). *IEEE*.
34. Ayon, S. I., Islam, M. M., & Hossain, M. R. (2022). Coronary artery heart disease prediction: A comparative study of computational intelligence techniques. *Taylor & Francis*.
35. Crucitti, D., Pérez Míguez, C., Díaz Arias, J. Á., et al. (2024). De novo drug design through artificial intelligence: An introduction. *Frontiers in Drug Discovery*.
36. Azeem, M., Ullah, A., Ashraf, H., Jhanjhi, N. Z., Humayun, M., Aljahdali, S., & Tabbakh, T. A. (2021). Fog-oriented secure and lightweight data aggregation in iomt. *IEEE Access*, 9, 111072-111082.
37. Khater, T., Alkhatib, S. A., AlShehhi, A., & Pitsalidis, C. (2025). Generative artificial intelligence based models optimization towards molecule design enhancement. *Springer*.
38. Xie, S., Zhu, H., & Huang, N. (2025). AI-Designed molecules in drug discovery: Structural novelty evaluation and implications. *Journal of Chemical Information and Modeling*
39. Hunter, F. M. I., Ioannidis, H., Bento, A. P., et al. (2025). Drug and clinical candidate drug data in ChEMBL. *Journal of Medicinal Chemistry*.
40. Shah, I. A., Jhanjhi, N. Z., Amsaad, F., & Razaque, A. (2022). The role of cutting-edge technologies in Industry 4.0. In *Cyber Security Applications for Industry 4.0* (pp. 97-109). *Chapman and Hall/CRC*.
41. Kim, S., & Bolton, E. E. (2024). PubChem: A large-scale public chemical database for drug discovery. *Wiley*.
42. Lee, S., Abdullah, A., & Jhanjhi, N. Z. (2020). A review on honeypot-based botnet detection models for smart factory. *International Journal of Advanced Computer Science and Applications*, 11(6).
43. Lyubishkin, N. R., Kardash, O. V., & Klenina, O. V. (2022). Virtual databases for drug discovery.

44. de Azevedo, D. Q., Castilho, R. O., & Gómez-García, A. (2025). *Molecular databases in computer-aided drug design*. Springer.
45. Brohi, S. N., Jhanjhi, N. Z., Brohi, N. N., & Brohi, M. N. (2023). Key applications of state-of-the-art technologies to mitigate and eliminate COVID-19. *Authorea Preprints*.
46. Blanco-Gonzalez, A., Cabezon, A., Seco-Gonzalez, A., et al. (2023). The role of AI in drug discovery: Challenges, opportunities, and strategies. *Pharmaceuticals*.
47. Pun, F. W., Ozerov, I. V., & Zhavoronkov, A. (2023). AI-powered therapeutic target discovery. *Trends in Pharmacological Sciences*.
48. You, Y., Lai, X., Pan, Y., Zheng, H., Vera, J., & Liu, S. (2022). Artificial intelligence in cancer target identification and drug discovery. *Signal Transduction and Targeted Therapy*
49. Zhang, X., et al. (2022). In silico methods for identification of potential therapeutic targets.
50. Schneider, P., Walters, W. P., Plowright, A. T., et al. (2020). Rethinking drug design in the artificial intelligence era. *Nature Reviews Drug Discovery*.
51. Vamathevan, J., et al. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*.
52. Zhavoronkov, A., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*
53. Mak, K. K., & Pichika, M. R. (2019). Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*.
54. Topol, E. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*.
55. Pushpakom, S., et al. (2019). Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery*.
56. Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A., Židek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., . . . Velankar, S. (2024). AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, 52(D1), D368-D375. <https://doi.org/10.1093/nar/gkad1011>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.