Convergence of Artificial Intelligence and Biotechnology

[1]Radhika S Nair, [1]Priya Swaminathan*
[1]Department of Biotechnology, College of Engineering and Technology, SRMIST, Kattankulathur- 603203, Tamilnadu
Corresponding Author

"Anything humans can do, machines will do too"- was the expectation that caused the creation and evolution of Artificial Intelligence(AI) as we know it today. Artificial Intelligence or AI, is merely a hypernym for creating intelligent systems to mimic and mirror tasks that require human intelligence. The process involves utilising computational power and data-driven insights to determine outcomes. It can be visualised as a mathematical equation: $x + y = a$. The outcome "a" is wholly based on the values fed into the algorithm by the user, i.e., x & y.

AI utilises techniques to extrapolate that idea to utilize the computational power of multiple equations, analysing vast datasets, with the system developing the capacity to determine the most viable route to reach the required outcome over time. For example, Machine Learning takes this principle of finding the relationship between various factors to determine the outcome for previously unknown instances. This ability of pattern recognition enables the decision-making capacity of machines.

This capacity is "Artificial Intelligence", aided by two crucial factors: increasingly available data sets and more computation power. This capacity has applications in every aspect of human existence.[1]

This chapter will focus more on AI's impact on Biotechnology, familiarising the reader with the methods and terms surrounding Artificial Intelligence and subsequently relating to applications of AI in biotechnology. Considering the multi-faceted applications of biotechnology - cosmetic industry, healthcare, and food industry etc, the societal impact of the field is sublime. The process involves harnessing cellular and biological processes and modifying living systems (e.g. GMOs) to develop technologies and products.

Scaling up the research to meet the requirements of practical applications mandates precision and efficiency. While current methods utilize engineering, immense manpower and extensive resources, integration with AI allows biotechnological research to obtain accurate results with optimised methods in an expedited and sustainable manner.

Furthermore, the unprecedented nature of the progression of AI incites a massive amount of scrutiny. The need for ethical and responsible methods has always been highly regarded in Biotechnology, considering the field utilizes living organisms. Therefore, the social impact of integrating AI and biotechnology must be thoroughly analysed.

This chapter prompts deliberation on critical aspects of research, such as the limits of inquiry. As the boundaries of research expand, it is imperative now than ever, to resolve such questions.

1. Foundations of AI/ Taking Humans out of the loop

Since its inception, Artificial Intelligence has been interpreted through many disciplines. It is essential to note that AI is a concept that motivated the origination of machine learning and deep learning techniques to enable machines to imitate human cognitive functions such as learning and problem-solving.

The ultimate aim of machine learning is the artificial generation of knowledge from experience. Data is processed through models/algorithms to elicit predictions. Over the years, the technique has evolved to incorporate various theories and practices to provide optimum results.

1.1 Knowledge extraction

In the current scenario, data exists in many forms - structured, unstructured, semi-structured, etc. Before techniques under artificial intelligence are applied to analyse and interpret the data, the input data must be extracted and processed.

The correct data representations can ensure that the right quantity and quality data is extracted from the input data.

1.2 Machine Learning

Machine learning or ML uses the foundations of Statistical Learning Theory or SLT to study learning, inference, data extraction, prediction modelling and decision-making from given data. Consequently, SLT enables the construction of better learning algorithms.

Machine learning is primarily based on probabilistic reasoning. Machine learning utilizes input data to establish data from various factors and predict the outcomes with the highest probability. When this method is combined with "artificial generation of knowledge or knowledge from experience", machines gain the capacity to make independent yet accurate predictions.

The availability of "Big Data," gathered by satellites, telescopes, high throughput machines, sensor networks, smartphones, and other devices, helped the area advance significantly over the past two decades with applications successful in various domains, ranging from astronomy to zoology.

