Case Study

Presented by: Chahat (2022meb1046@iitrpr.ac.in)

PROBLEM STATEMENT

The field of pharmaceuticals faces a critical challenge in leveraging real-world evidence for data-driven decision making. With vast and diverse sources of data, ranging from clinical records to genomic information, ensuring data quality, conversion, and curation has become a daunting task. Errors and inconsistencies in these datasets hinder the timely and accurate analysis of patient outcomes, treatment effectiveness, and disease progression. I need to design an AI solution that will help solve this problem.

Dissecting the Case

STAKEHOLDERS : Patients, Doctors, Pharmaceutical Companies, Medical Authorities, Scientists/ Researchers, Hospitals & Medical Institutes

Pharmaceuticals Market size was valued at USD 209.85 billion in 2021 and is poised to grow from USD 222.4 billion in 2022 to USD 352.98 billion by 2030, growing at a CAGR of 5.9% in the forecast period (2023-2030).

With this vast and continuous growth, the competition has increases as well as the need to ensure that correct drugs and medicines are being developed.
In recent times we have seen how blunders have been caused by some drugs.

Emergence of COVID taught that the current systems are not efficient and we need better and faster strategies to provide solutions at a faster rate.

There is a need for a system that can enable fast and safe development of drugs with minimal harm and also maintaining cost effectiveness.

A lot of studies suggest that current methods are not completely reliable and misdiagnosis still happens especially in case of severe diseases. A way has to evolved that can give patient specific treatment.

According to McKinsey, operating efficiencies attainable from scaling the impact of advanced analytics auch as AI and ML in pharmaceutical industry range as high as 15 to 30 percent of EBITDA over five years, accelerating to 45 to 70 percent over a decade gave the potential impact of predictive modelling in discovering and optimisng new blockbuster therapies.

This has huge potential and will also be a driving force for upcoming healthcare application using AI and Data Science.

Deciphering the Problem

Data Availability & Integration

Lack of accurate and reliable data is the foremost hindrance in our case study. In the pharmaceutical industry, data comes from various sources such as electronic health records, clinical trials, and Research studies. The data coming from these sources is stored in different formats (structured/unstructured) and systems and cannot be used for analysis in raw form. Also, ensuring that the data we get is accurate in a timely manner is quite difficult as of now.

Inability in developing the right treatment for patients :

Due to lack of accurate and sufficient patient records, we currently lack an efficient system for diagnosing diseases quickly and developing the appropriate treatments. Various geographical regions experience distinct health issues and diseases. Due to unavailability of any unified platform to store the medical data of patients, even if we get data from some sources, it cannot be guaranteed if it's accurate and updated.

Drug Trials on Animals and Human Subjects

We are aware that while developing any new drug it has to go through various trials in which animals as well as humana are subjected to trials which are fatal. If the effects of a particular drug can be simulated and tested in an artificially generated environment using the previous data and records available to us could save lives of trial subjects.

Privacy and Regulatory Compliance Data Security

The pharmaceutical industry often contains sensitive patient information. Ensuring compliance with data protection regulations like HIPAA or GDPR, while still making data accessible for analysis, is a big challenge. Maintaining data security and protecting against breaches and cyber threats is critical. The use of data in pharmaceuticals raises ethical concerns, especially when it comes to patient privacy and consent.

Proposed Solutions

Natural Language Processing

NLP can extract insights from unstructured clinical notes, research papers, and electronic health records, enabling a deeper understanding of patient conditions, disease progression, and treatment effectiveness. This data can be further modeled so that we can use it for predictive analysis.

Drug Adverse Event Detection

AI can monitor adverse events associated with medications in real-time by analyzing patient reports and data. AI can predict potential drug-drug interactions and adverse effects to improve medication safety that will alert us of the harm that could have been caused. AI can identify suitable patient cohorts and optimize trial protocols to reduce costs and accelerate the development of new therapies.

