Multi-Criteria Compound Prioritisation and Go/No-Go Signal Modelling: Machine Learning Frameworks for Enhanced Decision-Making in Drug Development

Dr. Małgorzata Michalewicz, Associate Professor of Computer Science, Warsaw University of Technology (WUT)

1. Introduction to Drug Development Decision-Making and the Role of AI

Drug development is a complex field with important decision-making processes that directly lead to the success or failure of pharmaceuticals. There is increasing interest in employing artificial intelligence (AI) to assist in these decision-making processes. AI technologies generally claim to improve the efficiency and accuracy of these processes. AI is seen to be able to scroll through vast datasets, find patterns and associations that are missed by humans. While it is clear that AI has great potential to improve the process of decision-making, it is important to match the AI technology to the process and the stakeholders involved in this process. These stakeholders are diverse, including funders, researchers, patients, physicians, regulators, and payers—some of whom have different interests in the use of AI technology in decision-making.

A number of approaches for decision-making currently exist; these approaches generally involve trying to accurately simulate the decision-making process of stakeholders. AI can address several of the limitations in these traditional methodologies, in particular, those methodologies that focus on optimization—namely decision analysis and stochastic modeling. AI can enable us to learn directly from data without the need to implement highly structured experiments and can potentially account for complex interactions between variables. However, the use of AI can introduce a range of additional challenges to decision-making, such as outputs, the need for highly skilled labor, and the resource implications of large datasets. The impact of AI on drug development decisions may occur at several time points throughout the drug development cycle. The aim of this paper is to critically review the literature on the use

of machine learning within drug development and evaluate its potential impact on the speed and efficiency of the different stages.

2. Fundamentals of Machine Learning in Pharmaceutical Research

In 2023, the global pharmaceutical industry spent roughly $174 billion on research and development, and approximately 75% of this investment did not result in a new drug. Modern machine learning techniques have shown promise in reducing the amount of time, energy, and monetary resources wasted by giving researchers the tools they need to make better-informed decisions before drugs proceed to more expensive clinical trials. This technique relies on predictions made by algorithms constructed through example pairwise similarities and differences extracted from large, high-throughput biological datasets. Machine learning can be broken down into three approaches: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is used to predict an output value when given an input. Unsupervised learning is designed to find hidden structures in unlabeled data. Reinforcement learning, the least common machine learning approach in pharmaceutical research, involves interacting with an environment to achieve a goal.

The success of the model depends on the quantity and quality of available data and the phosphorylation state of the protein. In biomedical research, building a dataset at the beginning upon which to base a model is not as simple. Data must be accumulated from independent, high-throughput laboratory studies. Unfortunately, when data is considered, the concept of big data emerged, so too did the importance of data programming to filter out human bias and check the quality of the data. A custom support vector machine-based model is automatically retrained with patients' new data daily to investigate and predict cancer disease progression. Machine learning models excel at finding complex and often hidden patterns among inputs, which is appropriate when looking at data that does not show an obvious connection between inputs. An algorithm's effectiveness can be evaluated and optimized by comparing predicted outputs against observed outputs in the training data. This process is known as model validation. The most common machine learning algorithms used in pharmaceutical research are decision trees, followed by neural networks, then support vector machines, and least commonly random forests. In pharmaceutical research, high-quality data with minimal missing values, errors, and noise leads to models with greater accuracy that

perform well on an external dataset. Frequently, standard techniques are applied to handle missing data points. There are various guardrails to consider when developing machine learning models for pharmaceutical research. Two commonly encountered guardrails that impact the performance of the model are overfitting and model interpretability. In pharmaceutical research, overfitting often occurs when data volume is too low to effectively train the model. Given a small dataset, a model can study the idiosyncrasies of the training data and produce an overly complex model that appears to perform well on the development data while actually failing to generalize to a new dataset.