While statistics attempted to give a human the means to analyse data manually, the goal of machine learning (ML) was to replace humans and to automatically learn from data to make decisions similar to human cognitive behaviour as seen in **Figure 1.**

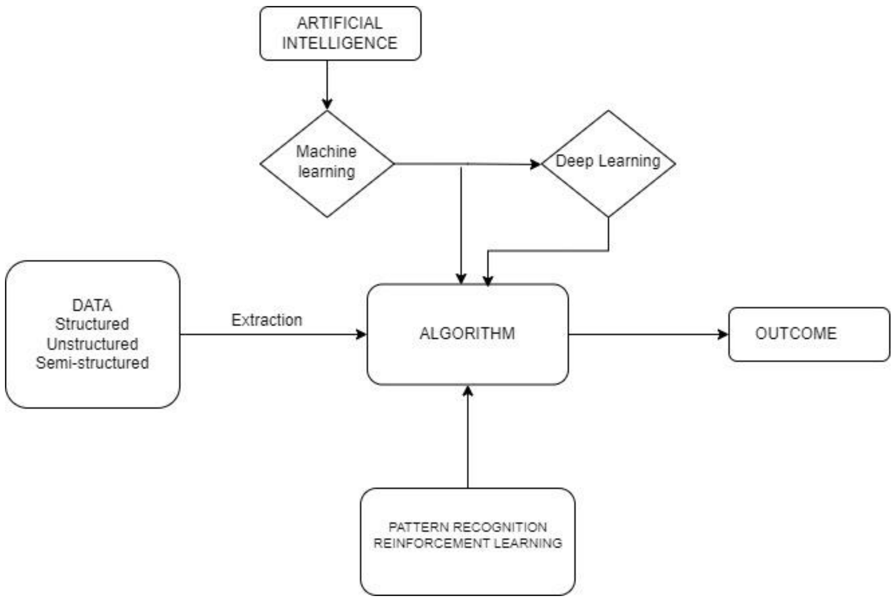

Figure 1: Artificial Intelligence uses machine learning and deep learning methods to extract, analyse and predict outcomes.

1.3 Deep Learning

As previously mentioned, Deep learning is a subset of machine learning. It utilises neural networks, a computational model inspired by interconnected neurons in the human brain. Artificial neurons or interconnected nodes mimic the information processing of the brain. During training, each node is associated with a weight that determines the node's impact on the output. The propagation of information utilises input, hidden and output layers. The information is propagated through the layers, where different computations occur at every node, resulting in the final output - predictions, classifications and many more. With training the network learns to minimize the difference between its predictions and actual results.

Unlike traditional neural networks, Deep learning utilizes stacked multiple hidden layers to extract more intricate features from the input data - hence "deep" learning.

Deep learning network architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are currently used in imaging and data analysis.

Recently, artificial neural networks were used to predict the microbial composition of Activated Sludge (AS) as part of the wastewater treatment method. As seen in **Figure 2,** the study utilised Artificial Neural Networks (ANNs) to model complex relationships between microbial communities and environmental variables. The method allowed us to account for non-linear associations between variables and relationships between predictors.

Predictive models were constructed for alpha diversity indices such as the Shamon-Weiner index, Ielou's evenness index, species richness, and Faith's phylogenetic diversity.

Garson's connection weight approach was used to assign a vital weight value to each environmental element throughout the model training process for forecasting the alpha-diversities of AS systems. When the model was used to forecast alpha diversities, the components with greater relevance weights were more informative. ANN helped observe that certain environmental factors had a more significant impact in predicting alpha-diversities of microbial communities.[2]

