Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ31044

Bridging Algorithms and Big Pharma: Review of Machine Learning Techniques and AI-Biotech Collaborations with Multi-national Companies

Jianuo Zhu

Jserra Catholic High School, San Juan Capistrano, USA
Jianuo.Zhu@jserra.org

Abstract. The use of artificial intelligence (AI) and machine learning (ML) have revolutionized biopharmaceutical research and development. AI has facilitated drug design, increased clinical-development efficiency, and led to more personalized patient care. At the same time, advancement of ML algorithms ranging from margin-based classifiers to deep neural networks have makes it possible to learn from heterogeneous high-dimensional biomedical data, which includes omics profiles, medical images, and real-world evidence. To convert algorithms into actual clinical impact, it is important to understand the technical underpinnings of ML and the emerging collaboration models between AI biotech firms and MNCs. In this review work, we identify core ML methods widely used in biopharma and give a brief introduction to each algorithm in the first part. Examples include support vector machines (SVM), random forests (RF), and artificial neural networks (ANNs). In the second part, we explore representative collaborations between AI-focused biotech companies and major MNCs, including Pfizer, Bristol Myers Squibb (BMS), Roche, AstraZeneca, and Janssen across drug discovery, clinical development, digital pathology, and real-world data analytics and show how they are changing the biopharma value chain.

Keywords: Machine learning, Artificial intelligence, Bio-medicine

1. Introduction

Artificial intelligence (AI) has become an indispensable force in transforming the landscape of modern biotechnology and pharmaceutical research. At the heart of this transformation are machine learning (ML) models—ranging from traditional algorithms such as linear regression, decision trees, and support vector machines, to more advanced architectures like convolutional and transformer-based neural networks. Machine learning learns patterns from labeled training datasets, and the resulting trained model can then be used to predict previously unseen new samples. These models enable the rapid analysis of vast biological datasets, uncovering patterns that would otherwise remain hidden through conventional experimental approaches. By learning from molecular structures, genetic data, and clinical outcomes, ML systems can now predict drug-target interactions, optimize compound design, and accelerate clinical trial processes with unprecedented precision.

DOI: 10.54254/2755-2721/2026.TJ31044

In recent years, collaborations between AI-driven biotechnology startups and multinational pharmaceutical corporations (MNCs) have emerged as one of the most promising forces shaping the next generation of drug discovery and healthcare innovation [1]. Startups bring cutting-edge algorithmic expertise, agile data pipelines, and deep specialization in AI model development, while MNCs contribute large-scale biological databases, regulatory experience, and global commercialization capacity. The synergy between these two sectors has led to breakthroughs in areas such as generative drug design, digital pathology, biomarker discovery, and real-world evidence integration. Through joint ventures, strategic partnerships, and multi-year research alliances, AI-biotech firms and MNCs are jointly redefining the speed, cost, and accuracy of pharmaceutical innovation.

This review paper begins by outlining the fundamental concepts of machine learning models relevant to biomedical applications, establishing the technical foundation necessary to understand their role in drug discovery and development. Support vector machine (SVM), Random Forrest and ANN are introduced as example models in the following section. It then provides a comprehensive analysis of key collaborations between AI biotech companies and major pharmaceutical MNCs, examining how these partnerships combine data, computation, and biological expertise to create transformative value across the biopharma ecosystem.

2. Machine learning techniques review

2.1. Structure

Support Vector Machine (SVM) is a method that enables computers to learn to "distinguish things", such as differentiating apples from oranges [2]. It will draw a line on the graph, making this line exactly separate the two types of data, and it should be as middle as possible. The points closest to the dividing line are called support vectors, which determine the position of the line. If the two types of data cannot be separated by a straight line, SVM will use a kernel function technique to transform the data to a higher-dimensional space, such as from two dimensions to three dimensions, where a line that can separate them can be found.

2.2. Random Forest (RF)

Random Forest is an ensemble learning method that accomplishes predictions by establishing multiple decision trees and having them vote together [3]. Decision trees are the core component of Random Forest, equivalent to its basic unit. It is a model that makes predictions by splitting data layer by layer. During this process, a tree will determine which branch the sample belongs to base on the input features and make the judgment step by step until reach the final label end. Decision tree also can improve the prediction accuracy mainly because it divides the characteristics of substances layer by layer and decomposes complex data problems into multiple simple decision-making steps, thereby reducing cumbersome work. Each tree uses different data aspects and features during training, so the overall model can reduce simulation time, improve efficiency and stability.

