AI in Drug Discovery and Development

Prof. Rajnish Kumar
Dept. of Pharmaceutical Engineering and Technology
IIT-(BHU), Varanasi

Week-02
Lecture-10

Welcome to the course "AI in Drug Discovery and Development." In today's session, we will talk about advantages of integration of AI in drug discovery and development. Earlier we have seen how we can integrate AI in drug discovery and development steps. And then we saw the several tools and techniques which can be used for you know accelerating drug discovery and development by the integration of artificial intelligence. And in today's lecture, we will talk about what are those advantages of integrating AI in drug discovery and development pipeline. So, by the end of this lecture, you will be able to understand the key advantages of AI integration in drug discovery and development.

Also explore some real-world examples showcasing cost and time savings and then also analyze how AI improves success rates, efficiency and affordability in drug development. So, as we have seen that the traditional drug discovery is highly challenging. It's like it costs around, you know, ranging from 1 to 2.6 billion US dollars per drug, which reaches the clinic.

And for some drugs, this cost can increase as well. For example, CNS drugs, where you have a lot of challenges to make the drug available in the brain. And then the average timeline was like really long. So it took like around it takes around 10 to 12 years. And then the failure rate was very high.

And so almost 90% of drugs they fail in clinical trial. And then the failure is was mainly due to lack of efficacy as well as occurrence of side effects of that drug actually. Some major role of AI, how AI can accelerate drug discovery in development is that it accelerates the timelines. So, instead of you know spending 10 to 12 years on discovering a drug and making it available in the clinic to the patients. We can squeeze this time to maybe somewhere around like 2 to 4 years, again that depends on case to case actually.

And it also reduces the cost because if we are doing it quickly, so it reduces the manpower cost. It reduces the experiment cost as well because in some experiments we can completely replace the experimental methods with the AI algorithms. And it can improve the success rates as well. Like we saw that in clinical trials, we can design a better clinical trial. We can also predict the outcome of the clinical trial.

We can differentiate between clinical trial responders and non-responders as well. So, that is how it improves the success rate of the clinical trial. So, these are basically three things where AI plays a very important role. So, now we will see the major advantages of integrating AI into drug discovery steps. As well as we will see some of the real world examples where people have used AI, integrated AI into drug discovery and development and they have improvised the process.

So, in the case of target identification, so the advantages are faster target discovery. So, because AI has a capability along with the hardware infrastructure we have now. So, it has a capability to rapidly sift through multiomics data that is again genomics, proteomics, transcriptomics, metabolomics, lipidomics to identify disease linked targets. and then it can also identify the hidden pattern so where the ML can uncover non-obvious target which are disease associations from the massive data set. So, it is able to identify the pattern and identify the disease association from massive data sets and thus helps us in target identification.

And then we have seen that it helps us in protein structure prediction as well, where it predicts the protein structure, exposing new druggable sites, even for the previously undruggable targets. Here we also talked about the cryptic pockets. So, with the help of all these protein structure prediction tools, we can identify those cryptic pockets as well, those allosteric pockets, cryptic pockets, which was not possible earlier actually. So, some of the real world examples are like AlphaFold, the development of AlphaFold and AlphaFold2 by DeepMind. So, it predicted three dimensional structures of around 200 million proteins.

So, now all those proteins whose 3D structures were not available like mostly like G protein coupled receptors or membrane bound receptors or large proteins. So, now the structure has become available. So, now it is accelerating the target discovery for cancer, rare diseases and more predicted structures for doing structure based drug discovery. And then InSilico medicine, so they are using its AI platform Pharma.AI and with that they have identified a novel target for fibrosis within weeks using AI driven multi-omics analysis.

And here they have used panda omics and this we will talk in detail in later sessions. So, this is also a real world example where the InSilico medicine, a pioneering company using AI in drug discovery, they have identified these novel drug targets within weeks using the multiomics data and the AI algorithms. And then we have the benevolent AI, which partnered with AstraZeneca to uncover new CKD targets, chronic kidney disease targets, leveraging AI to prioritize promising options faster. And if we talk about the target prioritization, so some of the advantages which AI has are data driven target ranking. So, now if we have, for example, 20 to 30 targets which we have identified in the target identification step.

