ARTIFICIAL INTELLIGENCE IN LIFE SCIENCES - THE NEXT FRONTIER, DECODING THE SECRETS OF LIFE

Dr. S. Vanitha
Associate Professor
Department of Biochemistry & Nutrition
Bhavan's Vivekananda College of Science, Humanities and Commerce
Sainikpuri, Secunderabad, India

Ms. A. Vijaya Lakshmi & Mr. Sagar Dakua
Department of Biochemistry & Nutrition
Bhavan's Vivekananda College of Science, Humanities and Commerce
Sainikpuri, Secunderabad, India

https://doi.org/10.34293/ctbtls.2025.ch007

Abstract

The scientific landscape is being reshaped by emerging technologies and their cutting-edge applications superseded by the State-of-the-art techniques. One such technology revolutionizing the world is Artificial Intelligence (AI). It is a field that encompasses a wide array of computational approaches and algorithms capable of mimicking complex cognitive skills and functions of human. AI encompasses Machine Learning (ML), and ML encompasses Deep Learning (DL), whereas Next Generation Sequencing (NGS) gets benefitted from AI/ML/DL for data analysis. Machine learning has applications in both personalized medicine and gene editing. It can predict patient-specific treatment responses based on genomic data. Moreover, they are instrumental in optimizing gene editing technologies such as CRISPR-Cas9 by predicting on- and off-target effects in test datasets, thereby facilitating the design of guide RNAs (gRNAs) that minimize off-target activity. ML can also be applied to organoid technology. Organoids are three-dimensional structures, cultivated from stem cells, that aim to recreate the intricate cellular arrangements and 3D architectures characteristic of natural organs, there by simplifying the study of organ complexity. Next-generation sequencing (NGS), facilitates the simultaneous sequencing of millions of DNA fragments, as it plays a crucial role in deciphering the genetic heterogeneity of diseases. DL leverages AI to analyse healthcare data and drive improvements in patient care. By leveraging real-world data (RWD), AI can also streamline clinical research and bridge the gap between research and real-world practice. Deep learning models learn from real-world data, enabling them to recognize patterns and predict outcomes in similar real-world situations. AI, encompassing machine learning and deep learning, is transforming various fields, including healthcare, where it complements powerful tools like next-generation sequencing (NGS) to advance disease understanding, personalized medicine, and drug discovery, improving early detection and streamlining clinical research with real-world data.

Keywords: *Artificial intelligence, Clinical research, CRISPR Cas9, Deep Learning, Machine learning, Organoid*

1. Introduction

AI/ML is revolutionizing healthcare by providing actionable insights from complex data. These technologies predict DNA sequence function from next-generation sequencing data using methods like Convolutional Neural Networks (CNNs). They also advance genome editing by predicting effective gRNAs (guide RNA) for Clustered Regularly

Interspaced Short Palindromic Repeats (CRISPR). Additionally, AI/ML compares organoid and real organ RNA-sequence data and improves NGS-based diagnostic accuracy. When combined with RWD, which reflects actual patient experiences and outcomes, these technologies can provide a comprehensive view of disease, leading to more personalized and effective treatments.

AI can help in maintain and even increase crop yields in the face of increasingly unpredictable environmental conditions by enabling farmers to adapt their practices to challenging climate change (Holzinger *et al*., 2023). Artificial Neural Networks (ANN) method is used to predict mortality after cardiac surgery, which will also identify and rank the most influential risk factors (Nilsson *et al*., 2006)

Five machine learning algorithms, namely deep neural network, support vector regression, decision tree regression, random forest regression, and gradient boosting regression, for real time prediction of process attributes during continuous production of monoclonal antibodies has been reported (Nikita *et al*., 2022).

According to Bloomberg Technology, Microsoft has been working on Project Hanover, an artificial intelligence system that aims to support oncologists by processing extensive cancer treatment databases. Hanover's goal is to predict the most effective combination of drugs for each patient, based on their individual case (Agrawal., 2018). *In silico* Medicine's AI-designed cancer drug, ISM3412, has received FDA approval for clinical trials, marking a milestone in AI-driven drug discovery.

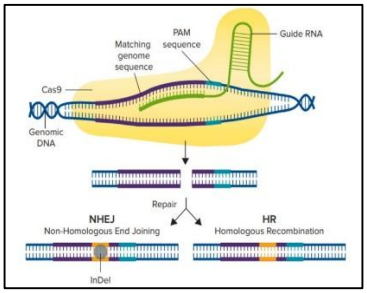
2. Methodology

The methodology involved a systematic approach in literature search, and selection and data extraction. The primary databases utilized for this search included PubMed, Google Scholar, and ScienceDirect. The search was conducted using a combination of keywords such as AI in organ transplantation, machine learning for analysing next generation sequencing data, AI in organoid technology, CRISPR-Cas9 and artificial intelligence (AI). This comprehensive search aimed to capture a wide range of studies relevant to the application of AI and ML in various healthcare domains.