It is not just one tree making decisions, but many trees participating together and voting to determine the result. Each tree will be trained on randomly selected data and features to obtain different judgment methods. Finally, combine the results of all the trees and select the one that gets the most votes. This mechanism can reduce errors and make predictions more stable and accurate. Even if the final judgment of some trees is wrong, the judgment of a large number of trees can compensate for it.

DOI: 10.54254/2755-2721/2026.TJ31044

2.3. Artificial Neural Network (ANN)

• Artificial Neural Network (ANN) is an algorithm that mimics the way the human brain thinks [4]. It is composed of many neuron nodes, each of which receives information, performs calculations, and then passes it on to the next layer, and so on. By iteratively adjusting the weights of these connections, the model can learn to recognize patterns, such as distinguishing between different categories. Specifically, this optimization method is called backpropagation, that is the machine will perform forward calculation to obtain a prediction result, and then compare the prediction result with the label to calculate the difference. Then, gradient descent will be carried out to diminish this difference in a backward direction. In simple terms, it is like a learning brain that can identify patterns and make judgments on its own from a large amount of data. For instance, if people want to use artificial neural networks to recognize faces, they need to show ANN many photos of themselves. ANN will automatically learn the features, such as eyes, nose, mouth, and other basic features, and finally be able to determine whose face photo it is.

3. MNCs collaboration with AI biotech companies

3.1. Pfizer

Pfizer, as a company which focus on Oncology, Immunology Inflammation, Vaccine, and AI digitalization. It already had multiple corporations with other companies. Nowadays Pfizer trending to utilize AI as an auxiliary strategic tool [5]. Now Pfizer already had three main corporation directions, Small-molecule design, modeling and discovery; Disease and patient modeling, real-world data, and clinical development; Earlier AI initiative and enabling platforms.

The first one small-molecule design, modeling and discovery is about utilizing AI to reduce the drug research and development cycle. Pfizer with PostEra corporate this project from 2020 to now, they are multiple year generative chemistry partnership, and they are trending their new works on ADCs [6]. ADCs, which is Antibody Drug Conjugate, connect Antibody and cytotoxic drug(chemotherapeutic drugs), target on cancer cell, trying to use antibody grab the cancer cell and use the poison to kill the cell, it is efficient to kill the cancer cell rather than the normal cell. In addition, Pfizer and XtalPi had corporation since 2018, now they are expanding to enhance Pfizer's small-molecule platform. Their direction core is using Quantum mechanics/Molecular mechanics some Physic approaches to combine with machine learning and deep learning, to predict for solubility, stability, conformation/energy, interaction and something relate to the drug ability. At the same time, they also collaborated to feature and model the platform for Pfizer's vast proprietary chemical space, generating more accurate property/structure-activity predictions with higher throughput for molecular prioritization, design iteration, and process feasibility assessment. If they success, Ai-physics hybrid modeling can significantly reduce the number of molecules to be tested, saving 1 to 2 years compare to traditional medicine research, and AI can predict solubility, stability, and toxicity in the preclinical stage, eliminated the doomed molecule and reduce the risk of costly failure in the later stage.

The second direction is Disease and patient modeling, real-world data, and clinical development, which is use AI to set Mathematical/data-driven modeling of disease processes and patient characteristics, predict which molecules and Immune cell subsets have a core affection during the disease development, then they use the real-world data to verify their opinions and check in the clinical trial. Pfizer and CytoReason corporate begun in 2019, The core technology of CytoReason is to establish computer models of the human immune system and diseases based on multi-omics and

