ISSN 2744-1636
Review Article

Bioengineering Studies
Vol. 5, No. 1, July 2024, pp.29-35
https://doi.org/10.37868/bes.v5i1.id317

Artificial intelligence in biology – a ashort review

Adna Hrapović [1*]
[1] Genetics Bioengineering, International University of Sarajevo, Sarajevo, Bosnia and Herzegovina

*Corresponding author: adna99@gmail.com

Received Jun. 30, 2024
Revised Jun. 15, 2024
Accepted Jun. 30, 2024

Abstract
Artificial intelligence is the development of computers and machines capable of mimicking human problem-solving and decision-making. Essentially, computer scientists wanted to create a system that acts like a human. Over time and with machine learning algorithms, AI has gained predictive power by using statistics and training data. It was able to "learn" and sort out data based on available research and calculate results. It started to grow and be frequently used by scientists in fields of genomics, proteomics, the pharmaceutical industry, and healthcare. AI has a bright future in applications of diagnosing cancers, offering an effective treatment, simulating metabolic processes, and predicting DNA and RNA binding sequences. Most importantly, the application of AI in the pharmaceutical industry can lead to lower costs and even eliminate clinical trials. Even though machine learning in biology is an extremely useful tool, there are some challenges, like insufficient financing and ethical concerns about data insufficiency and safety. Lastly, artificial intelligence should not replace human decision-making; it should make it more precise and easier. Artificial intelligence is still in need of development of law and ethical policies in order for it to collaborate with the healthcare and pharmaceutical industry.

Keywords: Artificial intelligence, genomics, AI in healthcare, proteomics, bioinformatics, drug design

1. Introduction

Artificial intelligence (AI) is a development of computers and machines that are capable of mimicking the problem-solving and decision-making capabilities of humans. The father of computer science, Alan Turing, defined artificial intelligence for the first time during his seminal work "Computing Machinery and Intelligence", published in 1950. His main question was "Can machines think?" From there, he proposed many experiments to see whether humans can distinguish computer work from human work. It is an important part of artificial intelligence history, since it started the ongoing concept of AI philosophy. For Alan Turing, artificial intelligence would be a "system that acts like a human", which means that ideally, machines would think and act rationally. Artificial intelligence is a field that combines computer science and datasets to enable problem-solving. It is connected with machine learning and deep learning, allowing the system to make predictions and classify input data [1] [2]. Nowadays, artificial intelligence programming brings cognitive skills such as learning, reasoning, self-correction, and creativity into focus. It has the power to change how we live and work since it is effectively used in business as an automation tool for humans, like customer service work, fraud detection, and quality control, and even completes a lot of competitive, detail-oriented tasks. It is used in a lot of successful companies, including Alphabet, Microsoft, Apple, and Google [3].