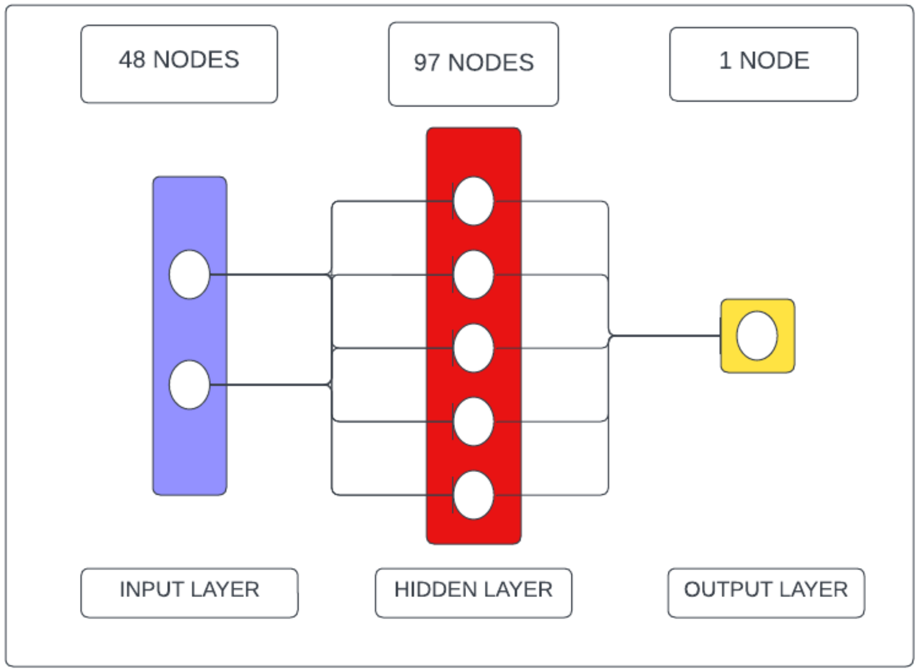

Figure 2: The Artificial Neural Network model framework was utilised to predict microbial diversities in activated sludge, with the input layer consisting of environmental factors and the output layer consisting of alpha-diversities functional groups.

1.4 Big Data

The biomedical industry and other fields produce vast amounts of data that, when harnessed, can optimise practice methods to a large extent, especially in the healthcare industry. [3]

Big Data is usually described using five features -

1. Volume: Vast amount of unaccounted and accounted, complex and heterogeneous data being produced at every instant that exceeds the capacity of conventional methods for storage and analysis.

2. Velocity: The data is being produced, exchanged and processed at an alarming rate.
3. Variety: As time progresses, data is generated in different forms- structured, semistructured, and unstructured.
4. Veracity: The data's precision, significance and prediction capacity are unknown and can only be identified through thorough processing.
5. Value: As data becomes publicly available, converting data for commercial purposes increasingly becomes relevant.

Big Data analytics begs the need for Artificial intelligence for its computational and reasoning capacity in data management. Due to this Big Data along with AI are ubiquitous across society. An excellent example is how AI-based tools are used by the government and healthcare policymakers to prevent and control outbreaks of illnesses- FINDER is a machine-learning model to diagnose foodborne illness diagnosis in real-time via anonymous and associated web search and location information.[4]

2. APPLICATIONS:

2.1 Imaging

AI has been used for various imaging applications such as radiologic applications in classification, object detection, image segmentation, image generation, and image transformation.

The convolutional layers of the CNN identify patterns, layers and edges of an image.

The CNN architecture contains kernels (small matrices) capable of vision tasks such as segmentation and classification. The kernel convolves or slides across the input image, detecting different image features from different regions, resulting in values in the output feature map. CNNs contain multiple kernels, producing feature maps that collectively capture different visual elements. After convolution, pooling reduces spatial dimensions and extracts precise information from each feature map. During training, CNNS undergo successive layers of convolution and pooling allowing the model to learn hierarchical representations and capture complex abstract features.

CNNs require large datasets during training to achieve maximum accuracy. Data augmentation resolves this issue by utilising image transformation methods such as rotation, translation, scaling etc. to generate images from limited or unbalanced datasets. The process aids in classification, object detection, image segmentation, image generation and image transformation, especially for biomedical applications.

2.1.1 Image classification:

Deep Convolutional Neural Networks (DCNNs) can accurately classify the presence or absence of diseases from given datasets in a study conducted to evaluate the efficacy of DCNNs in detecting tuberculosis from chest radiographs. The datasets were divided into three categories: training (68.0%), validation (17.1%), and test (14.9%). AlexNet and GoogLeNet, two separate DCNNs, were utilised to categorise the images as either having pulmonary TB symptoms or as healthy. It was observed with a high area under curve or AUC value of 0.99 that tuberculosis can be correctly classified using deep convolutional neural networks from chest radiography.[5]

2.1.2 Object Detection:

With convolutional neural networks, a two-stage object detection algorithm provides the highest accuracy of object detectors. Yet, single-stage object detection algorithms with regression-based models provide more rapid results. Retinanet recently developed has been proven to improve the accuracy rates of single-stage methods.