DOI: 10.54254/2755-2721/2026.TJ31044

clinical data, simulate disease mechanisms, and predict drug responses in patient populations. Pfizer uses this platform to enhance the speed of target discovery, optimize patient stratification and clinical trial design, thereby enhancing R&D efficiency and reducing failure rates. This collaboration not only helps Pfizer to advance drug development in areas such as immunology and inflammation more quickly and accurately, but also establishes an AI-driven closed loop of "disease modeling - data validation - clinical development" for it, becoming an important support for its R&D strategy. In addition, Pfizer's collaboration with Tempus began in 2023 and is a multi-year strategic research agreement. The core objective is to leverage Tempus's AI-driven precision medicine platform and its large-scale clinical and genomic data resources to drive Pfizer's drug development and clinical trial optimization in oncology. Tempus possesses a vast amount of real-world data (RWD), covering information such as electronic medical records, gene sequencing, imaging, and treatment responses, and transforms these data into actionable insights through machine learning tools. By leveraging this platform, Pfizer can make more efficient decisions in areas such as target validation, patient stratification, clinical trial design and enrollment, thereby facilitating the advancement of candidate drugs to clinical trials and enhancing the success rate of trials. This collaboration is a key link in Pfizer's closed loop of disease and patient modeling - real-world data - clinical development, demonstrating its strategic layout of deeply integrating AI into the entire drug research and development process.

The third one earlier AI initiative and enabling platforms, which is according to corporate with technology company, build fundamental digital and intelligent R&D infrastructure and faster than the drug research in the later stage. Pfizer and IBM Watson Health corporate since 2016. Pfizer utilizes the Watson for Drug Discovery platform in Immuno-Oncology research. Watson’s core advantage is Nature language processing and machine learning, it can read and analyze massive amounts of unstructured data quickly. Through this platform, Pfizer hopes to find new research directions more quickly in the complex immune pathways and molecular interaction networks, accelerating the exploration of drug discovery and combination therapies. Essentially, this collaboration is an early attempt by Pfizer to explore its AI knowledge mining capabilities, aiming to shorten the R&D cycle, reduce the inefficient steps of manual screening, and increase the success rate in the development of oncology drugs. Pfizer and AWS started their corporation in 2019. They set a team called Pfizer-Amazon Collaboration Team(PACT), focus on scientific data cloud, integrate multimodal data from hundreds of experimental and manufacturing devices, it giant reduce scientists spend the time from data acquisition to insight output. In the innovation prototyping mechanism, AWS and Pfizer build multiple AI prototype applications, and the result is reducing the prototype development cycle, and reduce about 16000 hours for finding every year.

3.2. BMS

The main business of BMS is concentrating on the four therapeutic domains which are oncology/hematological oncology, immunology, cardiovascular and neuroscience. The company continuously strengthens its pipelines through a multimodal platform, such as cell therapy, targeted protein degradation, ADC, radiopharmaceuticals, etc. AI and data science are not independent fields but rather a universal facilitator that runs through "discovery - clinical - diagnosis - commercialization", aiming to increase hit rates, shorten cycles, and reduce uncertainties, which has the same orientation with other MNCs like Pfizer.

In the early discovery stage, the strategy of BMS is to combine "rapid exploration" and "high-precision constraints" within DMTA (Design - Synthesis - Test - Analysis) cycle. Meanwhile, the Exscientia platform focuses on using knowledge graphs and generative algorithms to propose high-

DOI: 10.54254/2755-2721/2026.TJ31044

potential molecules in a vast chemical space. The academically complementary BMS and Exscientia soon reached a collaboration [7]. They combined active learning and parameter optimization schemes to enable the model to learn from each round of experimental data and improve hit quality. In addition, BMS has also collaborated with Schrodinger's Company before. Schrodinger excelled in physical modeling and ML fusion, predicting binding affinity, stability, solubility/permeability, etc. Through the calibration of molecular mechanics and quantum chemistry, Schrodinger makes the generated candidate substances more practical. The complementarity of the two collaborators lies in that Exscientia helps BMS quickly propose enough reasonable "prototype molecules" to drive the project forward. Schrodinger helps the BMS to screen out in advance those designs that were physically unreliable or risky, reducing the need for expensive expenditures in later development stages. From the perspective of BMS, this tripartite cooperation is also conducive to the parallel processing of difficult-to-drug targets and selective/safe trading and also can share prior and parameter experience with BMS's existing targeted protein degradation (TPD), macromolecule /ADC projects. More importantly, it transforms the closed loop from model to experiment and then back to model again into a reusable process. Therefore, the pace of research and development is more stable and predictable.