3. Discussion

3.1 Crispr Cas and Artificial Intelligence (AI)

Introduction of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) associated protein 9 (CRISPR-Cas9) based gene editing tools helped in improving the accuracy and precision in gene editing. CRISPR Cas is a method that uses gRNA [Fig 1].



(Image Source- https://images.app.goo.gl/k1ogmZDceHVprxMhu5)

Fig 1. Illustration of CRISPR Cas 9 Editing and Repair Method that Involves gRNA, Cas Protein, Protospacer Adjacent Motifs (PAM) and Matching Genome Sequence. The Repair Method Involves Two Ways 1) NHEJ (Non-Homologous End Joining) in which the Non-Homologous End is Joined without Homologous Template 2) HR (Homologous Recombination) in which the Broken Ends are Joined with a Homologous Template.

But it is generally observed that, sometimes gRNA will show off-target effects. Due to this, the guide RNA (gRNA) mistakenly directs the Cas9 enzyme to cleave DNA at a site in the genome that is not the intended target, leading to unintended genetic modifications at a similar but not identical sequence, potentially causing adverse effects due to this unintended DNA damage.

Implementation of AI tool to overcome this, can enhance the accuracy and precision in gene editing. ML models can be created to predict and analyse various on and off target effects. There are various algorithm which have been developed to predict and maintain the on-target efficacy of CRISPR like azimuth, CRISPRScan, DeepCRISPR, DeepCas9 etc (Bhat *et al*., 2022)

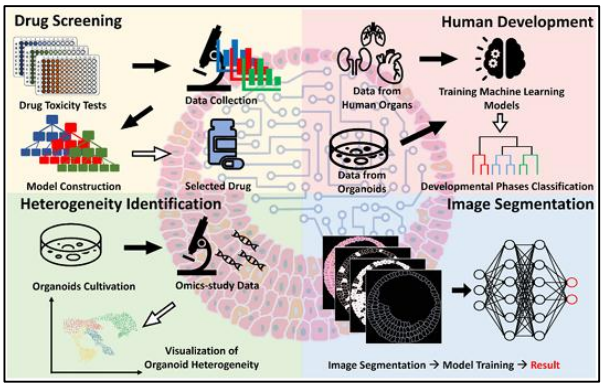
3.2 Machine Learning in Organoid Technology

Organoid technology is the 3D cultivation of humanlike structures of organ or tissues from the stem cells or progenitor cells. This mimics the complex human tissue and provides the understanding of these structures *in vitro*. For example, the human induced pluripotent stem cell (*hi*PSCs) derived cardiac organoid was done using spatially patterned Polyethylene glycol (PEG) to understand the early development of cardiovascular tissues. The application of machine learning has elevated the confines of understanding this technology.

The Machine Learning algorithms used in this technology can be divided into two parts

- Supervised Machine Learning
- Unsupervised Machine Learning

For supervised machine learning, algorithms train models using labelled data, where inputs are paired with corresponding target labels or outcomes. During the late 1950s, one of the early supervised learning models emerged with a single-layer neural network tailored for classification tasks. As computational capabilities advanced, a diverse array of supervised machine learning models, including Decision Trees, the k-Nearest Neighbour algorithm (k-NN), Support Vector Machines (SVMs), and Naive Bayes, were developed and refined. This refinement led to enhanced predictive accuracy for classification tasks. In contrast, unsupervised machine learning algorithms are trained on unlabelled data and focus on revealing patterns, structures, or relationships within inputs without predefined categories. The evolution of unsupervised machine learning algorithms is closely intertwined with the broader landscape of AI and machine learning. An early example, the k-Means algorithm, proposed by Stuart Lloyd in 1957, stands as one of the earliest clustering algorithms (Shi *et al*., 2024) [Fig 2].



(Image Source: https://images.app.goo.gl/YzqJ1UzaZAJ9mPnz5)

Fig 2. Illustration of Workflow of AI/ML in Organoid Technology for the Designing of Models from Organoid Data, Selecting Drugs from Data Collected from the Toxicity Test, Model Training from Image Segmentation and Showing Organoid Heterogenicity from the Omics Data Collected.