2. AI in biological applications

The idea of artificial intelligence itself dates back to ancient times, even 700 years B.C.in Greek mythology. Namely, giant Talus was a "creature" made of bronze created to protect Europa, the mother of King Minos in Crete. From then until now, artificial intelligence flourished with applications in academia and industry. From creating the first computer to complex rules in computer language that enable machines to do coordinated operations, such as playing the game of chess on their own. The chess-playing computer, Deep Blue, is a well-known example of IBM's computer from 1997. Deep Blue managed to beat Garry Kasparov, the world chess champion during that time, making this the first case of connecting AI with biology. Before this, computers were not capable of performing fast computations, let alone outpacing a well-trained human brain. Back then, the main difference between artificial intelligence and biological intelligence was that AI uses a prior established set of rules in its decision making, while biological intelligence is based on learning or acquired information [4]. The Deep Blue checkers program was seen as a big effort for evolutionary computing. Computing usually uses an automatic method of generating and evaluating success until a good solution evolves. This inspired another AI computer, the IBM 701 computer, which helped design a virtual rat's neural network that could be trained to find a way out of the maze. The creator of the IBM 701 computer was John Holland. After Holland graduated and earned his PhD degree, he joined the faculty of Michigan, where he automated evolutionary computing for four decades. He came up with the process called *genetic algorithms.* Genetic algorithms, among other important applications, essentially enabled the design of models of single-cell biological organisms. Another computer inspired by Deep Blue was the Logic Theorist. The Logic Theorist was made to prove theorems from "*Principia Mathematica"*, written in three volumes by Alfred North Whitehead and Bertrand Russell, British philosophers and mathematicians. The computer ended up writing a proof for the theorem in a more "elegant" way than the given proof from the books [5]. This inspired other projects, such as the GPS computer, which solved puzzles using a trial-and-error approach and an English-speaking computer that simulated a human therapist. This meant that the computer could have memorized intelligent conversations. Biology benefited from this project, especially in 1954 with the creation of artificial neural networks. Created by Farley and Wesley of MIT, the computer had more than 128 neurons running in a network. These neurons were able to recognize simple patterns. Their research showed that up to 10% of neurons can be randomly destroyed and that it would not affect a good, trained network. This sheds a glimpse of light on the brain's ability to tolerate damage that is inflicted by accidents, surgery, or even disease [6]. Two decades ago, biology was a descriptive science, explaining the systems and morphologies of beings. The discovery of DNA as a repository of inherited and coding information shed light on DNA recombinant technology, giving biology the power to not only explain systems, but to design and redesign them by developing genetic engineering. The main goal was to design systems and give them certain needed specifications, for example, to produce medical drugs or target drugs for the invasion of specific cancer types. The transition allowed industrialization of synthetic biology, which improved the health of humans and production of biofuels in order to combat climate change [7]. The main issue in biological sciences is the fact that scientists mostly cannot predict the behavior of a biological system, for example, how a genotype expresses into a phenotype or how a drug behaves on large-scale systems from small-scale experiments, etc. With machine learning AI, scientists are capable of providing the prediction without the details or mechanical understanding of the system. However, this process requires a lot of data, which can be produced through synthetic biology and automation, allowing AI to "act" as a biological system [8]. Predictive power of machine learning gives us detailed regularities of experimental data by linking statistics and training data, or in other words, it uses sets of input and output data of models that express almost any relationship without biased results. It is frequently used to predict metabolic pathway dynamics – diagnosing cancer, detecting tumors, predicting DNA and RNA binding sequences, etc. Generally, biologists can change DNA in model systems however they intend, but the cell behavior as a result is unpredictable [1] [9]. However, this type of biology is very "data hungry", meaning that it is difficult to produce both quality and quantity of recent data for machine learning systems without automation [8].

2.1. Machine learning techniques for data analysis and genetic algorithms

Machines learn using patterns, where they identify given data and correspond with a similar pattern. Machine learning can be categorized into three types: unsupervised learning, supervised learning, and reinforcement learning. Supervised learning is based on external assistance, like an expert who inputs information. Unsupervised learning requires no external assistance, and reinforcement learning, meaning that the decision is made based on a more positive outcome. All of these types are used in applications of machine learning in biology (Chakraborty et al., 2017; Uddin et al., 2020) [10][11]. Today, most machine learning processes require big data. Big data describes extremely big datasets and their combination with analytics, which enables gaining knowledge. The data usually increases because of the storage of data, like in pharmaceutical manufacturing. The data is managed in three steps: extraction of a big chunk of heterogeneous data, data configuration ensures formatting, and data analysis with different platforms for a good final output, which informs decisions about compounds or medicines that are about to be developed or to improve efficiency. Due to advancements and a large amount of research, meaningful insights were hard to generate, AI was enabled to actively adopt and tackle the data. It was designed to notice crucial issues in drug discovery and development while looking for the latest progress, technology, and novel studies by researchers in the field. Systematic search used recent literature with keywords and filtered scientific papers of this industry based on abstracts, relevant context, and methodology [12] [13].