In terms of clinical development, the focus of BMS is not on large models, but on transforming real-world data (RWD) and trial operations into reusable engineering capabilities. The value of the collaboration between BMS and Flatiron Health lies in Flatiron Health's high-quality tumor real-world evidence (RWE) and the extraction and quality control of tumor electronic medical records, which are used to generate structured datasets for population characteristics, external comparisons, and outcome studies. Flatiron Health helps BMS get closer to "real clinical practice" in aspects such as protocol design, endpoint selection, scale analysis, and post-marketing evidence. In addition, ConcertAI has also had a long-term collaboration with BMS. ConcertAI places greater emphasis on the use of artificial intelligence throughout the entire trial operation process. The AI conducts pre-screening before enrollment and remote follow-up quality control based on previous registration records, treatment pathways, patient capture rates and GenAI. During the research process, biases are identified as early as possible through the framework of anomaly detection and reasoning. BMS addresses the source of evidence and controllability, while ConcertAI addresses execution speed and operational flexibility. These methods enable them to find suitable patients more quickly, making the surgery go more smoothly. For BMS, this tripartite cooperation approach has significantly increased the success rate of surgeries for patients in the third stage and also benefited the general public. Furthermore, this approach can be extended from oncology to immunology, cardiology and neuroscience, forming a company-level universal clinical data capability.

The strategic significance of digital diagnosis lies in pushing the task of finding suitable patients to the very top of the medical pathway and integrating it with drugs. Viz.ai is a company that embeds artificial intelligence into the real-time workflow of hospitals. For instance, by using ECG-based algorithms to automatically screen and classify suspected patients and push high-risk individuals to professional teams, the time window from identification to referral and then to initial treatment has been significantly shortened. This has formed a natural connection with BMS's therapeutic products in the cardiovascular field, reducing missed diagnoses and delays, expanding the potential beneficiary population, and thus a cooperative relationship was established between the two sides, which has continued to this day. On the other side, the fusion of Optellum's imaging and clinical variables can help doctors identify high-risk lesions earlier and more confidently, thereby increasing the diagnosis rate of early-stage lung cancer and optimizing the starting point of treatment. Moreover, they have a large amount of patient data and have initiated a collaboration with

DOI: 10.54254/2755-2721/2026.TJ31044

BMS in April 2025. For BMS, the value of Optellum lies not only in identifying patients but also in their possession of real-world evidence. When screening, image analysis, referral and subsequent treatment are digitally recorded, enterprises can more systematically assess the impact of early detection and early treatment on outcomes and costs and reapply this set of evidence to research and development and the market. In the long run, as multi-modal learning becomes basically mature, these diagnostic entry points can also be coordinated with companion diagnostics/biomarker strategies to form a closed loop from diagnosis to treatment and then to follow-up. This not only enhances clinical value but also increases the certainty of pipeline commercialization.

3.3. Roche

Roche's overall positioning in the field of artificial intelligence is similar to that of BMS and Pfizer. They view artificial intelligence as an accelerator spanning drug discovery, diagnosis, and clinical evidence generation, rather than as an independent business. At the pharmaceutical end (Genentech), they promote molecular design and target discovery through large-scale computing platforms and generative models. The Diagnostic End integrates multiple algorithms into the pathological workflow through an open ecosystem, creating supervised and implementable digital diagnostic capabilities. The evidence side takes Flatiron as the core, which also have a close collaboration with BMS, combining high-quality real data with clinical development. The trinity integration of platform, ecosystem and evidence enables Roche not only to enhance the speed of scientific research iteration at the front end, but also to support clinical and access decisions with RWE at the back end, and to continuously progress through an open pathological ecosystem.

Since the end of 2023, Roche's Genentech division has established a strategic partnership with NVIDIA to apply generative biological models such as BioNeMo and DGX aiming to enhance computing capacity for drug discovery [8]. Roche's key does not lie in large models, but in expanding the training of Genentech's proprietary data and algorithms to form reusable model machines, thereby completing the closed loop from protein/small molecule to candidate molecule screening more quickly. The integration of native data, customized models and dedicated computing power not only increases the number of molecules generated and evaluating, but also unifies the accuracy and speed of early computational screening onto a single engineering pipeline. On the other hand, Genentech established a multi-year partnership with Recursion in 2021, leveraging Recursion's Maps of Biology and OS high-content imaging combining with machine learning systems to conduct large-scale cellular lead mining in the fields of neuroscience and oncology. They utilize the single-cell perturbation data from both sides to enhance the possibility of new target and pathway hypotheses. This approach combines the two paths of generating improve computing capabilities and phenotypic and mechanism extrapolation, jointly enriching and stabilizing the R&D chain of Genentech.

