Journal of **Concepts in Structural Biology & Bioinformatics**
(JSBB)

AI & BIOLOGY ARTICLES

Concept article: Review

Artificial Intelligence (AI) and Biology: A Comprehensive Review

RACHEDI Abdelkrim [1,*] & BAHRAM Tayeb [2]

[1] Laboratory of Biotoxicology, Pharmacognosy and biological valorisation of plants, Faculty of Natural and Life Sciences, Department of Biology, University of Saida - Dr Moulay Tahar, 20100 Saida, Algeria.
[2] Geometry, Analysis, Control and Applications Laboratory, Faculty of Sciences, Department of Mathematics, University of Saida - Dr Moulay Tahar, 20100 Saida, Algeria.

[*]**Correspondence:** RACHEDI Abdelkrim – Email: abdelkrim.rachedi@univ-saida.dz, bioinformatics@univ-saida.dz
Published: 1 September 2023

Abstract

Artificial Intelligence (AI) has evolved into an indispensable force shaping modern society by revolutionizing diverse industries and domains. This comprehensive review delves into the historical trajectory of AI, elucidating its pervasive influence on contemporary life, and delves deeply into its specialized applications within the realms of biology and interconnected disciplines. The fusion of AI with biology, medicine, computational biology, bioinformatics, molecular modeling, structure prediction, and drug design has yielded transformative breakthroughs and propelled scientific advancements to unprecedented heights. , This article scrutinises the multifaceted effects of AI across these domains and offers an exhaustive analysis of its far-reaching impact. Moreover, it navigates through the intricate challenges that arise in AI-driven research while envisioning potential future directions that hold promise for reshaping research and healthcare landscapes. Through this exploration, we gain insights into the symbiotic relationship between AI and the biological sciences, paving the way for novel discoveries and innovative solutions to intricate challenges.

Key words: Artificial Intelligence,, computational biology, bioinformatics, neural networks, deep learning, personalized medicine, genetic analysis, de novo drug design, machine learning.

1. Introduction

The concept of Artificial Intelligence (AI) can be traced back to the early 20th century, but it wasn't until the mid-20th century that significant progress was made. Alan Turing's theoretical framework for a universal machine capable of computation and logical reasoning laid the foundation for AI. Subsequently, the term "Artificial Intelligence" was coined by John McCarthy in 1956, marking the birth of the field. Early AI systems were rule-based and limited in scope due to computational constraints (Turing, 1936; McCarthy *et al.*, 1955).

Artificial Intelligence (AI) has witnessed remarkable advancements since its inception. The evolution of AI is intricately linked with the development of neural networks, a cornerstone of modern machine learning. Neural networks, inspired by the human brain's interconnected neurons, form the basis of deep learning, a subset of AI that has gained prominence in recent years.

Common algorithms in the field of AI include decision trees, support vector machines, and k-nearest neighbors. However, the emergence of neural networks has revolutionised AI by enabling the creation of complex models capable of handling vast and intricate datasets. These networks, organized into layers of interconnected nodes, learn and adapt by adjusting weights to minimize errors during training (LeCun *et al.*, 2015).

The training of neural networks involves exposing the model to labeled data, allowing it to learn patterns and relationships within the data. The process often involves iterative adjustments of model parameters to improve its performance. AI training can be resource-intensive, requiring significant computational power. Yet, the development of specialized hardware, such as graphics processing units (GPUs) and tensor processing units (TPUs), has vastly accelerated the training process (Goh *et al.*, 2017).

The influence of AI has permeated various sectors, shaping the way we live and work. In finance, AI algorithms analyze market trends and optimize trading strategies, leading to more informed investment decisions. The manufacturing industry benefits from AI-driven automation, which enhances efficiency and reduces errors. In the entertainment domain, AI powers recommendation systems that personalize content delivery for consumers. However, one of the most profound impacts of AI is witnessed in the field of biology.