Convolutional neural networks developed for biomedical image segmentation are also highly utilised for landmark prediction. For example, U-Net, a CNN developed by the University of Freiburg is being utilised to detect landmarks in spine sagittal X-rays before spine surgeries. U-Net segments the landmarks in the images from the region of interest and further, corresponding coordinates (x and y) were evaluated.[6]

2.2 MOLECULAR PROFILING / METABOLOMICS

Metabolomics is the study of metabolites or molecules found in biological systems that can be used to comprehensively assess the physiological processes, diseases and system dynamics. This comprehensive assessment is critical in disease diagnosis and therapy. Metabolomics is often associated with phenotype, reflecting the data modulated and expressed by upstream genetic regulation. Consequently, metabolomics directly influences the areas of biomarker discovery.[7]

At the expression level, metabolic disorders are indicated by biochemical abnormalities, which can be detected by biomarker identification or screening methods. However, accurately associating a single biomarker with a disorder or disease requires extensive datasets. Different branches of omics cater discovery of biomarkers. Targeted analysis, metabolite profiling and metabolic fingerprinting are the significant approaches in metabolomic studies for analysis and profiling. The approaches generate an extensive amount of complex data. The processing and interpretation algorithms and high computational power should prioritise background noise elimination, peak identification, alignment and normalization. Integrating metabolomic studies with AI allows rapid and accurate data analysis, pattern identification and predictive modelling.

2.2.1 Case Study On Identification Of Severe Coronavirus Cases Based On Molecular Signatures Of Proteins And Metabolites:

The COVID-19 pandemic urged extensive research and investigation into disease heterogeneity and progression**. Figure 3** illustrates a particular study, "Proteomic and Metabolomic Characterization of COVID-19 Patient Sera" conducted at Westlake University attempted to combine proteomic and metabolic profiles of 31 COVID-19 patients to create a Machine Learning molecular classifier, which can screen potential blood biomarkers associated with severe COVID-19. Out of the patients, 13 individuals were severely affected by the disease.[8]

Figure 3(A) summarises the groups of COVID-19 patients considered. A Random Forest model was constructed, which successfully identified 29 variables of priority - 22 proteins and seven metabolites. In the training set, the model had an area under the curve of 0.957. The model was tested with an independent cohort of 10 patients, as shown in figure 3(B) and resulted in the accurate identification of 9 patients as illustrated by figure. The confounding factors such as the prolonged administration of non-traditional medicine, might have misidentified the one patient.