3.3 Next Generation Sequencing (NGS)

Next-generation sequencing (NGS) is a technology that analyses genetic material by sequencing DNA or RNA. It can sequence millions of DNA fragments at once, which allows researchers to study the genome in greater detail. The DNA is obtained from the nucleus of the cell of any region or tissue of interest. NGS works in the below workflow [Fig 3],

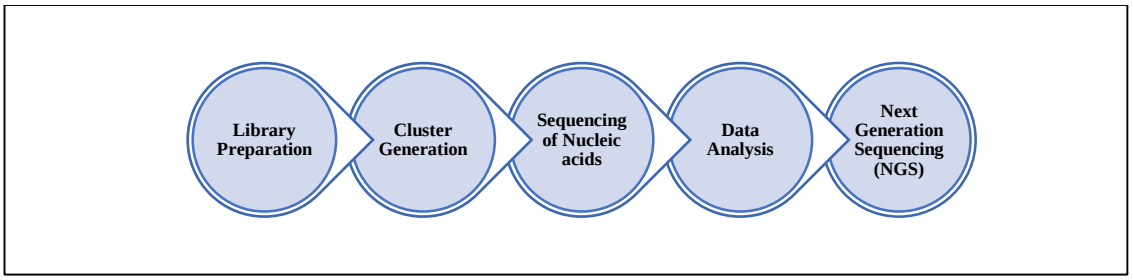


Fig 3. Workflow of NGS

Library Preparation

There is binding of the adapters to the end of the DNA fragments. The adapters are including sequencing initiation sites, indexes for molecular barcoding and region that will bind to complementary strand on the flow cell.

Cluster Generation

It is the process in which there is amplification of the individual fragment on the flow cell surface. Flow cell surface is coated with two types of oligonucleotides. The DNA fragments (library) will hybridize by binding to one oligo with the help of polymerase. This process is called exclusion amplification.

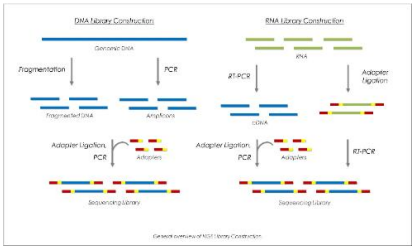
Now the other end will hybridize with another oligo creating a complementary strand this creates two identical copies attached to the flow cells. Like this many copies are generated creating a cluster of same type of strands. After amplification the reverse strands are removed, and only forward strands are left. 3′ ends are blocked to prevent undesired priming.

Sequencing

It begins with extension of the first sequencing primers to produce the first read. With each cycle, a mix of fluorescently tagged nucleotides are available for the binding. Only the complementary nucleotide will bind to the strand of template and the light is emitted by the excitation *via* a light source which results in the emission of a characteristic light source. This is called as sequencing by synthesis. The number of cycles determine the length of the read. The emission wavelength along the signal intensity determines the base cell. All identical sequences in each cluster are read simultaneously. After completion of the first read, the read product is washed away and the index read one primer is introduced and hybridize to the template and generates an index read which is like the first read and then this read product is washed away. The template folds and bind to second oligo on the flow cell. Polymerase will create a reverse strand oligo forming a double strand bridge. Then the dsDNA is linearized, and the original forward strand is removed. Index two is read as same as index one with the introduction of index two sequencing primer and then the product is washed away. Read two begins with read two sequencing primers, as read one the steps are repeated till desired length is reached. This process generates billions of reads representing all the fragments

Data Analysis

Sequences from the pooled library are separated based on the unique indexes introduced during library preparation. Based on the library preparation and the chosen sequencing application processes, downstream processes can either assemble the sequences to reconstruct the original sample without the need of the reference sequences, map and align sequences to a reference genome to identify regions that differ from the reference genome to identify the variations or count molecules *via* sequencing read to assess abundance and monitor gene expression (Choon *et al*., 2024) [Fig 4].



(Image Source: https://images.app.goo.gl/naGWEm6uZrwquVdh7)

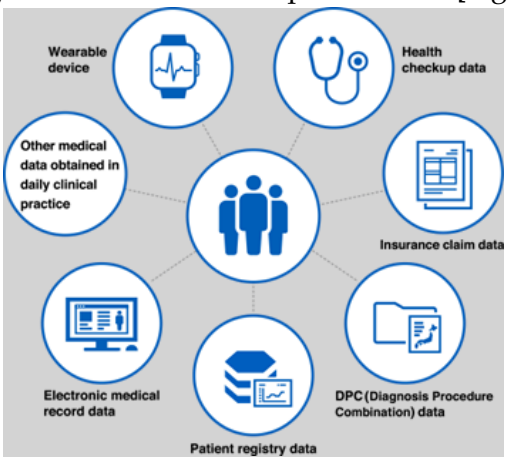
Fig 4. Illustration of the Workflow of the NGS Process. Library Construction of Two Types 1) DNA Library 2) RNA Library. In DNA Library, Fragmentaion Followed by Adapter Ligation and then Sequencing Library. In RNA Library, Adapter Ligation Followed by RT-PCR Followed by Sequencing Library Generation.