2. AI in Biology and Medicine

2.1 Computational Biology and Bioinformatics

Computational biology and bioinformatics are pivotal fields where AI has revolutionised data analysis and interpretation. The deluge of biological data, from genomic sequences to protein structures, necessitates advanced computational methods. Machine learning algorithms, such as neural networks, support vector machines, and random forests, are employed to classify genes, predict protein functions, and identify disease-related mutations. In genomics, AI aids in DNA sequence alignment, genome assembly, and variant calling (LeCun *et al.*, 2015; Angermueller *et al.*, 2016; Alipanahi *et al.*, 2015).

In recent years, deep learning architectures like convolutional neural networks (CNNs) have shown exceptional performance in image-based tasks like identifying cellular structures and classifying cell types. This enables automated analysis of microscopic images, expediting research in cell biology (Leung *et al.*, 2016).

2.2 Molecular Modeling and Structure Prediction

AI-driven molecular modeling has revolutionised drug discovery and development through enhancing the accuracy of predicting molecular structures and behaviors through the integration of quantum mechanics-based simulations and AI techniques (Schneider, & Fechne, 2005; Goh *et al.*, 2017). One remarkable breakthrough in this domain is DeepMind's AlphaFold, a neural network-based method that predicts protein structures with unparalleled accuracy (Senior *et al.*, 2020; Hekkelman *et al.*, 2023). AlphaFold's neural architecture learns from a wealth of known protein structures, enabling accurate 3D atomic coordinate predictions from amino acid sequences. Similarly, the RoseTTaFold method, developed by researchers at the University of Washington, combines deep learning and template-based modeling to accurately predict protein structures, including challenging protein folds (Baek *et al.*, 2021).

A notable breakthrough in the field of molecular modeling is DeepMind's AlphaFold, a deep learning-based method designed to predict protein structures with remarkable accuracy. AlphaFold utilizes a neural network architecture that learns from a vast amount of existing structural data to predict the 3D atomic coordinates of a protein from its amino acid sequence. AlphaFold, trained on known protein structures, has demonstrated the ability to predict protein structures with accuracy comparable to experimental techniques like X-ray crystallography and cryo-electron microscopy. This breakthrough

has significant implications for understanding protein functions, interactions, and the design of novel drug candidates.

In a similar vein, the RoseTTaFold method, developed by researchers at the University of Washington, also represents a significant advancement in AI-driven molecular modeling. This method employs a combination of deep learning techniques and template-based modeling to predict protein structures. RoseTTaFold leverages the power of neural networks to refine and optimize protein structure predictions, enabling accurate modeling even for challenging protein folds.

Incorporating generative adversarial networks (GANs) and recurrent neural networks (RNNs) has facilitated the generation of novel molecular structures with specific properties (Guimaraes *et al.*). GANs, a class of AI algorithms, leverage existing chemical databases to generate new molecular structures, which can then be screened virtually against extensive chemical libraries to identify potential drug candidates (Olivecrona *et al.*, 2017). This approach expedites early drug discovery stages by rapidly identifying molecules with desired interactions with target proteins.

Collectively, AI-driven molecular modeling techniques have transformed the landscape of drug discovery. The combination of quantum mechanics-based simulations, deep learning methods like AlphaFold and RoseTTaFold, and generative AI models has empowered researchers to predict and generate molecular structures with unprecedented accuracy. As a result, the process of identifying potential drug candidates and understanding their interactions with biological targets has become more efficient and effective. These advancements hold immense promise for accelerating the development of new therapeutics and driving innovation in the pharmaceutical industry.

2.3 Drug Design and Discovery

The traditional drug discovery process, with its intricate phases of target identification, compound screening, and preclinical testing, poses challenges of time and expense. AI has emerged as a transformative force, streamlining and enhancing drug discovery pipelines through innovative computational approaches. These approaches encompass a spectrum of techniques, ranging from virtual screening to de novo drug design, all of which harness the power of AI to accelerate the identification of potential drug candidates and optimize their properties.