In terms of digital pathology, Roche has adopted two approaches. Firstly, in 2021, PathAI and Roche officially collaborated, embedding the PathAI algorithm into the uPath/navi® digital pathology workflow to address the implementation challenges of algorithm availability. This means that both sides can invoke PathAI's computing power and algorithm production line, forming a channel from R&D cooperation to regulatory paths, promote the smooth operation of the algorithm closed loop. Secondly, Roche has established a partnership with AWS, a cloud computing service from Amazon, in 2021. Roche has migrated most of its infrastructure and health data workloads to AWS to better leverage cloud computing capabilities for drug research and development, data analysis, and the development of medical solutions. Both parties leverage AWS's artificial intelligence, machine learning, and high-performance computing capabilities to accelerate new drug

DOI: 10.54254/2755-2721/2026.TJ31044

discovery and development, enabling pharmaceutical companies, hospitals, and researchers to collaborate in a secure environment, speed up the transformation of innovation achievements, and build a more open ecosystem. In addition, both sides integrate and analyze large-scale health data and promote the development of personalized medicine.

Depending on world evidence, Roche collaborates with Flatiron Health to continuously improve the quality and regulatory capacity of tumor RWD. In 2025, NRG Oncology joins the partnership between Roche and Flatiron Health. In multi-center clinical trials, NRG's clinical trials utilized Flatiron's product clinical pipeline technology, which is a connection technology from electronic medical records (EMR) to electronic data acquisition (EDA). The goal is to simplify and promote the data entry and management process in clinical trials. This method not only saves time cost but also reduces the cumbersome manual input, helping the personnel at the trial site to focus more on the clinical judgment of patients. Meanwhile, Roche carried out strategic projects with The Alan Turing Institute in the UK for many years. The Alan Turing Institute focuses on modeling the differences in diseases, patients, and outcomes. The institute also has robust predictive and stratified methods and provides a new tool for precise enrollment in chemotherapy. This advancing approach enables Roche to fill the gap in internal enterprise modeling at the external academic front end.

3.4. AstraZeneca

AstraZeneca regards artificial intelligence as a promotor for their entire value chain utilizing generative data at the front end to enhance target and molecular design’s speed. In the middle section, clinical development and operation are facilitated through clinical data engineering, RWE, and internal genAI tools. In the later stage, the use of digital health and remote monitoring helps to advance the identification and continuous management of appropriate patients onto the path of diagnosis and treatment. To ensure the normal operation of the functions of these three parts, AZ has specially established an internal digital department called Evinova to be fully responsible for the normal flow of these functions, which also provides a significant guarantee for AZ's important resources and cooperation.

With the existence of Evinova, AZ can collaborate with multiple companies. Among them, Benevolent AI is a long-term partner of AZ [9]. They have a unique data-driven goal approach. They first utilized multi-set learning to sort out disease pathways and candidate targets, and then transformed reading literature to search for clues into evidence chain search. Since its establishment in 2019, AstraZeneca has successively incorporated the AI goals of CKD and IPF into its product lines in 2021, indicating that this model can bring experimentally verified hypotheses to the forefront more quickly. Later, the cooperation between Benevolent AI and AZ expanded to more areas, such as immunology and cardiovascular diseases. Their essence is to use the system's data to reduce directional deviation and improve the migration accuracy from silicon to both in vitro and in vivo. In addition, AZ has also established a cooperative relationship with Absci which is adept at directly generating antibody sequences with target characteristics through zero-shot constraints, then rapidly screening and reflow data, iterating the model in a small, closed loop, and compressing the DMTA cycle. AZ and Absci share a common goal: one is to find the right place to hit the ball, and the other is to accelerate how to hit the ball, which is to make more reliable early decisions in fewer experimental rounds.

