BioBytes

NEWSLETTER

Introduction to CB

and more about BioBytes.

Aug 2019

001

Throughout the course of humanity, we humans have been constantly evolving our machines and tools to make our lives easier. Having gained total dominance over the silicon empire, in the past 30 years alone, we have managed to achieve innumerable feats with the advancements in computing technologies. We have looked to nature for inspiration and in recent years there has been a surge in the development of tools to understand the natural phenomena and to simulate them. Understanding the deep and intricate natural design has become relatively easier with the computers simulating billions of natural processes in seconds to produce an outcome which, for us, would have taken years to produce. Computational algorithms have been coded which encompass most of the biological world. This field of study, as an umbrella term, is referred to as computational biology. More formally -

Computational Biology is the development of models and algorithms to understand biological systems. Experimental Biology produces a vast amount of data and so arose a need to process this data using computational tools and thus computational biology is becoming an essential field because of the possibilities it gives rise to in fields like healthcare and drug discovery. Contrary to popular belief, one does not come across names like “Myrmekiaphila neilyoungi” on a daily basis and nor does one have to cut open a frog to find the cause of a disease, instead data analysis and modelling are a computational biologist’s subsistence.

We at BioBytes aim to showcase the research being undertaken in the field of computational biology and help students regardless of their knowledge of biology take part in some of them. On top of that, we would be interacting with prominent personalities in this field, organizing sessions on the various techniques being used and have long challenges which would require you to solve real-world problems by developing state of the art algorithms. We will also initiate team projects soon exclusive to the members of BioBytes where we would be mentored by professors of the CB department to be a part of their research projects. These projects would potentially be published in research papers or could even be further incubated into startups.

BioBytes

Research

1. Microsoft Research

At Microsoft's research labs around the world, computer scientists, programmers, engineers and other experts are trying to crack some of the computer industry's toughest problems. A subset of those scientists, engineers and programmers are trying to use computer science to solve one of the most complex and deadly challenges humans face: Cancer. Although the individual projects vary widely, Microsoft's philosophy toward solving cancer focuses on two basic approaches

One approach is rooted in the idea that cancer and other biological processes are information processing systems. Using that approach the tools that are used to model and reason about computational processes – such as programming languages, compilers and model checkers – are used to model and reason about biological processes.

The other approach is more data-driven. It's based on the idea that researchers can apply techniques such as machine learning to the plethora of biological data that has suddenly become available, and use those sophisticated analysis tools to better understand and treat cancer.

"The collaboration between biologists and computer scientists is actually the key to making this work," said Jeannette M. Wing.

"We've reached the point where we are drowning in information. We can measure so much, and because we can, we do," said Jasmin Fisher, who is a senior researcher. "How do you take that information and turn that into knowledge? That's a different story. There's a huge leap here between information and data, and knowledge and understanding."

"We can use methods that we've developed for programming computers to program biology, and then unlock even more applications and even better treatments," said Andrew Phillips.

Of course, none of these tools will help fight cancer and save lives unless they are accessible and understandable to biologists, oncologists and other cancer researchers. And researchers at Microsoft are taking great pains to make these tools accessible and easy to understand.

One approach Fisher and her team are taking is called Bio Model Analyzer, or BMA for short. It's a

Microsoft Research

3. Verily

Owned by Alphabet, Verily is a life sciences company previously known as Google Life Sciences. They develop tools to analyse health data to help in timely decision-making and effective interventions. By researching ways to predict and prevent disease onset and progression, they aim to transform the way healthcare is delivered.

Onduo is a virtual diabetes program by Verily that provides the tools, coaching and access to specialty doctors that you need to take control of your diabetes and learn what works for you. Onduo collects data from members directly, from members' wearable and other connected devices, and from the healthcare system. The data is used to understand your health and health risk factors, track progress, measure outcomes, and deliver meaningful insights.