2.2. AI in drug discovery

Another noticeable AI implementation is in drug manufacturing and drug discovery computation. Some of those technologies include metabolism, ADMET platform (tells information about the excretion of a drug from the system), *in silico* absorption, etc. Here, AI systems are designed to model pharmacokinetic and physicochemical points for drug manufacture, design interesting proteins, and recognize unknown protein models. Currently, this model is used by the Bayer company [14]. Even if the data comes out negative in research, AI has the ability to learn and improve the ongoing designs; however, the systems still need manual maintenance [15]. In artificial intelligence, there is a type of evolutionary computer algorithm that uses symbols such as genes or chromosomes. This is known as a "genetic algorithm". Genetic algorithm calculates a combination of symbols, analogous to the crossing over process of genetic recombination and mutation rates [16]. Since large amounts of data can be used, generations of algorithms are stored, providing processes like natural selection or the evolution of genetics. This field of research is called genetic programming. Genetic algorithms are also used in cellular automation research, artificial life, and neural networks. The system was initially written by John Holland and commercialized by the founder of Scientific Games Company, John Koza. Genetic programming was computationally hard, which unfortunately limited applications because of the lack of "supercomputers" that could process all data. However, personal computers became much stronger, which led to genetic programming success in quantum computing. An important part of the genetic algorithm is that it offers a stable and efficient path planning that can improve the accuracy and efficiency of samples [17]. The human genome project's completion signaled the start of a new era in biomedical research. The human genome has over 3 billion DNA base pairs, which helped determine the most precise sequences. Before the detection of the problem, the human genome must be precisely sequenced, and the analysis of such a massive amount of data is expensive. As a result, global genomics firms are creating methods of sequencing using AI. DNA variants and their analysis are important for genetic diagnosis, which AI can detect 80 times faster than a human. As a result, the clinicians get their results quickly and precisely, which helps them start clinical management accordingly [18]. Over the last ten years, machine learning has become increasingly popular. The scientific community had several papers published by 2014 which interpret genome data and machine learning. They recognized that understanding genes would help medicine around the world. Data patterns and a large amount of memory will allow problem-solving in an accurate, personalized, and faster way [19]. Ever since proteins were possible to obtain in purified form using Mass Spectroscopy and Blotting techniques, they came into the picture with AI analysis. High-throughput methods developed in protein-based studies are called proteomics. Machine learning expanded the applications in prediction, pattern recognition, feature selection, and automation. The majority of programs were using semi-supervised learning techniques [20] [21]. Protein structure prediction was essential for understanding biological processes and cell functioning. Protein structure and fold predictions had a big role in understanding the functions of proteins, and their data storage was important. Since the experimental process of protein structure prediction is very expensive and requires precise work, programmers used a Deep Learning AI in order to create computational techniques for folding and structure prediction of available protein sequences from public domain datasets. This developed a fast method and accurate automated tools that can be used for structure prediction. The three-dimensional shape of proteins and the attainment process of the shapes are very important to understand, because it essentially helps in fields like drug design, medicine, and the pharmaceutical sector [22][23]. Another AI application that is important for genome analysis is phylogeny. Over the past few years, bioinformatics has grown enormously. With simple techniques and a couple of algorithms, phylogeny expanded the data of evolution, organisms, genomic sequences, etc. Algorithms involved are sequence alignment, alignment check, distance computation, and validation. Artificial intelligence also helped evolve the next-generation sequencing by collecting, storing, and defining data, including published scientific experiments, experimental results, literature, and analysis from proteomics, genomics, metabolomics, microarray analysis, and the latest NGS data. Primary databases contain genetic sequences, and protein structure is stored in a secondary database. Some popular examples are GenBank, which stores genetic sequences, SwissProt, a protein sequence database, and PIR [24]. Finally, with AI, we are able to check expression rates of thousands of genes at the same time by using DNA Microarrays. This helped the expression profile of thousands of genes, helping the identification of diseases and developing their treatment. A microarray is an integral part of medical studies and molecular biology. The analysis of gene expression can help us find a cause of the disease by checking the disease-related genes in a patient [10] [25]. The drug discovery process is complex, requiring identification of active compounds, designing the drug candidates, and determining their safety through preclinical and clinical studies. The process took many years and billions of dollars, often resulting in large numbers of failures, even if the compound had promising results [26]. Recent improvements in genomics and disease diagnostics enabled the new and creative pharmaceutical products, based on approaches and analytic precision instruments and portions. Medicine is becoming more tailored to the patient’s needs, improving the efficiency of the drug industry [27]. Since the use of artificial intelligence is increasing, it is possible that it will change the way of clinical examination. Doctors can be a part of the development of the technology used technology and it will improve medical care. AI is used in the pharmaceutical industry in four main ways: prediction of the success of treatment, assistive technology during treatments, defining the severity of the disease, and it is able to determine the reason behind the use of medical instruments and to upgrade it in order to improve safety and efficacy. Machine learning and AI have the ability to lower costs or even eliminate clinical trials because of the offered simulations of the situation, providing researchers more information about different molecules [28][13].