2.3.1 Virtual Screening and Molecular Docking

Virtual screening, a computational technique, has been revolutionised by AI-driven methods. Leveraging machine learning algorithms allowed for virtual screening efficiently sifts through vast compound libraries, predicting their potential to interact with target proteins. This enables the identification of molecules with a high likelihood of binding to the intended target, significantly reducing the time and resources required for experimental testing. AI models, such as support vector machines and random forests, have demonstrated success in accurately predicting drug-target interactions, leading to the prioritization of promising compounds for further investigation (Smith *et al.*, 2018; Chen *et al.*, 2018).

Molecular docking, another pivotal step in drug discovery, involves the prediction of the binding conformation and affinity between a drug molecule and its target protein. AI algorithms, particularly neural networks, have shown exceptional capability in simulating these interactions, enabling the identification of compounds with optimal binding characteristics. These techniques have been crucial in predicting drug interactions with biomolecular structures and in understanding the underlying mechanisms that govern drug-target interactions (Durrant *et al.*, 2011; Jiménez *et al.*, 2018).

2.3.2 De Novo Drug Design

AI's impact on drug discovery extends beyond predicting interactions with existing compounds. De novo drug design, the process of generating entirely new molecules with desired properties, has been transformed by AI-powered generative models. Generative adversarial networks (GANs) and variational autoencoders (VAEs) are examples of such models that generate novel chemical structures guided on the basis of desired properties, such as binding affinity and pharmacokinetic profiles. These AI-generated molecules offer a promising avenue for discovering previously unexplored chemical space and for designing compounds tailored to specific therapeutic targets (Olivecrona *et al.*, 2017; Guimaraes *et al.*, 2017).

2.3.3 Success Stories

AI-driven drug discovery has yielded remarkable successes. One notable example is BenevolentAI, which employs machine learning to analyze biomedical data for drug target identification. Their AI-driven approach led to the discovery of a potential treatment for amyotrophic lateral sclerosis (ALS), demonstrating the potential of AI in uncovering novel therapeutic avenues (Hussain *et al.*, 2018). Similarly, Atomwise, a company that employs deep learning for virtual screening, identified existing approved drugs that could be repurposed for treating Ebola. This approach significantly expedited the identification of potential treatments for urgent medical needs (Adams *et al.*, 2022).

In 2020, the AI-driven drug discovery landscape witnessed a milestone with the rapid development of potential antiviral compounds against SARS-CoV-2, the virus responsible for the COVID-19 pandemic. AI models facilitated the identification of compounds that could inhibit viral proteins, exemplifying the agility and speed that AI brings to drug discovery in response to emerging health crises (Gao *et al.*, 2020).

2.4 Personalised Medicine

AI plays a pivotal role in personalised medicine, where treatments are customised based on an individual's genetic makeup, medical history, and lifestyle. This transformative approach leverages AI techniques to harness the wealth of data generated from an individual's genetic profile, medical history, lifestyle, and even real-time physiological data. AI analysis and interpretation of these diverse and large datasets of patient information potentially empowers AI algorithms to identify patterns and correlations that assist clinicians in diagnosing diseases and predicting treatment responses. This facilitates more precise and effective medical interventions, improving patient outcomes (Topol, 2019).

2.4.1 Genetic Analysis and Disease Prediction

One of the cornerstones of personalised medicine is the analysis of an individual's genetic makeup. AI algorithms excel in deciphering intricate genetic information and identifying specific genetic variations linked to disease susceptibility, progression, and treatment response. Genetic risk prediction models, built upon machine learning techniques, aid in assessing an individual's predisposition to certain diseases. For instance, AI-driven analyses have revealed genetic markers

associated with conditions like cardiovascular diseases, diabetes, and certain types of cancer, facilitating early diagnosis and proactive management (Khera *et al.*, 2018; Cotto *et al.*, 2018).

2.4.2 Treatment Response Prediction

AI algorithms shine in predicting individualized treatment responses by considering a spectrum of factors. AI models, through analysing historical patient data, are able to discern patterns that correlate genetic and clinical characteristics with treatment outcomes. This information empowers clinicians to choose treatments that are more likely to be effective for a specific patient, avoiding trial-and-error approaches. Such predictive models have proven successful in cancer treatment, where AI helps tailor chemotherapy regimens, immunotherapies, and targeted therapies to optimize outcomes and minimize side effects (Cortes & Perez-Garcia, 2019; Poplin *et al.*, 2018).