3. Challenges in Drug Development Decision-Making and the Need for AI Solutions

Pharmacological research and development often involve decision-making tasks at various stages of the drug development continuum. Regulatory decision-making aims to initiate or end drug development, label a new disease, expand or remove a treatment option, and adjust the guidelines and policies. In clinical development, major decisions include identifying compounds in pre-clinical and clinical research phases that have a lower possibility for successful approval due to target-linked safety concerns, using machine learning approaches to identify efficient ways of strategic recruitment, early stopping for futility, defining subpopulations, resizing the sample, endpoints, or surrogate efficacy links in confirmatory or exploratory trials, and continually evaluating the benefit-risk profile of a compound to anchor or describe it in its dose regimen or indications. These steps have to be of high quality at the best possible time.

A multitude of show-stopping blockades are seen, including the length of clinical and pre-clinical research, as well as a set of complex laws, that an increasing amount of drugs do not offer the chance to translate into a regulatory and economic reward. A concrete illustration comes from the field of cannabinoids or cannabinoid-associated treatment drugs, which is now a subject of study and drug development in multiple cases. Modus operandi includes several medications targeting drugs that are being removed or developed in many countries as novel term 1 or term 2 drugs. The work goes through the challenges of early-stage clinical preparation for some of these new drugs, where retrospective human efficacy and early safety studies indicate inappropriate risks or prognoses for successful vectors of clinical growth. Digitally analyzing data with prognosis and diagnosis purposes can be helpful.

4. AI Models for Evaluating Risks in Pharmaceutical Projects

Over the last decade, a number of AI technologies have been developed and applied specifically for the purpose of evaluating and managing such risks in the context of pharmaceutical projects. Machine learning, and especially predictive analytics, has undergone a rapid evolution over the last decade and promises superior predictability about future events. Beyond demographic factors and depending on the utilized methodology, these AI techniques can be distinguished into very different models and approaches. Most of the models, however, belong to machine learning or even deep learning techniques.

In a drug development context, source data, risk factors, or predicted outcomes can differ quite significantly from typical uses in which these algorithms have demonstrated good predictive performance. Indeed, in some instances, it can be rather challenging to reach standardized external validation or to disclose the modeling approach. There is still a body of evidence in real-life drug development and failures that underscores the high added value of predictive models to identify projects at risk or to anticipate situations where a pharmaceutical product might fail to show its efficacy or safety. Potential benefits include savings in terms of human time and material resources and earlier detection of potential pitfalls. In other words, it is important to try and quantify the different assessed risk factors in order to possibly enhance the basis for decision-making at the different steps of the drug development process. Several case reports illustrate the added value of such simple real-world applications for decision-making. As in any predictive modeling work, it is important to consider issues regarding accuracy, performance, and reliability. Measures of these predictions, including sensitivity, specificity, positive and negative predicted values, and accuracy should be considered. Despite improvements made in the prospective validation of bio-tissue predictive models, a number of issues exist in the real world and are related to the failures associated with the objectives of true predictive model validation, which explores the introduction of the real prognostic variable(s) that were initially included in the predictive model.

5. AI Models for Evaluating Benefits in Pharmaceutical Projects

This article describes the positive aspects of pharmaceutical projects, investigations of which can be used in decision-making. One crucial type of benefit is when investigating

'potential' within a pharmaceutical project. Indeed, when a potential benefit is large enough, it can justify taking on some potential risks. Hence, evaluating both the potential benefits and risks, and not just the potential benefits as was done in the past, can most effectively support decision-making. Most potential benefits come in the form of the so-called cash flow that is related to two different types of risks: the probability that the cash flow will or will not actually materialize and specific 'kind' of characteristics or values that the cash flow will assume. AI models of all kinds have been developed that can support decision-making by making it possible to predict and quantify these kinds of 'benefits'.

Such AI models are used to predict drug efficacy, patient survival, disease progression, biomarkers as surrogate endpoints, drug approval, identifying drug responders, drug uptake and sales, patient adherence, patient health, impact of health and diseases, disease biomarkers, disease epiphenomena, disease associations, quality and probability of obtaining data of interest, etc. Several case studies and examples are reviewed where AI models were developed and applied to predict and quantify the benefits of doing drugs, ones with cancer and other drugs - in terms of their efficacy, conduct, marketing, profits, etc. It is freely acknowledged that we are developing models of greed - all the benefits we describe can either be used directly to develop drugs or can be used to market diagnostic tests that will identify those patients who would most benefit from being treated with such drugs. Both cases will maximize 'our' short and long-term sustainability; otherwise, investors will no longer support our pharmaceutical efforts. We also discuss a wide array of outstanding research barriers: easily and quickly creating different types of models successively for multiple stakeholders, including multinational reimbursement agencies; including different supportive parameters within AI models; promoting transparent communication between data and model developers and model users; 'explainability' versus interpretability of models; and other issues. Despite these and other barriers and outstanding work, this direction of research promises to revolutionize pharmaceutical projects.