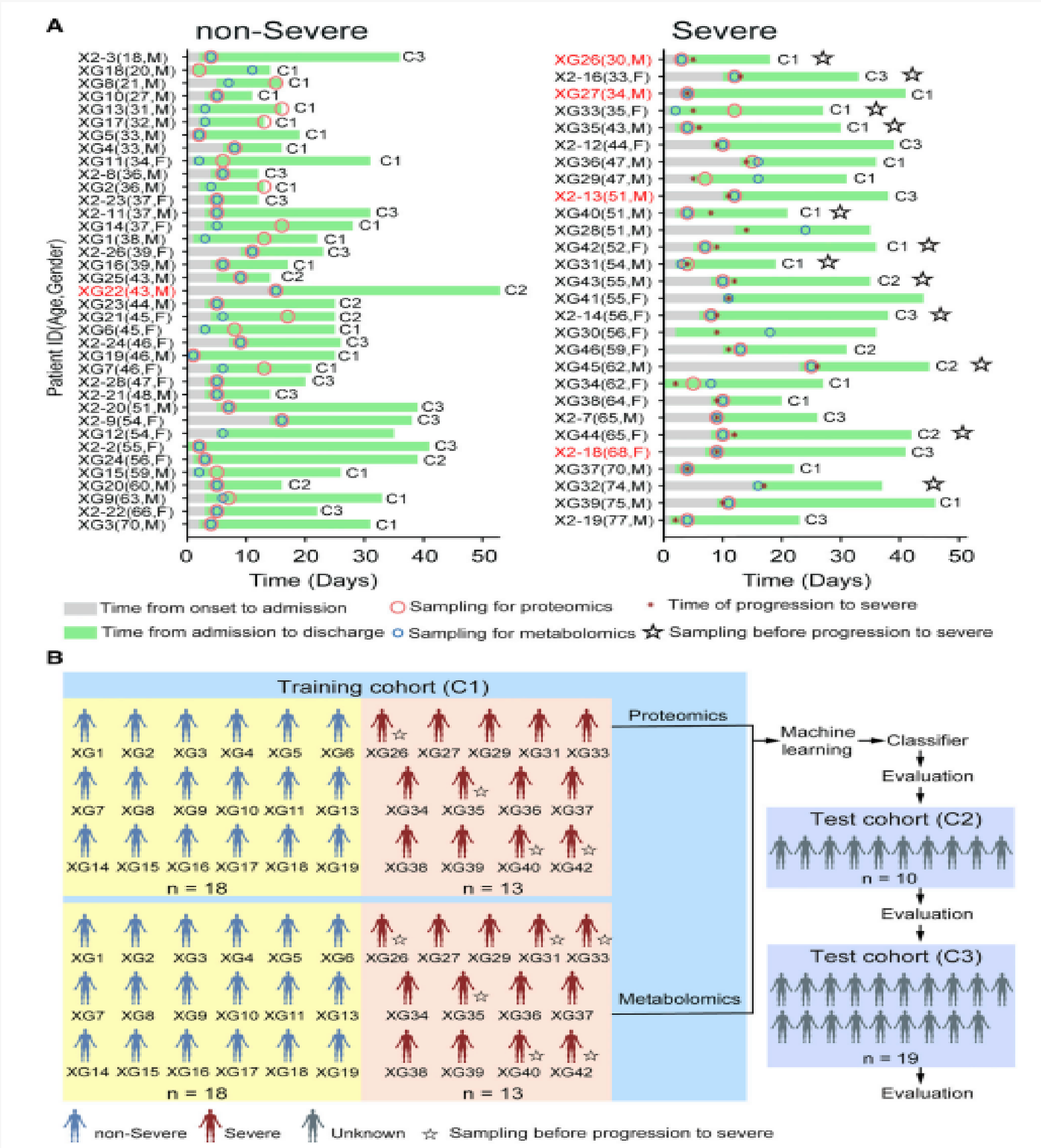

Figure 3:(A) Summary of COVID-19 patient groups, including non-severe (n = 37) and severe (n = 28) cases, with specific details provided in Table S1. Notably, patients with chronic hepatitis B virus infection are highlighted in red on the y-axis.(B) The outlined procedure for developing a machine-learning-based classifier to identify severe COVID-19 cases. This involved obtaining samples in a training cohort (C1) for proteomic and metabolomic analysis, followed by validation in an independent test cohort (C2) and a subsequent test cohort (C3).

2.3 DRUG DISCOVERY/ AI in the pharmaceutical industry:

The impact of AI in the pharmaceutical industry has been increasingly more relevant. AI-assisted methodologies are used in drug discovery, development and development as displayed in **Figure 4**. The integration of AI improves pharmacological designs in terms of efficiency and personalization while reducing human interference to a great extent.[9][10]

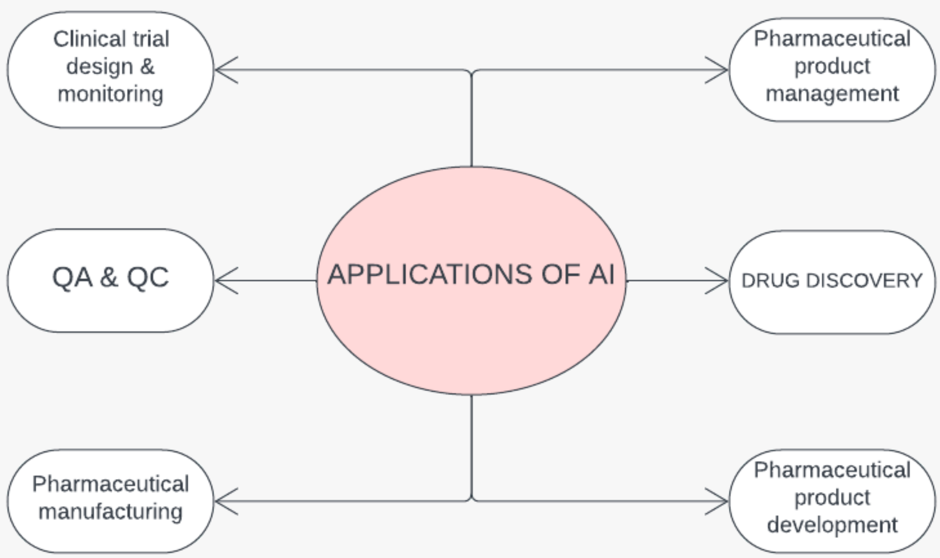

Figure 4: The utilization of artificial intelligence (AI) across various sectors within the pharmaceutical field, ranging from drug exploration to the management of pharmaceutical products.

2.3.1 Quantitative Structure-Activity Relationships (QSAR)

It is a modelling approach used to establish relationships between structural features and biological activities of molecules. It relies mainly on electronic, hydrophobic and steric attributes and other structural, quantum and mechanical descriptors.