AI has a bright future for applications and enormous potential for the pharmaceutical industry. Novel therapeutic target structure-based molecules can be discovered, and AI allows researchers to solve challenging issues like quantitative and predictive epidemiology, host-pathogen interaction, and precision-based medicine [29][30]. It is valuable to include differences in genetics, ecology, and lifestyle while diagnosing a patient. Due to the large advantages of AI, an era of personalized medicine is tailoring drugs based on the body’s needs and adaptability. Researchers can use DNA, RNA, and protein studies to visualize the effects in certain patient groups. Computer algorithms can detect metastatic breast cancer in lymph node biopsies from slide images, and the results are 91% accurate. Combining these results with a physician's diagnosis, the accuracy can reach as much as 99.5%. Another proven AI application is risk analysis for heart malfunctioning diagnosis that includes automated monitoring. [31] [32]. AI tools are used to detect the primary type of cancer from a liquid biopsy sample, which was very hard in the past. Genome fragments in the liquid biopsy sample, in combination with AI models, can provide a good approach for the screening, including good detection and monitoring of human cancer. It is significant for cancer treatment to identify the type of cancer and the time of progression. AI-based models were applied to datasets of genomics from breast, lung, colorectal, and many other cancer types, which helped predict the kind of cancer that can progress in a patient. AI-based models and genomics data can identify cancer subtypes, new marker discoveries, therapeutic targets, and overall better knowledge of cancer-related genes, which is crucial for personalized patient management. It is possible to identify subpopulations of genetic profiles that are likely to respond positively to a particular treatment and track how genetic makeup responds to treatment [33]. No field of science remains unaffected by artificial intelligence, not even psychiatry. Novel system diagnostics and therapeutics with AI are possible due to proteomics data and algorithms like "environtome" and "social proteome" that study external elements of the host [20].

2.3. Challenges

Even though machine learning in biology is an extremely useful tool, there are some challenges. For example, it is a challenge not to allow the AI computer to be an open-access one, since a lot of patients' medical information can be exposed, which is unethical. Secondly, occupational and skillset immobility in the pharmaceutical industry represents a problem because there are not enough staff members currently working in the pharmaceutical industry who are qualified to operate AI systems. It requires being a specialist in both data science and molecular chemistry or biology. The third problem is insufficient financing for the development of AI in the industry because of skepticism about the results of artificial intelligence. This, unfortunately, can lead to slower and less thorough research. These problems are in the way of the true development of artificial intelligence in the pharmaceutical industry [34] [28]. Another challenge is that healthcare data is sometimes incomplete or inaccurate, which leads to the production of errors. AI in healthcare raises some other ethical concerns because it is crucial to understand that AI should make human decision-making easier and more precise, instead of replacing it [35]. For solving the challenges, it is needed to invest time and resources into implementation processes and develop laws and policies that regulate the collaboration of AI across healthcare and industry [36].