6. AI Models for Evaluating Feasibility in Pharmaceutical Projects

Feasibility has a strong correlation with project outcomes. Successful operations start with a feasibility study to prevent potential problems at the very beginning of drug development. Project feasibility considerations can be divided into a variety of

categories; for the purpose of examining various potential applications of AI and machine learning in drug development decision-making, we will classify them as technical, operational, and economic. Figuratively, AI can be used equally to develop data analysis and feasibility project predictions in a modern business environment and in healthcare industries, including pharmaceuticals. In business, AI is giving organizations a radical edge over their competitors, while making clinical trials more cost-effective. In healthcare, we can and do use real-time analytics and AI for predictive modeling. Real-time analytics consist of real financial analyses that use information from the pipeline and targets based on personalized insights.

One area with potential to embed AI is in evaluating project feasibility actions. In pharmaceuticals, project feasibility factors include elements that do not change and some that can change as development continues, due either to internal actions or to changes in the world outside the organization. AI proposals are put forward for each of these feasibility types, intended to extend knowledge about the basic requirements of integrating AI models into feasible actions. It is important to understand that the more we know about a part of a project, the easier it is to plan resources. There are similarities and differences in the application of AI to identify both kinds of foundational work beyond technical solutions to integrating with TLF. AI, including machine learning, can and is being used to better assess the meaning of complex preclinical and clinical data. More particularly, there are appropriate models for decision-making in developing programs. This includes judging both the likely space in the market and the financial viability of the product throughout development, up to the completion of the program, at which point the company has to decide whether to proceed with registration, out-license, or stop the development.

7. Case Studies and Applications of Machine Learning in Drug Development Decision-Making

Here, we illustrate individual project iterations of no more than three pages that machine learning can be successfully applied as a decision tool at the pharmaceutical R&D interface. "Decision tool" can refer to predictive/response modeling, risk assessment, benefit valuation, or another step in a decision process. Off-label applications, such as pharmacoepidemiology and precision medicine, have also been presented. The output of the decision tool can be integrated with others. The decision

must have been made by someone from the pharmaceutical division or the AI or machine learning professionals themselves based on direct orders from R&D or marketing.

The previous experience on the job did not need to be successful as "lessons learned" can be more interesting. The descriptions should be clear about methods, the power of machine learning, and the results. A clear structure is also important, and the reader should understand a few key points: what was done; how the project was potentially transformative - including what was tried before (why it is groundbreaking); what was learned. The role the AI professionals had in the decision-making process is also important. What is trending: integration with existing workflows.

8. Ethical and Regulatory Considerations in Implementing AI in Pharmaceutical Projects

AI, particularly machine learning, is increasingly being used in various aspects of drug discovery and development. The effectiveness of these machine learning models is dependent on the quality and quantity of the data employed. Using AI and other technologies comes with important ethical considerations, such as transparency, accountability, and privacy. Moreover, AI algorithms are often designed to draw on large amounts of data that are often uncurated, and even when they are free of errors, they can perpetuate implicit biases. At present, little is known about how data will be used in the process of algorithm selection and how consent will be obtained from those whose data is used during drug development. This report assesses the state of the field in AI, outlines key ethical dilemmas and questions that are raised by the use of AI in pharmaceuticals, and puts forward a series of practical recommendations for governance.