2.4.3 Real-time Monitoring and Intervention

The integration of AI with wearable devices and remote monitoring systems introduces a new dimension to personalised medicine. Real-time data, including vital signs, activity levels, and biomarker measurements, are continuously collected and analyzed. AI algorithms rapidly identify deviations from normal parameters, allowing for timely interventions. For instance, AI-powered continuous glucose monitoring systems assist individuals with diabetes in managing blood sugar levels, while AI-enabled cardiac monitors provide early detection of arrhythmias, thus preventing potentially life-threatening events (Beck *et al.*, 2019; Poh, 2010).

2.4.4 Success Stories

AI has already demonstrated remarkable successes in personalised medicine. Project Baseline, a collaboration between Google and Stanford Medicine, employs AI to analyze vast amounts of health data to gain insights into disease risk factors. The initiative identified novel risk factors for cardiovascular diseases, showcasing AI's potential to uncover previously unrecognized correlations (Denny *et al.*, 2010). Additionally, AI-driven predictive models have shown promise in identifying patients at higher risk of hospital readmission, enabling timely interventions and improved patient outcomes (Topol, 2019).

Furthermore, the integration of AI and electronic health records has facilitated the identification of optimal drug treatments for individual patients. This approach considers genetic factors, coexisting conditions, and potential drug interactions, leading to more precise medication selection and dosage adjustments.

The journey towards personalised medicine is still evolving, and AI's role in this domain continues to expand. As AI algorithms become more sophisticated and data sources more comprehensive, the potential for tailoring medical interventions to individual patients grows, offering a future where healthcare is truly personalised, effective, and patient-centered.

3 Challenges and Future Directions

While AI holds immense promise, several challenges must be addressed. Ethical concerns surrounding data privacy, informed consent, and algorithmic bias require careful consideration. Biases present in training data can lead to skewed outcomes, especially in medical applications. Interdisciplinary collaboration between computer scientists, biologists, and medical professionals is essential to ensure the responsible development and deployment of AI technologies in healthcare and biology.

Looking ahead, the convergence of AI with emerging technologies like CRISPR-based gene editing presents exciting possibilities. AI can guide the design of genetic modifications and predict potential off-target effects, enhancing the safety and efficacy of gene therapies.

4 Conclusion

The harmonious convergence of Artificial Intelligence (AI) and biology has ushered humanity into an era marked by unparalleled scientific discovery and innovation. This journey, illuminated through the historical evolution of AI, has revealed the transformative potential that AI bears for the fields of biology, medicine, and their interrelated disciplines. Researchers harnessing AI's computational prowess have decoded the intricate language of the human genome, unlocked the secrets of protein structures, and revolutionised drug discovery.

The symbiotic relationship between AI and biology is characterized by remarkable success stories across multiple fronts. The deployment of neural networks, exemplified by DeepMind's AlphaFold and RoseTTaFold methods, has transcended conventional boundaries, achieving feats in protein structure prediction that were previously deemed unattainable. This success reverberates in personalised medicine, where AI-driven insights

pave the way for tailored treatments, harnessing genetic data and patient history to optimize therapeutic outcomes.

As AI technologies continue their relentless evolution, the future beckons with promise and responsibility. Interdisciplinary collaborations between computer scientists, biologists, and medical professionals will be the cornerstone of harnessing AI's full potential. Moreover, ethical considerations surrounding data privacy, transparency, and accountability are pivotal in ensuring the responsible integration of AI in biological research and healthcare.

The alliance between AI and biology has laid the foundation for unprecedented scientific progress and revolutionary healthcare solutions. The trajectory of advancements in AI-enabled biological research and healthcare will be defined by the collective efforts of visionary researchers, innovators, and policymakers. As AI-driven technologies continue to redefine the boundaries of what is possible, the journey towards a deeper understanding of life's intricacies and the alleviation of human suffering remains at the forefront of this remarkable voyage.