So, now we wanted to prioritize which of those target we shall go further with. So, we can use AI. So, the AI can rank targets based on druggability, disease relevance and pathway involvement. And then we can then another advantage is the functional network analysis where the ML techniques they map protein-protein interactions to prioritize the key disease regulators. And then risk assessment as well where it can predict the target related safety risk or off target effects early on so that we reduce the chance of a failure of a drug in later clinical trial phases.

And some of the real-world examples are like open targets, which is prioritized cancer and cardiovascular disease targets by integrating genomics, chemical and clinical data set. And then we have the CanSAR, where they merged structural biology, drug screening and cancer genomics to rank oncology targets more effectively. After the target prioritization next comes is the target validation. So, some of the advantages of you know integrating AI in target validation are the in silico validation. So, we can use the AI models which can simulate target ligand interaction to confirm binding potential before the wet lab experiments And then we can have the functional assay prediction as well where the AI supports phenotypic screening and predicts cellular effects for faster target validation.

And then we have the CRISPR data integration. We can use the machine learning techniques which can enhance gene knockout screening analysis to validate the essential targets. And thus, these all advantages, they are very crucial for the target validation. And some of the real world examples are that like DepMap which integrated CRISPR knockout data and AI to validate essential cancer targets. And then you have the cell painting from Recursion Pharmaceuticals.

So, where they have developed an AI tool which analyzed high content imaging data to track phenotypic changes and then validating targets faster. Okay after the target identification and validation, so the next step is the hit identification. So, some of the advantages which AI has in the area of hit identification are the high throughput virtual screening. Because as we discussed earlier that if you wanted to screen you know compounds in high throughput screening or in physical assay, So, you need to invest a lot of money actually and not only that it takes a lot of time as well. So, just imagine you have to screen 1 million compounds in an assay setup.

So, it may take several you know months or based on your facility which we have if you have automated setup. So, it may it can be done in you know weeks as well, but still you need a lot of investment to procure all those compound libraries and then use them in screening actually. So, it can screen millions as well as billions of compounds against the target structure in days instead of months actually and even in hours actually if you are

using for example, ligand based virtual screen tools. So, you can screen billions of compounds in hours to get you know hit compounds. And then you can do the generative design because all those hit molecules which you are identifying from already existing library there might be some IP issues actually because those compounds might have been patented by some other companies.

So, you do not have a very good scope in you know getting a patent or IP rights on those molecules. Instead, if you design a molecule from scratch by using de novo generality modeling, so then you have a better chance of getting in IP rights on those molecules. So, AI can create new molecules optimized for binding affinity and drug likeness. And then you can, another advantage is the activity prediction, where the machine learning models they can forecast biological activity, reading out false positive very early in order to increase the chance of success of those molecules in the POC studies and further in the clinical trials. And some of the real-world examples are the Atomwise.

So, they screened 8 million compounds for Ebola inhibitor in days using AI powered virtual screening. and InSilico medicine so they designed AI generated molecule targeting fibrosis within weeks, progressing to preclinical studies in 18 months. So that traditionally takes more than you know, four years. So, you can see like they have reduced the time from you know, target identification to preclinical studies from 4 plus years to the to just 18 months actually. And the next is the hit to lead where the advantages are predictive SAR modelling.

So, it can rapidly map structural activity relationship to guide the chemical modification. And the lead expansion where it can generate novel analogues from hit structures, optimising for potency and selectivity as well as binding affinity prediction like those deep algorithms. They can refine the hit structures to improve binding strength and stability. And the real world examples are GraphDTA, which uses graph neural net to predict binding affinity for hit compounds, prioritizing the best leads. and the Exscientia, which optimized lead series for a cancer drug, balancing potency, selectivity and safety using the AI models.

And then you have the then next step is the lead optimization. So, several advantages of utilizing or incorporating AI into lead optimization are the multi objective optimization where it can simultaneously optimize potency, toxicity, solubility and synthetic feasibility, which is creating a balance between competing properties And then the predictive toxicology also so ML models, they can flag potential toxicity risk early on reducing costly failures. And then the retrosynthesis planning, so it can predict the best synthetic route to ensure lead compounds. These are scalable and cost effective to produce. And the real world examples are the REINVENT, for example, it used reinforcement learning to refine lead compounds for oncology and improving the drug likeness and bioavailability.