Using AI technologies in pharmaceutical development raises important ethical issues, primarily around reliability, trust, data use, consent, and bias. Transparency is vital to public and professional understanding. However, there are several reasons why companies are likely to find it difficult to be more open about the AI tools they are using in drug development. One reason is commercial. Many data processes in drug development are undertaken using internal data, which companies will see as proprietary, crucial to trade secrets, and whose very value lies in its exclusivity. Trust in AI in drug development is going to depend to some extent on regulatory decision-

making. Many regulatory processes for drug development rely on proprietary data submitted by companies with little or no external scrutiny. Whether regulators will continue to trust existing AI systems is important from an ethical point of view. Other important stakeholders in this field are those working at health security agencies and other national and international bodies. This paper outlines some of the key ethical challenges facing AI in pharmaceuticals and proposes initial areas where guidance and rules can be developed. Ethical concerns surrounding the increasing use of AI in drug development are not limited to these areas, but are a starting point to stimulate discussion and debate in this area. Furthermore, AI greatly impacts various aspects of drug development, from understanding disease mechanisms to improving patient access to medicine. Drug regulatory agencies play a critical role in overseeing and providing guidance for incorporating AI technologies into the drug development process. It is important to facilitate multidisciplinary dialogue and help inform future regulatory decision-making around AI in drug development.

9. Future Direction

Potential future directions may include addressing trends such as the increase in personalized medicine and the use of real-world evidence, the exponential growth of omics data including multi-omics, the increased availability of AI applications in predicting all kinds and classes of biological activities, employing advanced predictive modeling and simulation approaches, and using next-generation pharmaceutical technologies. There is a real need to focus on the integration of these new technologies to improve decision-making throughout the R&D process. Moreover, it will be essential to continuously improve the machine learning techniques in response to how these new technologies and the pharmaceutical landscape change.

Capacity building in academia, industry, and regulatory bodies should take place to keep up with these advancements, and there will likely be a need for greater interaction between these different bodies as these methods could be largely dependent on academia for new methods, industry for the application of methods, and regulatory bodies in terms of acceptance of new guidelines. Many regulatory bodies are already working on tools or guidelines that would allow their organizations to handle the accountability and transparency needed for machine learning predictions. However, there is little relevant knowledge on the topic among regulatory bodies, which poses a

major hurdle to the implementation of such methods. Future challenges in applying these techniques are also expected to include legal issues regarding the sharing of patient data, ethical issues surrounding the use of machine prediction tools, and increased training to improve the use of machine learning among pharmaceutical professionals. The use of machine learning-based approaches in the pharmaceutical industry allows for the use of a vast number of tools that can deal with big data sets in a far less time-consuming fashion. The implications of the findings outlined here could have significant global applications in healthcare.

10. Conclusion

In this essay, we discussed the role of machine learning in enhancing decision-making for drug development. The high cost of development, coupled with the many risks and high rates of failure, paint a landscape in need of innovative solutions. We believe that artificial intelligence can address the challenges of precompetitive nature, such as changes to recruitment criterion, considered too high-risk or expensive in a standard drug development environment. Drug development is characterized by a complete dearth of data derived from similar previous decisions – no major pharma company develops exactly the same drug as another one. Therefore, when decisions inevitably turn out badly, this represents a significant area of learning that supports this field. We believe that, as these technologies mature, machine learning is capable of transforming the drug development and manufacturing industries.

It is clear that machine learning can help optimize the complex process of drug development decision-making. But no decision in drug development is made in isolation. Stakeholder collaboration is the only way to optimize the drug development processes and anticipate the needs of decision-makers more accurately. This technology is still reactive, essentially "learning from history" more effectively. Going forward, the challenge identification and significant further investment in eliminating these issues is critical. With active rather than reactive learning, adaptation to policy changes is needed. It is critical that machine learning further research in the development process. This work should be conducted with a focus on ethical considerations, such as patient privacy and decision reproducibility, and regulatory compliance. Further work is also needed to demonstrate the real-world health outcomes and the value of this information in the whole pharma lifecycle. It is clear that the use of AI in health care has a profound

longer-term vision of the efficiency of medical decision making around precision medicine and drugging. We hope that the maturing of these tools can also provide an opportunity for an innovative way to develop new areas through earlier access to more efficient potential drugs and accelerates their development. Gone are the days of empirical drug development and the "three failures rule" in drug associations. The way needs to be defined in this section to establish the basis for machine learning in drug development and in medicines.