3.4 Real World Data (RWD)

Real world data is the collection of extensive data collected or accumalated from various sources under the name of same individual. It is a data relating to the patient's health status. Healthcare routine collected data sources includes

1. Clinical data which contains lab results, procedures, histology etc.
2. Medication which a patient is administered, told to be taken, or the therapies the patient is undergoing.
3. Molecular profiling that contains the genomic and genetic data, with other omics data.
4. Family history that will give an idea of the individual family health history, allergies being inherited on from generations.
5. Environmental data that will show the climatic factors, pollutants, or the lifestyle that could have contributed to this condition.
6. Social media that will show the blog or other social media accounts that is being used by the individual (Liu & Panagiotakos., 2022).

These data are collected by various sources is depicted in the [Fig 5]



(Image Source: https://images.app.goo.gl/LSAgMVMA7tCKpSGo8)

Fig 5. Sources of Collection of data of a given individual

The data being collected can be of different types such as

- Data collected in a controlled manner
- Data collected can be unstructured that contains text, image, graphs etc
- Data that is collected are of high frequency provided by the wearables.

4. Conclusion

CRISPR Cas9 can be benefitted by using AI algorithms and using programs and software applications to enhance the accuracy of the process. In addition to its use in gene editing, ML can also be applied to organoid technology, for data analysis using Next Generation Sequencing (NGS) all gets benefitted from AI/ML/DL algorithms. It can predict patient-specific treatment responses based on genomic data.

Real world data can provide real time data of the respective patients in an extensive manner. With the help of AI and ML the process with tedious efforts and time-consuming procedures can be completed in a very less time with much easier way without any hassle. The complex technology can be understood in an easier way and accurately with the use of AI.

Acknowledgement

Our endless thanks to our Management, Principal and Dept of Biochemistry & Nutrition, Bhavan's Vivekananda College for supporting us in presenting this poster. Special thanks to our guide Dr. Vanitha S for her support and help in completing this paper.

References

1. Agrawal, P. (2018). Artificial intelligence in drug discovery and development. *Journal of Pharmacovigilance*, *6*(2), 1000e173. DOI: 10.4172/2329-6887.1000e173
2. Bhat, A. A., Nisar, S., Mukherjee, S., Saha, N., Yarravarapu, N., Lone, S. N., ... & Haris, M. (2022). Integration of CRISPR/Cas9 with artificial intelligence for improved cancer therapeutics. *Journal of translational medicine*, *20*(1), 534. https://doi.org/10.1186/s12967-022-03765-1
3. Choon, Y. W., Choon, Y. F., Nasarudin, N. A., Al Jasmi, F., Remli, M. A., Alkayali, M. H., & Mohamad, M. S. (2024). Artificial intelligence and database for NGS-based diagnosis in rare disease. *Frontiers in Genetics*, *14*, 1258083. https://doi.org/10.3389/fgene.2023.1258083
4. 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. doi.org/10.1016/j.nbt.2023.02.001
5. Liu, F., & Panagiotakos, D. (2022). Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Medical Research Methodology*, *22*(1), 287. https://doi.org/10.1186/s12874-022-01768-6
6. Nikita, S., Thakur, G., Jesubalan, N. G., Kulkarni, A., Yezhuvath, V. B., & Rathore, A. S. (2022). AI-ML applications in bioprocessing: ML as an enabler of real time quality prediction in continuous manufacturing of mAbs. *Computers & Chemical Engineering*, *164*, 107896. doi.org/10.1016/j.compchemeng.2022.107896
7. Nilsson, J., Ohlsson, M., Thulin, L., Höglund, P., Nashef, S. A., & Brandt, J. (2006). Risk factor identification and mortality prediction in cardiac surgery using artificial neural networks. *The Journal of thoracic and cardiovascular surgery*, *132*(1), 12-19. doi.org/10.1016 /j.jtcvs.2005.12.055
8. Shi, H., Kowalczewski, A., Vu, D., Liu, X., Salekin, A., Yang, H., & Ma, Z. (2024). Organoid intelligence: Integration of organoid technology and artificial intelligence in the new era of in vitro models. *Medicine in novel technology and devices*, *21*, 100276. https://doi.org/10.1016/j.medntd.2023.100276