Cost Effectiveness

AI can enhance drug manufacturing and supply chain management by predicting demand, optimizing logistics, and reducing production costs. One of the major goals of any business is to save resources and costs wherever possible.

Patient Stratification

AI can segment patient populations into subgroups based on genomic, clinical, or demographic factors, allowing for more personalized treatments and clinical trial design. In COVID, we saw that different patients give different responses to the same medicine, therefore, the patient stratification will give us an upper hand in strategic planning. In addition, AI can analyze real-world patient data, such as electronic health records, to assess treatment effectiveness and track disease progression outside of controlled clinical settings.

Drug Discovery and Development

AI models can make drug discovery faster by analyzing molecular data, predicting potential drug candidates, and simulating the effects of new compounds on biological systems. We will require all the past research work done and data of similar trials if any for discovering a drug related to that particular disease. AI can also analyze large-scale genomic data to identify genetic markers associated with disease susceptibility and treatment response, aiding in evolving precise treatment for patients.

Risk Benefit Analysis

In order to analyse the solution ideas and their prioritisation , I have done a Risk-Benefit analsyis for each solution that will give us a clear idea about how to proceed futher along with minimizing the cost and efforts and maximizing the ROIs (Return on Investment). e
I will futher elaborate on the technicalities of the top three solutions in next slide.

Natural Language Processing	The input data should be reliable or else the results will not be promising	• Research papers, prescription notes and clinical trial data all can be processed to generate useful information • Physical capital is not required	#1
Drug Adverse Event Detection	We are analysing real time patient reports but due to the difference in body types and genomic composition, results can't be used universally, subjective analsyis will still be required.	• Life of humans and animals will be saved • Since, the effects will be analysed on the basis of simulations, waste products will not be generated as well as real drugs won't be wasted, thus monetory benefit too	#3
Cost Reduction	If too much focus is paid on reducing costs on the sole basis of AI predictive and planning models, it can lead to bad consequences	• Resources and capital will be saved. Overall advantageous for the pharmaceutical companies	#4
Patient Stratification	It is extremely difficult to gather varied data that can cover majority of the world population, so our results would only be useful for some segment	• Distribution of drugs in enough amount will ensure its availability to the patients across the world. • Patient specialised treatment procedures could be evolved.	#5
Drug Discovery & Development	While using AI models, we have a margin of error , so if due to any inaccuracy the results deviate it can cause huge loss	• A lot of time will be saved since all previous research insights are readily available • Process will become faster and easier	#2

Feasibility & Technicalities of Solutions

Gathering and Processing Data

- First, we need to work on our primary aim that is collection of acurate data from different sources and systems and consolidating it at one place that is secure enough and is only accessible to designated authorities and individuals.
- A cloud based application can be designed that will contain all the gathered data along with tools for AI/ML modelling. Only registered individuals and authortites will be able to access and use the data. The users will be assigned a role based on which they will be granted access to particular data and functionality.
- The collection of data will be done with the help of NLP. Using Large Language Models, we can find relations between all sort of unstructured data and create a knowledge graph for the same which is easier and more effective to work and this processed data could be futher used with various other AI models.
- **NLP** has been explored a lot and implementing it won't take excessive time. The tricky part is collection of data. Summarising it, I would say its is feasible as well will give high ROI and investment will be less in comparison to the benefit we will get.
- The inaccuracies in the data can be seen in form of outliers after processing and can be tackled accordingly.

Drug Discovery and Development

- **Reinforcement learning** can be used for drug designing and molecular modelling processed through computational simulations and then suggesting chemical modifications in the drug to improve it's properties.
- **Generative Adversarial Networks (GANs)** can be used to generate molecular structures with desired properties that can help in designing new potential drugs.
- For image based drug discovery, **Convolution Neural Networks (CNNs)** can be used . They analyse cells and tissues and identify the target for drugs and also analyse the structure of tissue and cells to predict the properties required in the drug.