And then we have Schrodinger FEP+, which predicted the binding free energy and stability of leads, improving drug potency and selectivity. So after the lead optimization, the next step is the preclinical studies. So some of the advantages AI has in the area of preclinical studies are the faster preclinical validation where the AI models simulate drug behavior in biological systems, speeding up early efficacy testing. And the reduced animal testing because now we are replacing those animals with those computational models, AI/ML models. So, we are reducing the use of those animals in those studies actually.

So, the virtual models they predict drug responses reducing reliance on animal studies. And the further improved disease modeling where AI built more accurate disease simulations enhancing drug efficacy predictions. And the real-world example, which simulated drug behavior across multiple organs, predicting efficacy and toxicity outcomes before animal testing. And then Evotec, it used AI-powered cell modeling to predict drug performances in human tissues, reducing failed animal studies.

The next step is the ADMET studies. The first advantage or the biggest advantage I would say is that early ADMET property prediction because if before doing the assay, we know that this molecule will be soluble, permeable, BBB permeable or bioavailable that that will you know increase the chances of success to great extent. So, the AI can predict all these ADMET properties like absorption, distribution, metabolism, excretion and toxicity from molecular structure alone and so that it can reduce the late stage failures as well. Because as we have talked earlier that poor ADMET properties these are the major reason for failure of those drugs in clinical trials especially in phase 2 and phase 3 clinical trials. So, if those ADMET properties are poor ADMET properties are flagged earlier, so we are saving time and resources. And then the third advantage is the faster bioavailability optimization where the AI identifies molecular modifications to improve the solubility and permeability.

So, some of the real-world examples are, for example, AstraZeneca, which leveraged ADMET predictor to model oral bioavailability and hepatic clearance, avoiding costly lab testing. Another company called Genentech, so they used AI to flag an oncology drug's hepatotoxicity risk early. So, they stopped development before the clinical failure of that drug candidate. Coming to the safety pharmacology, so some of the advantages are like early toxicity detection. So likewise, if we can detect the toxicity early on, so that will be a savior for the failure.

So, the AI can predict cardiotoxicity, hepatotoxicity, and CNS toxicity before the clinical stages. As well as we can we can have an enhanced off target prediction means side effect predictions as well where the ML models they flag unintended protein interactions that may cause side effects. And then reduced safety related failures as well where the safety pharmacology signatures predict and prevent high risk compounds from progressing

further in the clinical trials. And some of the real world examples are like Novartis and Certara. So, they created AI models that predicted hERG channel cardiotoxicity, preventing compounds from causing QT prolongation.

And then you have the Toxicity.ai, which used machine learning to predict liver, kidney and CNS toxicity profiles early in the development. And then the Pfizer company, they built an AI based safety signature to flag off target binding, reducing late stage safety failures in the clinical trials. And then coming to the process development manufacturing scale up, so some of the advantages which AI offer are the optimized synthetic pathways where the AI predicts the most efficient chemical synthesis route. And bio process optimization where those ML models can fine tune the cell culture conditions to maximize the yield of your product. And predictive scale of modeling where it can simulate how lab-scale processes will perform at industrial scale.

And the real-time process monitoring where the AI-powered sensors detect variability in temperature, pressure and pH, ensuring consistent batch quality. And the supply chain optimization where AI forecasts raw material demands, optimizing supply chains and preventing manufacturing bottlenecks. So the real world examples are Chematica, which uses AI to design efficient synthesis pathways. Ginkgo Bioworks, which applied AI to engineer microbes for high yield production of biologicals. And Biogen, which implemented AI for real time process monitoring, enhancing, ensuring batch consistency.

Coming to the IND enabling studies, investigational new drug enabling studies, so some of the advantages it offers are the toxicity and safety prediction where machine learning models they can forecast long term toxicity risk. And then PK/PD simulation where the AI models they can predict drug absorption, distribution, metabolism and elimination in virtual human population, reducing the need for animal studies. Again, we are, you know, reducing the need of animals for doing all these preclinical studies. And stability and degradation prediction where AI forecasts how a drug degrades under different conditions.