QSAR employ hydrophobicity-related descriptors, as hydrophobicity is crucial in many biological processes such as receptor-ligand interactions. Selecting a key descriptor representing a ligand is critical in QSAR studies and associating patterns with activity (predictive fingerprints). AI approaches such as feature selection, pattern recognition, classification and clustering are applied to QSAR. Machine learning methods such as hierarchical divisive clustering have been utilised to cluster receptor proteins based on structural similarity, improving docking studies and drug design.[9]

Other methods used in QSAR studies include:

- ML-aided clustering, consensus k-nearest neighbour(kNN), Particle swarm optimization (PSO) and Naive Bayes methods.
- Neural Networks in virtual screening combined with docking

An approach that combined neural networks with docking was implemented to virtually screen anti-HIV-1 RT and anti-HIV-1 PR inhibitors from the Thai medicinal plant's database. The study generated a self-organizing map for evaluating potential inhibitors against HIV-1 for two specific targets: HIV-1 RT and HIV-1 PR. The map of reference structure was superimposed on feature maps of screened compounds to isolate structures having similar features to reference compounds. Self-organizing maps reduced the number of candidates for testing HIV-1 inhibitory affinities. This approach reduced and allowed successive screening with increased efficiency.[11]

2.4 Precision Medicine

Precision medicine, commonly called personalised medicine, is a cutting-edge method of medical care and treatment that considers each person's unique genetic makeup, environmental factors, and way of life. Instead of using a one-size-fits-all strategy, precision medicine aims to customise medical care and therapies for each patient. Precision medicine and artificial intelligence (AI) are interwoven and can complement one another's abilities as seen in **Figure 5**. AI technologies, such as machine learning and data analytics greatly aid the development of precision medicine. Various disorders and diseases such as cardiovascular issues, involve the associations between factors such as gender, lifestyle, environmental and genetic factors. Predictive modelling, an AI-assisted approach is highly viable in such scenarios that require detecting complex relationships from a large dataset (factors). Using AI to detect phenotype characteristics from electronic health records or images integrated with genetic data may expedite the diagnosis of genetic diseases.

AI-enhanced live monitoring for evaluating intrapartum stress can impact the choice between cesarean and regular vaginal deliveries during labour. This method reduces human error in the interpretation of data, significantly decreasing perinatal complications and stillbirths. This approach is also being utilised in colonoscopy for detection and characterisation. AI-driven image analysis is revolutionising disease risk prognosis and diagnosis.

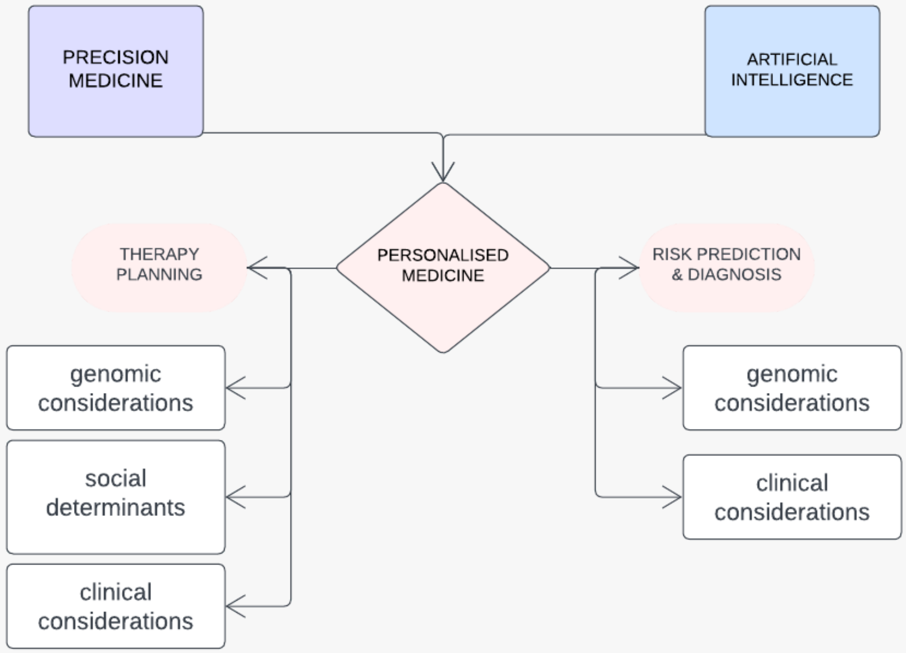

Figure 5: The collaboration between AI and precision medicine involves five critical ways of customizing care: creating treatments based on clinical, genomic, social, and behavioral factors; as well as predicting and diagnosing risks using genomic or other factors.