3. Conclusion

The existence of artificial intelligence allows humans to perform tasks in a much more precise, efficient, and faster way. Artificial intelligence can "learn" to assist humans in many scientific fields, including biological, clinical, genomic, and drug research. It is a huge achievement with the capability to mimic human intelligence. AI holds a lot of potential to solve issues in medical research, agriculture, the pharmaceutical industry, etc. Despite challenges, including data errors and non-consistency, machine learning capabilities are able to learn how to overcome any problem. In order to avoid AI misuse, it is necessary to form laws and policies that regulate AI and healthcare industry collaboration. It is believed that artificial intelligence will be increasingly used in the future, providing tools for genetic research, epidemiology, and cancer research, as well as industrial and healthcare fields.

References

[1] Z. Costello and H. G. Martin, "A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data," *NPJ Syst. Biol. Appl.*, vol. 4, no. 1, 2018, doi: 10.1038/s41540-018-0054-3.

[2] M. V. Johnson, K. Garanger, J. O. Hardin, J. D. Berrigan, E. Feron, and S. R. Kalidindi, "A generalizable artificial intelligence tool for identification and correction of self-supporting structures in additive manufacturing processes," *Addit. Manuf.*, vol. 46, 2021, doi: 10.1016/j.addma.2021.102191.

[3] Y. Leviathan and Y. Matias, "Google AI Blog: Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone," 2018.

[4] M. Usak, "ARTIFICIAL INTELLIGENCE IN BIOLOGY EDUCATION," *Journal of Baltic Science Education*, vol. 23, no. 5, 2024, doi: 10.33225/jbse/24.23.806.

[5] J. E. Force, "The confidence of British philosophers. An essay in historical narrative," *Hist. Eur. Ideas*, vol. 1, no. 4, 1981, doi: 10.1016/0191-6599(81)90030-9.

[6] E. Atalay, O. Özalp, Ö. C. Devecioğlu, H. Erdoğan, T. İnce, and N. Yıldırım, "Investigation of the Role of Convolutional Neural Network Architectures in the Diagnosis of Glaucoma using Color Fundus Photography," *Turk. J. Ophthalmol.*, vol. 52, no. 3, 2022, doi: 10.4274/tjo.galenos.2021.29726.

[7] C. Sequeiros, C. Vázquez, J. R. Banga, and I. Otero-Muras, "Automated Design of Synthetic Gene Circuits in the Presence of Molecular Noise," *ACS Synth. Biol.*, vol. 12, no. 10, 2023, doi: 10.1021/acssynbio.3c00033.

[8] P. Carbonell, T. Radivojevic, and H. García Martín, "Opportunities at the Intersection of Synthetic Biology, Machine Learning, and Automation," 2019. doi: 10.1021/acssynbio.8b00540.

[9] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, 2017, doi: 10.1038/nature21056.

[10] I. Chakraborty and A. Choudhury, "Artificial Intelligence in Biological Data," *J. Inf. Technol. Softw. Eng.*, vol. 07, no. 04, 2017, doi: 10.4172/2165-7866.1000207.

[11] F. Uddin, C. M. Rudin, and T. Sen, "CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future," 2020. doi: 10.3389/fonc.2020.01387.

[12] A. T. Greenhill and B. R. Edmunds, "A primer of artificial intelligence in medicine," 2020. doi: 10.1016/j.tgie.2019.150642.

[13] P. Solanki, D. Baldaniya, D. Jogani, B. Chaudhary, M. Shah, and A. Kshirsagar, "Artificial intelligence: New age of transformation in petroleum upstream," 2022. doi: 10.1016/j.ptlrs.2021.07.002.

[14] I. Abu-elezz, A. Hassan, A. Nazeemudeen, M. Househ, and A. Abd-alrazaq, "The benefits and threats of blockchain technology in healthcare: A scoping review," 2020. doi: 10.1016/j.ijmedinf.2020.104246.

[15] A. H. Göller *et al.*, "Bayer's in silico ADMET platform: a journey of machine learning over the past two decades," 2020. doi: 10.1016/j.drudis.2020.07.001.