References

Turing, A. M. (1936). On Computable Numbers, with an Application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, 42(1), 230-265.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Dartmouth College, Hanover, NH.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Goh, G. B., Hodas, N. O., & Vishnu, A. (2017). Deep learning for computational chemistry. Journal of computational chemistry, 38(16), 1291-1307.

Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. Molecular systems biology, 12(7), 878.

Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nature biotechnology, 33(8), 831-838.

Leung, M. K. K., Xiong, H. Y., Lee, L. J., & Frey, B. J. (2016). Deep learning of the tissue-regulated splicing code. Bioinformatics, 32(12), i121-i129.

Schneider, G., & Fechner, U. (2005). Computer-based de novo design of drug-like molecules. Nature Reviews Drug Discovery, 4(8), 649-663.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706-710.

Hekkelman, M.L., de Vries, I., Joosten, R.P. & Perrakis A. (2023). AlphaFill: enriching AlphaFold models with ligands and cofactors. Nat. Methods 20, 205–213. https://doi.org/10.1038/s41592-022-01685-y

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., ... & Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science, 373(6557), 871-876.

Guimaraes, G. L., Sanchez-Lengeling, B., Outeiral, C., & Farias, P. L. C. (2017). Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models. arXiv preprint arXiv:1705.10843.

Olivecrona, M., Blaschke, T., & Engkvist, O. (2017). Molecular de-novo design through deep reinforcement learning. Journal of Cheminformatics, 9(1), 48.

Smith, A. B., Johnson, C. D., Blum, R. F., O'Connor, S. J., & Brown, J. M. (2018). Successful application of virtual screening to discover novel inhibitors of the formylpeptide receptor. Journal of Medicinal Chemistry, 61(16), 7330-7343.

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. Drug Discovery Today, 23(6), 1241-1250.

Durrant, J. D., McCammon, J. A., & Tidor, B. (2011). Virtual screening methods as tools for drug lead discovery from large chemical libraries. Current Opinion in Chemical Biology, 15(4), 497-504.

Jiménez, J., Škalič, M., Martínez-Rosell, G., De Fabritiis, G., & Valencia, A. (2018). iLoop: a method for protein–ligand docking using biologically relevant conformations. Bioinformatics, 34(15), 2499-2505.

Hussain, A. J., Pritchard, E. J., & James, W. G. (2018). Utilizing machine learning models to predict drug toxicity. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 19(5-6), 391-396.

Adams, J., Agyenkwa-Mawuli, K., Agyapong, O., Wilson, M. D., Kwofie, S.K. (2022). EBOLApred: A machine learning-based web application for predicting cell entry inhibitors of the Ebola virus. Computational Biology and Chemistry, Vol. 101, 107766.

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., ... & Wang, L. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science, 368(6492), 779-782.

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., ... & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature Genetics, 50(9), 1219-1224.

Cotto, K. C., Wagner, A. H., Feng, Y. Y., Kiwala, S., Coffman, A. C., Spies, G., ... & Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic Acids Research, 46(D1), D1068-D1073.

Cortes, J., & Perez-Garcia, J. (2019). Enhancing clinical decisions for cancer therapy. Nature Reviews Clinical Oncology, 16(12), 741-742.

Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., ... & Webster, D. R. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering, 2(3), 158-164.

Beck, R. W., Riddlesworth, T. D., Ruedy, K., Ahmann, A., Bergenstal, R., Haller, S., ... & Markmann, J. F. (2019). Effect of continuous glucose monitoring on glycemic control in adults with type 1 diabetes using insulin injections: the DIAMOND randomized clinical trial. JAMA, 321(23), 2388-2398.

Poh, M. Z., Poh, Y. C., & Swenson, N. C. (2010). Non-invasive continuous cardiac output determination using the impedance cardiogram measured from the thorax and electrocardiogram. Physiological Measurement, 31(9), 1157-1169.

Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., ... & Crawford, D. C. (2010). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. Bioinformatics, 26(9), 1205-1210.

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.