AI-integrated wearable monitoring devices are also rapidly changing how we interpret disease monitoring concerning health conditions such as diabetes, epilepsy, pain management, Parkinson's disease, cardiovascular disease, sleep disorders, and obesity. Digital biomarkers facilitate the monitoring of diseases remotely beyond the confines of a healthcare institute. Continuous monitoring with this approach allows the identification of minimal residual disease and helps track the progression of the disease. Subsequently, these approaches support decentralised clinical trials.[3]

2.5 Cognitive Agriculture (Cognac)

With the rapid increase in global food demand, against the volatile climatic conditions, the field of Agriculture requires revolutionary methods to enhance productivity, sustainability and resource management. The integration of AI attempts to meet these goals while addressing resource scarcity and environmental impact.

Eight Fraunhofer institutes developed a project called "Cognitive Agriculture".[12] The research attempted to optimise the production of agri-products prioritizing resource management, efficiency and sustainability. The project focused on three main areas of innovation:

- Agricultural data space(ADS): Development of a data space specifically for the agricultural sector to increase accessibility to data. It allows open data exchange in a secure digital network system. ADS will utilise AI methods for data analysis and processing to enable users to make more informed decisions.
- Innovative sensor systems: Methods and processes in the agricultural sector are highly dependent on cause-effect relationships impacted by various geographical factors. The collection of comprehensive data is crucial to optimise production methods. High-resolution measurement data through airborne and terrestrial systems and automated interpretation allow continuous data collection and field management. Such systems are already being applied for seismic imaging of soil compaction, analysis of soil nitrogen content, and classification of soil areas. Innovative automation The concepts aim to implement field robotics to aid autonomous and continuous data collection. Swarm intelligence using small robots and drones can be integrated with innovative sensor solutions to enable monitoring of crops and fieldwork.

Cognitive agriculture goes beyond the project, focusing on holistic methods to improve decision-making and resource utilisation through automated intelligent sensors, predictive analytics, and precision agriculture.[13]

2.6. Plutchik

As research progresses, large amounts of data are being processed every day. The ability to search across the databases for clinical and research practices is critical and has extreme potential. Plutchik is a voice-enabled Artificial Intelligence chatbot that can conduct highly technical medical searches across National Center for Biotechnology

Information (NCBI) databases. Plutchik enables AI-driven gene editing and drug discovery.
Project Plutchik is named after Robert Plutchik (1927-2006), a renowned psychologist. The chatbot is designed to retrieve, analyze and communicate critical information in a medical forum. Moreover, the chatbot is also capable of interacting with healthcare providers in both natural language and through voice via 3D facial expressions and gestures.[14]

3. **The past, the present, and the future**

Artificial Intelligence has seen immense growth in the past decade. From revolutionising the scientific process of research and discovery to enabling the design and creation of novel products and services, AI has impacted the biomedical industry immensely. Integration of AI not only provides unprecedented opportunities but also counters the risks of human error in critical decision-making and processes.[3]
However, the integration of AI also enables multi-modal data integration, security issues and federated learning which raises concerns regarding security, privacy and human rights. The situation can be analysed through three main principles:

- Data & Security: As the incorporation of AI in the industry rapidly increases, we must establish transparency with data and trust the data analysis and processing methods utilised by AI.
- Analytics & Insights: The capability of AI to process data from structured and unstructured sources and to establish reasoning and critical thinking at a staggering rate, might promote AI replacing humans. It is critical that augmented intelligence support humans' capacity and not replace them. While Artificial Intelligence may increasingly gain expertise in cognitive characteristics such as understanding, reasoning, learning and empowering, humans hold the higher ground regarding common sense, empathy, morality and creativity.

- Shared expertise: As mentioned before, knowledge transactions with Artificial Intelligence must be applied to create complimentary relationships to enable advances in research.