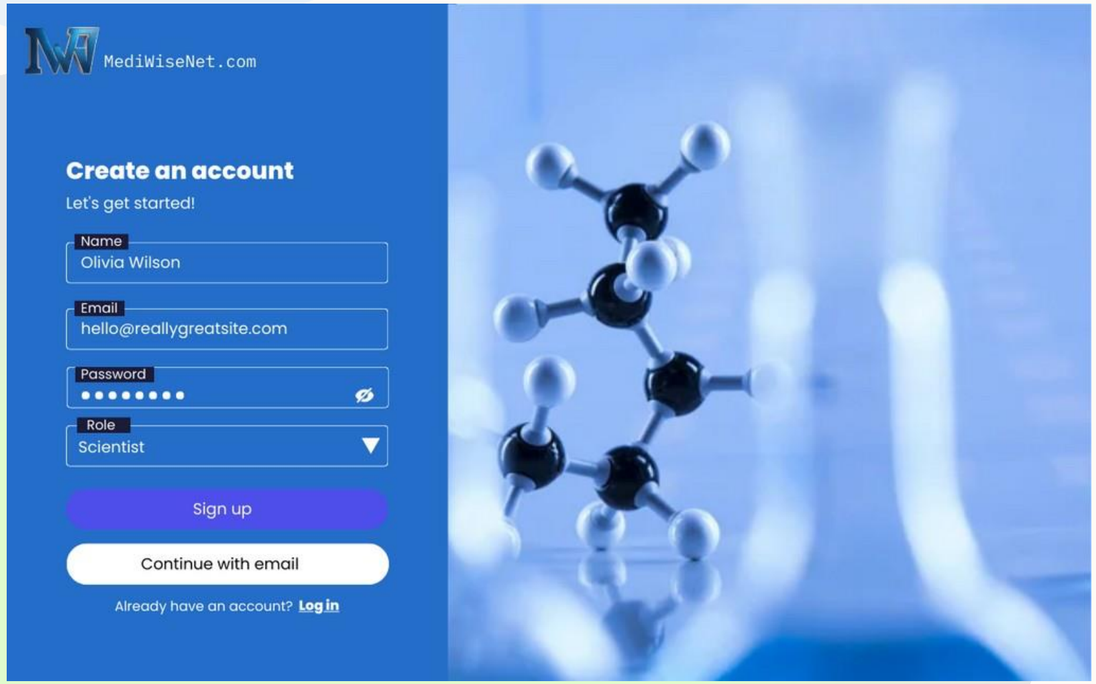
Drug Adverse Event Detection

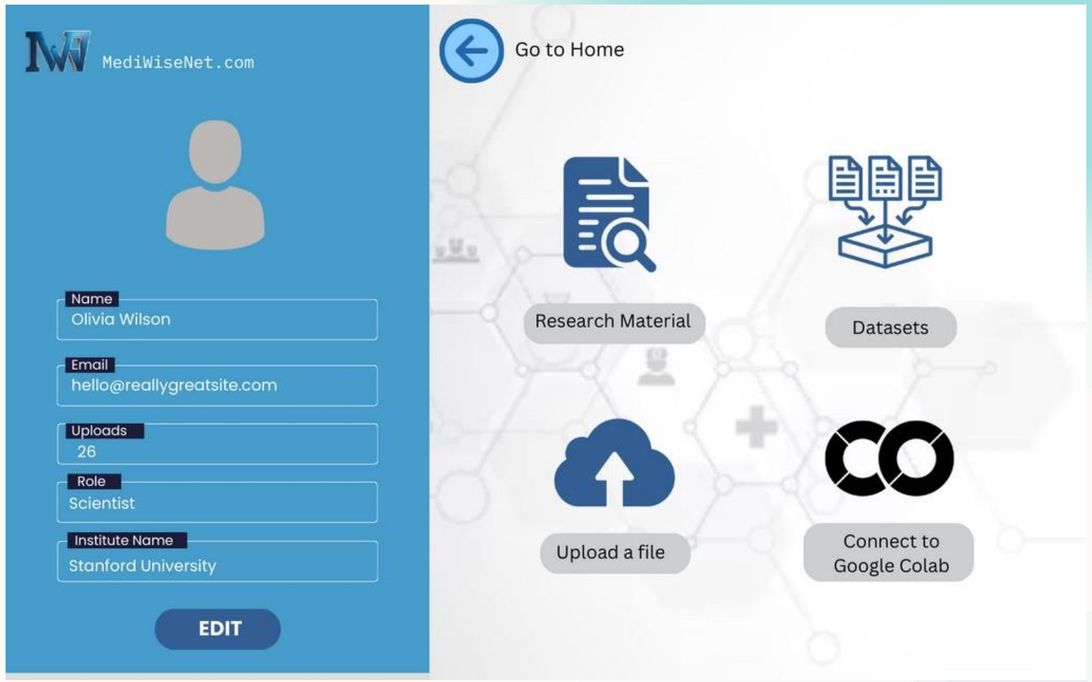
- **Physiologically-Based Pharmacokinetic (PBPK) Modeling:** These models simulate how drugs are absorbed, distributed, metabolized, and excreted in the body. They can predict potential interactions by simulating drug concentrations and metabolism rates in different tissues. This data can predict the effects that drugs generate in the body.
- **Structure-Based Screening :** There are **molecular docking algorithms** that can predict drug-drug interactions by analyzing the 3D structure of the drugs and their interaction with the target proteins in the body.

WIREFRAMES

I am presenting the mockup of screens that will be present in the application that I discussed previously that I have named as **“MediWiseNet.com”**. We will have different roles and there will be a role of admin that will be strictly available to selected people. Only they will be able to make important changes.

Following are the roles that will be available while signing up – Scientist, Researcher, Doctor, Associate at Pharmaceutical company, Hospital Administrator (Provide data from hospitals) , Admin (One has to specify the organization he/she is working for)





This is the Sign up page, from where one can login/signup. Account will be created only after successful authentication and verification from the admin.

One needs to provide appropriate proofs such as license for the company or institution with whom he is working.

Based on your role you will be granted access to the functionalities.

Given above is the page for the role of a Scientist. Similarly we will have pages for other profiles too

It has the options – Access to Research material, Data, Connect to Google Colab(or similar environment), Upload Material (Can be of different types – Code, Research paper, Clinical trails, Datatset)

Here is the link to another health application whose UI was designed by me -