In the clinical field, AZ's philosophy is to transform data, processes and tools into auditable and reusable engineering systems. Just a few years ago, AZ and AWS from Amazon established a partnership. Together, they developed the Development Assistant, which transformed critical clinical and R&D data into a quarriable assistant. From scheme comparison to focusing on rate, from

DOI: 10.54254/2755-2721/2026.TJ31044

security monitoring to metric extraction, all these applications can be completed on the same interface. They plan to expand its usage to a thousand people scale by 2025, minimizing the time consumption from discovering numbers to using them and then to interpreting them. Meanwhile, AZ collaborates with Tempus/Pathos to establish a multimodal basic model for cancer treatment. They leverage real-world and clinically diverse data to support target discovery, patient stratification, and design risk reduction. This intermediate platform capability can connect discovery layers forward and conduct review communication and indication expansion backward, enhancing the efficiency of trials and external effectiveness. The significance of the joint efforts of the two companies lies in, on the one hand, connecting data and tools to consolidate the operational foundation, and on the other hand, accumulating transferable evidence and models to ensure that clinical development is both rapid and stable.

Outside the hospital, AZ collaborated with Humana and shared data, transferring the existing management platform. Humana has expanded its capabilities in areas such as respiration through SaMD and decentralized trials, making symptom recording, adverse reaction management, and efficacy evaluation more continuous and hierarchical. The company has also accumulated real evidence from research and development over the long period of time. Moreover, it integrates remote monitoring (RPM) into the experimental process to set alerts for specific toxicities (such as ILD) and conduct rapid intervention. Early detection and response between two offline visits can not only protect the safety of the subjects but also improve the efficiency of experimental operations. AZ tends to focus on wide coverage and long-term operation, while Humana emphasizes deep integration and process closed-loop management. Both sides are jointly committed to establish an integrated channel from early identification to rapid referral and then to stable management and feeding high-quality data back to R&D and market access.

3.5. Janssen

Janssen has utilized artificial intelligence to facilitate the generation of clinical trials and evidence, and has applied it in many aspects, such as using multimodal data and machine learning at the front end to speed up target discovery and early research in data science. In the middle part, the engineering capabilities of clinical development and RWE are utilized for verification and reuse. In the later stage, Janssen advanced the search for suitable patients to the clinical front line through digital imaging and algorithmic diagnosis. To support the operation of these functions, Janssen focuses on privacy-secure data collaboration and the co-construction of an industrial ecosystem, ensuring that these functions not only play a role in individual projects but can also be reused in a standardized and large-scale manner.

In the field of early drug discovery, Janssen focuses on data-based and general modeling as well as professional research on diseases. Intelligent Omics/Intellomx initiated a collaboration with Janssen in 2018 on the discovery of hematological tumor targets, with emphasizing both the search for potential drug targets through multi-omics and network biology [10,11]. Subsequently, this company continued its collaboration with Janssen in the field of lung cancer, demonstrating that its method can metastasize across different cancer types. The two companies have clear divisions of labor. Intellomx focuses on solidifying mechanisms at the disease level to enhance the conversion accuracy from silicon to in vitro and in vivo. Coupled with Janssen's cross-pharmaceutical company joint learning experience, the engineering capabilities of data collaboration and privacy compliance have gradually developed and got more mature.

In terms of diagnostic entry points and trial stratification, Janssen emphasizes the ability to transform algorithms into clinical usability and collaborated with Paige in 2022. The content of the

DOI: 10.54254/2755-2721/2026.TJ31044

collaboration is to introduce H&E slide direct reading AI biomarkers into bladder cancer, observing changes within seconds. This technique is beneficial for doctors to rapidly metastasize and capsule, reduces the reliance on whole exome sequencing, alleviates the pressure on tissue samples, and ADAPTS to the real workflow. In addition, Janssen's internal platform collaborates with multiple companies, integrating the Domino Data Lab and NVIDIA's entire pathology section (WSI) for deep learning. The assistance from these companies has helped Janssen to increase the model training speed approximately tenfold, and it is expected that in some trial scenarios, it will be more conducive to recruiting suitable people more quickly.

