

# **Technical Review: Startup companies in Medical Industry and AI technology used**

Manasa Gangaiah ([manasag3@illinois.edu](mailto:manasag3@illinois.edu))

## **Introduction:**

In over a decade or so, AI technology in the medical industry has brought significant advancement in tackling some of the issues with patient experience, disease prevention, drug discovery, and more. Some of the startup companies like Babylon, Zymergen, Olive, Insitro, Exscientia, Owkin are working on those issues. In this review paper, will look at the background and machine learning technology used in drug discovery startups like Insitro.

## **Problem:**

Drug discovery is a very time-consuming process it takes around a decade or so and costs nearly 1 billion dollars. Drug discovery process involves multiple trails over a long period of time and finally narrowing down the best drug for a disease or infection. Usually for testing the drugs, at the initial stages pharmaceutical industries use animals like mice. These animals do not even get the diseases/infections that humans get, so they must inflict animals as those diseases or infections to try if the drug they are researching works. Humans cannot be used for drug testing at the initial stage, and when they provide the medicines to people as a trial medication, they should have done significant testing on animals. Since they have tested on animals does not mean that it will work on humans as the biology of human is so different from animals and they might react differently for the drug. So, each drug approved takes a long time and at a huge cost. Patients need to wait for availability of the medications for so long and pay high cost for their treatments.

## **Data Collection:**

Machine learning techniques are used to help drug discovery and reduce the time taken along with the reduce cost of drug discovery. To create models for drug discovery we need huge human biological data. Each human being is defined by their genetic data (genome), and they are unique. Different genomes define different phenotypes. Understanding genetic data and mapping them to each human clinical traits is vital as this information would be used for modelling. In recent years, huge data has been created/collected like in UK bio bank. There are 100 of millions of genome sequences are obtained and measured phenotypes from the participating people. These needs to be tied to clinical traits. Another data that are collected are in genome wide association studies where in disease and healthy traits across various diseases are collected and causal basis for these diseases are also collected. These data are high content biological data as these diseases will have 1000's of genes contributing to a disease and to identify which gene causes what effect is a difficult task.

In Insitro: Instead of using animals for testing, human skin cells are used to create relevant biological systems from patients and these cells are reverted to pluripotent stem cell status. From stem cells generated they can create any biological systems like neurons, cardiomyocytes to create disease cells. Perturb those cells using genome engineering techniques such as CRISPR to create even more disease-causing genes/phenotypes cells and measure them with healthy cells.

### **Machine Learning Techniques:**

Microscopic images of healthy and disease cells data obtained are in terabytes and is high dimensional data. We need high performing machines to process and model. AWS works with Insitro to provide them high performing machines. Computer vision techniques will be used to extract meaningful information from the images (as these images are very complex) for identifying the disease and healthy cells/genes from the images by classifying those images. The cellular phenotype data is high dimension manifold, this will be converted to low dimension to identify similarity semantics so that cells/genes can be classified as healthy or disease causing and finally it will be mapped to clinical traits. Now new drug in research will be used to improve or cure the disease(cell) and capture those images after drug is introduced. CNN will be used on these data images as well to see if they classified to healthy cells. If so, the drug is helping if not we can move on to research more drugs. This data will be fed back to bio data and dataset will keep improving. Again, the updated data will be fed to machine learning to improve modelling. This is continuing process.

### **Conclusion:**

Since, a drug showed an improvement on diseased human cell (or organ grown in lab), it has a higher chance of success in diseased humans. Turn around time for researching these drugs are significantly reduced, as the number of trials is reduced. Cost of the drug discovery is also reduced. Innocent animals need not be used for drug discovery as well. With high advancement in machine learning, companies like Insitro, Exscientia, Owkin are able improve better drug discovery processes.

### **References:**

<https://www.insitro.com/>

<https://www.medicalstartups.org/top/ai/>