Accelerating stability studies. And dose escalation strategy where ML models, they can simulate dose response curves to optimize first in human dosing strategies. So some of the real world examples are Formulation.ai where an optimized drug formulation for a poorly soluble API improving bioavailability without extensive lab testing can be performed. And the GastroPlus which is from the company Simulation Plus so it can simulate the human PK/PD profiles to guide IND dosing decisions cutting months of trial planning. So, these have been used extensively the GastroPlus for you know simulating PK/PD analysis.

And then next is the IND submission. So some of the advantages of integrating AI are the regulatory intelligence, like it can track the regulatory trend to help with evolving

guidelines. The gap analysis, it can identify missing sections, inconsistencies or weak points in the IND dossier before submission. Fast tracked regulatory writing where the NLP models they can assist in generating common technical document summaries for quicker approvals. Automated regulatory documentation which ensure the faster error free IND submissions and data consistency checks where machine learning ensures data integrity. So some of the real-world examples are RegDoc365, which has automated the preparation of regulatory documents for faster IND filing.

IQVIA Smart Solve, which used AI to ensure data consistency across preclinical and manufacturing reports, reducing the risk of rejection. And Cunesoft, which is an AI-powered platform that monitors regulatory changes, helping companies adapt IND submissions to new guidelines. And then after the IND submission, so the next step is like clinical trial design and planning. So, some of the advantages AI offers are the AI driven protocol design where we can simulate the trial scenarios to optimize study design endpoints and patient stratification. So even before doing it, we can simulate the trial and then we can get the best solution for designing the clinical trial.

And then we can do the feasibility prediction as well like these ML models they can forecast enrolment rates site performance and dropout risks improving site selection. And we can do the virtual patient simulation as well those AI models they create diverse virtual patient cohorts predicting responses across different demographics and conditions. And some of the real-world examples are Trial Pathfinder, Genentech, with simulated trial designs, reducing planning time and optimizing endpoints for faster regulatory approvals. Another example is Antidote Technologies, which used AI to improve protocol feasibility and ensure inclusion of underrepresented patient populations. And we have the Phesi which leveraged AI to predict trial outcomes based on historical data guiding design decisions to reduce the trial failure rates.

Now we have designed the trial. So, next step is the recruiting the patients and retaining those patients in the trial. So, some of the advantages which AI offers are the faster patient match where AI can scan electronic health records and real world data to find eligible patients faster. And then also with the personalized engagement like ML models they predict dropout risk and tailors communications to improve the patient retention because as I said earlier as well that retaining a patient in a clinical trial is another challenge actually. And then we have the decentralized trial support where AI powers remote monitoring and virtual participation expanding trial accessibility. So, some of the real-world examples are Deep6 AI, which identified eligible patients in minutes by scanning unstructured medical records and that reduced the recruitment time by almost 50%.

And here the Medidata AI, which predicted dropout risks and tailored patient

communication strategies, improving the retention rates. And then we have the TrinetX, which analyzed real-world data to speed up recruitment for rare disease trials, finding more diverse participants. Okay, once we get the clinical trial data, so now the next step is the analysis of the data. So, some of the advantages of integrating AI in clinical trial data analysis are the real time monitoring, where we can detect those protocol deviations, adverse events and anomalies as data streams in enabling faster decision making. And biomarker discovery where those ML it uncovers biomarkers linked to treatment response guiding precision medicine strategies.

And the adaptive trial insight which where AI supports interim analysis enabling protocol adjustments without compromising statistical power. And the real-world examples are Medidata Rave, which enabled real-time data monitoring, identifying protocol deviations early to prevent costly trial delays. And the GNS Healthcare, which used causal AI to discover hidden biomarkers, optimizing patient stratification during trials. And the Saama AI, which streamlined clinical data integration analysis, reducing the data cleaning times by almost 40%.