[16] K. de Jong, "Learning with Genetic Algorithms: An Overview," *Mach. Learn.*, vol. 3, no. 2, 1988, doi: 10.1023/A:1022606120092.

[17] Y. Ning, F. Zhang, B. Jin, and M. Wang, "Three-dimensional path planning for a novel sediment sampler in ocean environment based on an improved mutation operator genetic algorithm," *Ocean Engineering*, vol. 289, 2023, doi: 10.1016/j.oceaneng.2023.116142.

[18] F. Glover, "Future paths for integer programming and links to artificial intelligence," *Comput. Oper. Res.*, vol. 13, no. 5, 1986, doi: 10.1016/0305-0548(86)90048-1.

[19] V. C. Müller and N. Bostrom, "Future progress in artificial intelligence: a poll among experts," *AI Matters*, vol. 1, no. 1, 2014.

[20] V. Özdemir *et al.*, "Personalized medicine beyond genomics: alternative futures in big data—proteomics, environtome and the social proteome," 2017. doi: 10.1007/s00702-015-1489-y.

[21] N. Savage, "Proteomics: High-protein research," *Nature*, vol. 527, no. 7576, 2015, doi: 10.1038/527S6a.

[22] J. Schmidhuber, “Deep Learning in neural networks: An overview,” 2015. doi: 10.1016/j.neunet.2014.09.003.

[23] R. Schmitt, F. Dietrich, and K. Dröder, “Big Data Methods for Precision Assembly,” in *Procedia CIRP*, 2016. doi: 10.1016/j.procir.2016.02.141.

[24] Y. Park, D. Heider, and A. C. Hauschild, “Integrative analysis of next-generation sequencing for next-generation cancer research toward artificial intelligence,” 2021. doi: 10.3390/cancers13133148.

[25] R. Spang, “Diagnostic signatures from microarrays: A bioinformatics concept for personalized medicine,” 2003. doi: 10.1016/s1478-5382(03)02329-1.

[26] E. Naresh, B. P. Vijaya Kumar, Ayesha, and S. P. Shankar, “Impact of Machine Learning in Bioinformatics Research,” 2020. doi: 10.1007/978-981-15-2445-5_4.

[27] J. Rantanen and J. Khinast, “The Future of Pharmaceutical Manufacturing Sciences,” 2015. doi: 10.1002/jps.24594.

[28] V. Patel and M. Shah, “Artificial intelligence and machine learning in drug discovery and development,” 2022. doi: 10.1016/j.imed.2021.10.001.

[29] X. Liu *et al.*, “A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis,” *Lancet Digit. Health*, vol. 1, no. 6, 2019, doi: 10.1016/S2589-7500(19)30123-2.

[30] A. İ. Tekkeşin, “Artificial Intelligence in Healthcare: Past, Present and Future,” 2019. doi: 10.14744/AnatolJCardiol.2019.28661.

[31] A. Bhardwaj, S. Kishore, and D. K. Pandey, “Artificial Intelligence in Biological Sciences,” 2022. doi: 10.3390/life12091430.

[32] J. Zhang, R. Han, G. Shao, B. Lv, and K. Sun, “Artificial Intelligence in Cardiovascular Atherosclerosis Imaging,” 2022. doi: 10.3390/jpm12030420.

[33] H. Abdelhalim *et al.*, “Artificial Intelligence, Healthcare, Clinical Genomics, and Pharmacogenomics Approaches in Precision Medicine,” 2022. doi: 10.3389/fgene.2022.929736.

[34] P. V. Henstock, “Artificial Intelligence for Pharma: Time for Internal Investment,” 2019. doi: 10.1016/j.tips.2019.05.003.

[35] Z. Wen and H. Huang, “The potential for artificial intelligence in healthcare,” *J. Commer. Biotechnol.*, vol. 27, no. 4, 2022, doi: 10.5912/jcb1327.

[36] L. Petersson *et al.*, “Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden,” *BMC Health Serv. Res.*, vol. 22, no. 1, 2022, doi: 10.1186/s12913-022-08215-8.