Friedman's fundamental theorem of informatics which stipulates how an individual with an information resource is better than an unassisted individual enables us to understand how scientific research and practice, especially in the healthcare system, may be better with the assistance of Artificial Intelligence.[15]

Despite the enormous progress that has been made in the field of AI in society and its contribution to bettering the therapeutic process, it is not available to all societies. Many low-income and developing nations continue to lack access to cutting-edge technology. It should be highlighted that there are several issues we face when employing AI, including ethical conundrums, privacy and data protection, informed permission, social gaps, medical consultation, empathy, and sympathy. Therefore, professionals and practitioners should consider the four medical ethics principles, including autonomy, beneficence, nonmaleficence, and justice in all facets of healthcare, before integrating artificial intelligence with the healthcare system.[16]

5. CONCLUSION

The integration of AI technologies in biotechnology was explored throughout the chapter, ranging from establishing management systems for data to providing active assistance to experts in decision-making and extending to cyber-physical interactions(e.g. targeted nanobots for drug delivery). Emerging computational advances in natural language processing (NLP), pattern recognition, effective search, prediction, and bias-free reasoning will only further AI capabilities that tackle now unsolvable issues. As we progress in research and development in Artificial intelligence-integrated science, we must recognise that humans hold a higher ground in ethical reasoning and empathy. Artificial intelligence must be utilised to support the capacity of humans. Moreover, it is necessary to ensure that accessibility to AI-assisted research does not become another facet of societal discrimination.

REFERENCES:

[1]Holzinger, A., Keiblinger, K., Holub, P., Zatloukal, K., & Müller, H. (2023). AI for life: Trends in artificial intelligence for biotechnology. New Biotechnology, 74, 16-24.

[2]Liu, X., Nie, Y., & Wu, X. L. (2023). Predicting microbial community compositions in wastewater treatment plants using artificial neural networks. Microbiome, 11(1), 1-15.

[3]Johnson, K. B., Wei, W. Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., ... & Snowdon, J. L. (2021). Precision medicine, AI, and the future of personalized health care. Clinical and translational science, 14(1), 86-93.

[4]Sadilek, A., Caty, S., DiPrete, L., Mansour, R., Schenk Jr, T., Bergtholdt, M., ... & Gabrilovich, E. (2018). Machine-learned epidemiology: real-time detection of foodborne illness at scale. NPJ digital medicine, 1(1), 36.

[5]Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology, 284(2), 574-582.

[6]Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H. J., & Kim, N. (2019). Deep learning in medical imaging. Neurospine, 16(4), 657.

[7]Galal, A., Talal, M., & Moustafa, A. (2022). Applications of machine learning in metabolomics: Disease modeling and classification. Frontiers in genetics, 13, 1017340.

[8]Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., ... & Guo, T. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. Cell, 182(1), 59-72.

[9]Duch, W., Swaminathan, K., & Meller, J. (2007). Artificial intelligence approaches for rational drug design and discovery. Current pharmaceutical design, 13(14), 1497-1508.

[10]Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. Drug discovery today, 26(1), 80.

[11]Sangma, C., Chuakheaw, D., Jongkon, N., Saenbandit, K., Nunrium, P., Uthayopas, P., & Hannongbua, S. (2005). Virtual screening for anti-HIV-1 RT and anti-HIV-1 PR inhibitors from the Thai medicinal plants database: a combined docking with neural networks approach. Combinatorial Chemistry & High Throughput Screening, 8(5), 417-429.

[12]Falcão, R. (2023, July). Fraunhofer IESE, Kaiserslautern, Germany {rodrigo. falcao, raghad. matar, bernd. rauch}@ iese. fraunhofer. de. In Software Architecture. ECSA 2022 Tracks and Workshops: Prague, Czech Republic, September 19–23, 2022, Revised Selected Papers (Vol. 13928, p. 483). Springer Nature.

[13]Pathan, M., Patel, N., Yagnik, H., & Shah, M. (2020). Artificial cognition for applications in smart agriculture: A comprehensive review. Artificial Intelligence in Agriculture, 4, 81-95.

[14]Bohle, S. (2018). "Plutchik": artificial intelligence chatbot for searching NCBI databases. Journal of the Medical Library Association: JMLA, 106(4), 501.

[15]Friedman, C. P. (2009). A "fundamental theorem" of biomedical informatics. Journal of the American medical Informatics association, 16(2), 169-170.

[16]Farhud, D. D., & Zokaei, S. (2021). Ethical issues of artificial intelligence in medicine and healthcare. Iranian Journal of Public Health, 50(11), i.