Okay further is the regulatory submission and approval. So, we can integrate AI and have the following advantages like automated submission preparation where the AI can generate clinical study reports and compiles the regulatory documentation faster. We can have the, you know, the data integrity validation where the ML models, they checks for inconsistencies across the preclinical, clinical and manufacturing data sets. And we have the regulatory intelligence as well where AI monitors evolving regulatory guidelines to ensure submissions, to ensure submissions stay compliant actually. And some of the examples are the Cunesoft which automated IND and NDA preparation reducing submission timelines. And the Celgenence CAPTIS which used NLP to ensure submission content adhered to regulatory standards reducing the risk of rejection.

And another aspect is the post market surveillance or phase four clinical trial. Where the advantage is offered by AI integration or the AI powered pharmacovigilance where we can track adverse events from real world data like electronic health records, social media, case reports to detect safety signals early. And the real world evidence generation where we can analyze the real world data to assess long term safety, effectiveness and off label use. And another advantage is the automated adverse event reporting where NLP extracts adverse event data from unstructured sources ensuring faster regulatory reporting. And some of the real-world examples are Aetion, which provided real-world evidence to support post-market safety assessment and label expansions.

Benevolent AI, which uses NLP to monitor literature and patient forms for emerging safety signals. And then we have Flatiron Health, which aggregated oncology real-world data to

assess drug performance and safety post-approval. We talked about it earlier as well that lifecycle management and drug repositioning is a very important, another important aspect of drug discovery and development. So, these are some of the advantages which AI offers. So, this is for example, AI power drug repurposing where we can use the AI for repurposing you know already approved drug for a new indication.

So it can identify new indications for existing drugs by analyzing multiomics data and literature. And as well as market intelligence where ML models they track competitor activity, pricing trends and patent landscapes to guide life cycle strategies. And the personalized treatment pathways where AI supports personalized dosing recommendations and label expansions. So, some examples are like HealX which used graph based AI to predict alternative disease indications for rare disease therapies. And Recursion Pharmaceuticals which combined AI and high content imaging to discover new uses of the shelved drugs.

So, in summary if we look at you know the impact of using AI as compared to the traditional methods. So, if you look at the cost per drug which is being you know developed and made available in the market for the patients to use so traditional methods they cost a lot of money like somewhere around you know 2.6 billion US dollars. While the AI driven methods based on if it is a kind of a repurposing or a full stack you know end-to-end drug discovery and development so it can range from 150,000 US dollar to 500 million US dollar. And some of the examples are like the InSilico medicine, they developed a fibrosis drug in less than 150,000 US dollar And the Benevolent AI so they discovered the ALS target.

The Exscientia, they designed cancer drug which costed around 500,000 US dollar. And then, Atomwise they developed an Ebola inhibitor so which costed around 250,000 US dollar. And again if you compare the time for development you know usually traditionally it takes 10 to 12 years but with the help of AI you can reduce the time maybe even up to 18 months. So, this kind of reduction of failure rate by just using AI is still, you know, you have wait for some time to see the real impact actually.

But for example Janssen's Trial360.ai platform or BERG AI platform so they can they can you know design better clinical trials and reduce the failure rate by using by integrating AI into the drug discovery and development. And then the target discovery timeline, so usually it takes 4 to 6 years, but you can reduce the target discovery step to maybe a few weeks to months. Like Benevolent AI, they identified COVID-19 target in weeks and Atomwise Ebola inhibitors, they also identified though they are targeting in very short period of time. And then likewise, the lead optimization timeline can be reduced from 2 to 3 years to just 6 to 12 months actually.

Like Excientia, they optimized lead for cancer drug in just 8 months. And InSilico medicine, they designed the AI leads for fibrosis in a very short period of time. So overall, we can see that there are huge advantages of using AI in drug discovery. But as I said, the real impact has yet to come actually and I think the combination of both AI and human intelligence that will be the you know a very good for the future of drug discovery and development. So, coming to the summary, so AI accelerates timeline by enabling faster target identification, virtual screening lead optimization, and also you know, designing clinical trials and post market surveillance. And significant cost reductions are achieved through predictive modeling and simulation techniques.

And the clinical trial success rates they improve with better patient stratification optimized designs by integrating the AI in drug discovery and development. So, in the end, I have a question for you to think on. So, how can we balance the benefits of cost savings and efficiency with ethical concerns around data privacy and model transparency in AI driven drug discovery? And these are a few, you know, references you can go through and learn more about, you know, this topic. With that, thank you.