Pitfalls & Future Enhancements

With the use of Blockchain, we can create a centralized data system which is extremely secure where one central authority can handle the entire system. Data sharing is much more secure with the use of blockchain, so it is also an feasible solution that can be implemented in near future.

Telemedicine has gained popularity in last few years, and therefore it can become a more plausible source of data. One of our issue in patient stratification is that we lack data from remote areas. If somehow, this data is gathered via help of telemedicine, stratification could be implemented globally then.

One of the major issues that we face while solving the problem is how to keep the data secure. The data used in the pharma industry is quite elaborate, from top level research to the genomic data of patients, everything is important and needs to be kept safe from hideous people who can use it for malicious purposes. Designing a secured system is still a challenge since the product we are trying to formulate will be accessed by various authorities.

When we use AI algorithms to make analysis, we need to keep in mind that the solutions we design should have high accuracy because even a small margin of error can give horrendous results that might lead to severe consequences.

Conclusion

We can implement some really effective solutions using Artificial Intelligence that will help in data driven decision making, and analysis of data that will help in solving many of the challenges faced by the pharmaceutical industry as of now. Our primary concern is to gather accurate data that can be processed further to get useful results for the research and analysis. According to my analysis, a major issue is the security and ethical concerns regarding the matter that should be worked upon in the future. As, I mentioned Blockchain as one of the potential solution, in future there might be some superior techniques that will make the system more secure.