About clinical and evidence aspects, Janssen's goal is auditable, transferable and reusable, and the company initiated a long-term partnership with Tempus in 2021. Both parties collaborate on the design of patient matching and early detection solutions and provide AI/ML and diverse tumor data services to directly embed algorithms into the pre-screening process. This is done with the aim of shortening the enrollment time, increasing the success rate of the three phases, and reducing the risks of the protocol. On the other side, Janssen and Owkin also began their collaboration in 2021. The focus of the cooperation is on the method of AI-based external control arm (ECA), using de-biased machine learning techniques to adjust high-dimensional confounding factors, providing a more reliable comparison basis for small sample, rare disease or early-stage research, and reducing uncertainties in the later treatment stage. These two types of cooperation, one that leans towards data and operation and the other towards methods and causality, have jointly thickenes the evidence materials. This strategy not only enhances the efficiency of recruitment and execution but also makes regulatory communication more evidence-based and efficient.

Insert figures and tables following the paragraphs where they are mentioned. All figures and tables must be referred to in the main body of the text using “Figure 1” “Table 1”. Do not use text boxes to arrange figures or tables. All figures and tables must be set in line with text.

4. Conclusion

In this review, we examine the role of artificial intelligence and machine learning in the biopharmaceutical R&D from both technical and industrial perspectives. The first part summarized representative models such as SVM, RF, and ANNs, illustrating how they learn from high-dimensional, heterogeneous biomedical data to support target discovery, molecular property prediction, and disease phenotype characterization. The second part demonstrates collaborations between MNCs including Pfizers, BMS, Roche, AstraZeneca, and Janssen with AI-focused biotech companies showing how AI now spans the full value chain from discovery and clinical development to diagnostics and real-world evidence. Across generative molecular design, physics-ML hybrid modeling, multi-omics disease modeling, digital pathology workflows, and large-scale RWD/RWE platforms, these partnerships share a common objective: to increase hit rates, shorten development timelines, and more precisely identify patients who can truly benefit from specific treatment, all under manageable risk and within regulatory constraints.

References

[1] L. K., Gholap, A. D., Jetha, K., Thakur, R. R. S., Solanki, H. K., & Chavda, V. P. (2023). Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics, 15(7), 1916. https: //doi.org/10.3390/pharmaceutics15071916

[2] Noble, W. S. (2006). What is a support vector machine? Nature Biotechnology, 24(12), 1565–1567. https: //doi.org/10.1038/nbt1206-1565

[3] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https: //doi.org/10.1023/a: 1010933404324

DOI: 10.54254/2755-2721/2026.TJ31044

[4] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536. https: //doi.org/10.1038/323533a0
[5] Kint, S., Dolfsma, W., & Robinson, D. (2024). Strategic partnerships for AI-driven drug discovery: The role of relational dynamics. Drug Discovery Today, 29(12), 104242. https: //doi.org/10.1016/j.drudis.2024.104242
[6] Garnsey, M. R., Robinson, M. C., Nguyen, L. T., Cardin, R., Tillotson, J., Mashalidis, E., Yu, A., Aschenbrenner, L., Balesano, A., Behzadi, A., Boras, B., Chang, J. S., Eng, H., Ephron, A., Foley, T., Ford, K. K., Frick, J. M., Gibson, S., Hao, L., . . . Lee, A. A. (2024). Discovery of SARS-CoV-2 papain-like protease (PL pro) inhibitors with efficacy in a murine infection model. Science Advances, 10(35), eado4288. https: //doi.org/10.1126/sciadv.ado4288
[7] Philippidis, A. (2023). BMS collaboration paying off for exscientia. GEN Edge, 5(1), 147–150. https: //doi.org/10.1089/genedge.5.1.31
[8] Genentech and NVIDIA enter into strategic AI research collaboration to accelerate drug discovery and development. (2023, November 21). https: //www.roche.com/investors/updates/inv-update-2023-11-21
[9] Visan, A. I., & Negut, I. (2024). Integrating artificial intelligence for drug discovery in the context of revolutionizing drug delivery. Life, 14(2), 233. https: //doi.org/10.3390/life14020233
[10] Intelligent OMICS Ltd in AI Driven research collaboration with Janssen | Johnson & Johnson Innovation. (n.d.). https: //jnjinnovation.com/news/press-releases/intelligent-omics-ltd-in-ai-driven-research-collaboration-with-janssen
[11] Love Business. (1970, January 1). Intelligent OMICS Ltd in AI-driven research collaboration with Janssen. Love Business East Midlands. https: //www.lovebusinesseastmidlands.com/love-business-news/leadership-innovation/2023/09/01/intelligent-omics-ltd-in-ai-driven-research-collaboration-with-janssen/