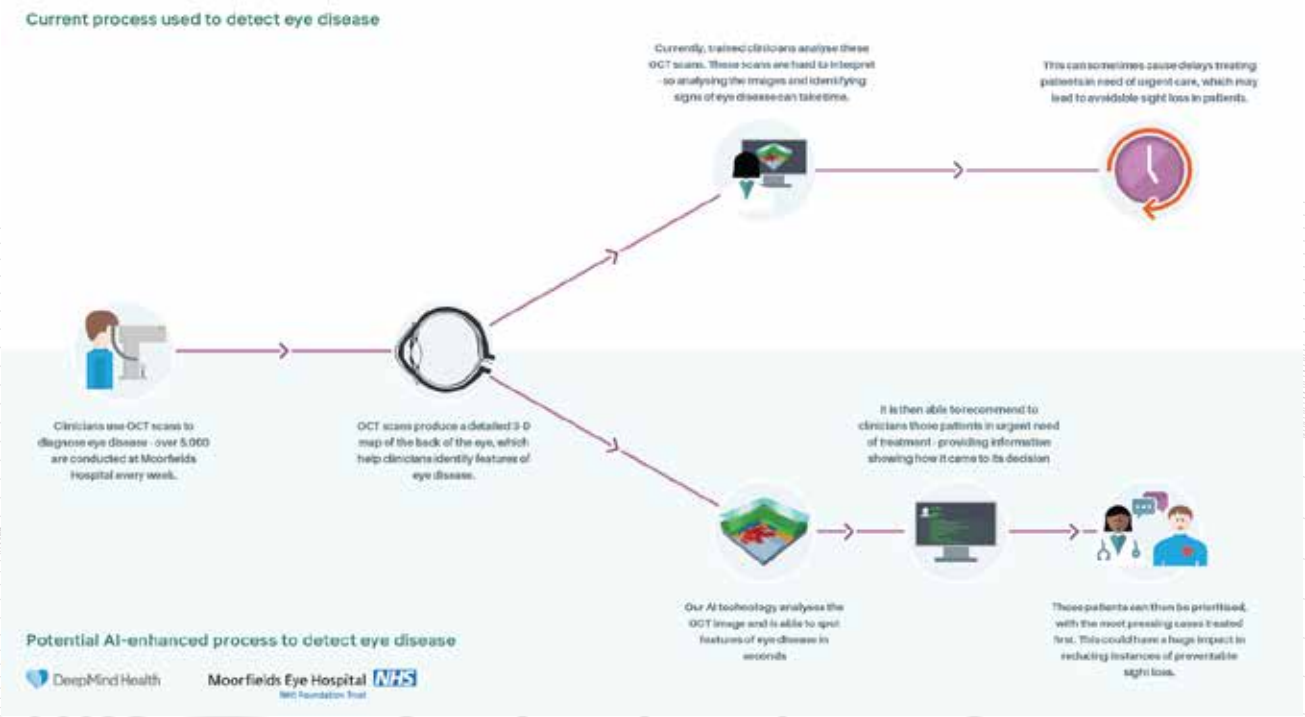
verily

4. Deepmind

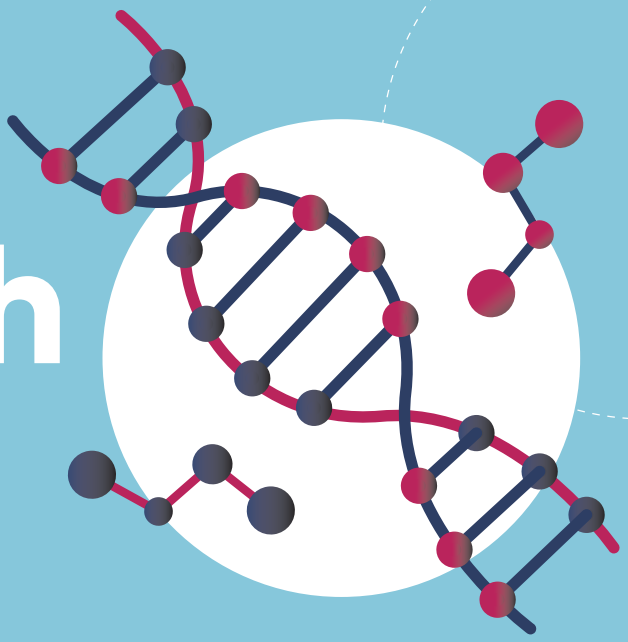
Owned by Alphabet, DeepMind seeks to solve real-world problems through the use of AI. One such field where DeepMind has managed to achieve excellence in is the healthcare domain.

DeepMind

In many clinical specialities, there is a relative shortage of this expertise to provide timely diagnosis and referral. For example, eye care professionals use optical coherence tomography (OCT) scans to help diagnose eye conditions. These 3D images provide a detailed map of the back of the eye, but they are hard to read and need expert analysis to interpret. And the worst part is that for the initial assessment of many of the sight-threatening diseases, OCT scans are indispensable. Artificial intelligence (AI) provides a promising solution for such medical image interpretation and triage. The system that has been developed seeks to address this shortage of skilled personnel. Not only can it automatically detect the features of eye diseases in seconds, but it can also prioritise patients most in need of urgent care by recommending whether they should be referred for treatment. This instant triaging process should drastically cut down the time elapsed between the scan and treatment, helping sufferers of diabetic eye disease and age-related macular degeneration avoid sight loss.The early results show that the system could handle a wide variety of patients found in routine clinical practice. In the long term, we hope this will help doctors quickly prioritise patients who need urgent treatment – which could ultimately save sight



Research Paper



Continued progress in technology over the past years has made transcriptome analysis of individual cells a reality. (The study of all the RNA molecules within a cell is known as transcriptome analysis. Many studies of the transcriptome focus on messenger (m)RNA molecules only, which reflect the genes that are being actively expressed (as protein products) in a cell or tissue at a given time or in a given situation).

The recent introduction of droplet-based transcriptomics platforms has enabled the parallel screening of thousands of cells. Large-scale single-cell transcriptomics is advantageous as it promises the discovery of a number of rare cell subpopulations. Existing algorithms (like RaceID and GiniClust) to find rare cells are unbearably slowly (or even stop working) as the sample size grows to the order of tens of thousands. Both RaceID and GiniClust use clustering in some form, as an intermediate step for detecting rare cells.

Single cell RNA-sequencing (scRNA-seq) screens gene expression at an individual level, as opposed to measuring their population-level average expression-signature. The data we get from a tissue is scRNA-seq data. The comprehensive characterization of all major and minor cell types in a complex tissue requires processing several thousand single cells, this is because a large number of cell type-specific transcripts are not detected in the sequencing. As a result, a small number of cell type-specific genes often fail to influence the downstream analysis sufficiently. However, the cost of profiling thousands of cells has been significantly reduced by using droplet based techniques.

The advent of single-cell transcriptomics has made rare cell discovery a mainstream component in the downstream analysis pipeline. Rare cells represent minor cell types in an organism. When the number of profiled cells are in the hundreds, even an outlier cell (singleton) deserves attention. Examples of rare

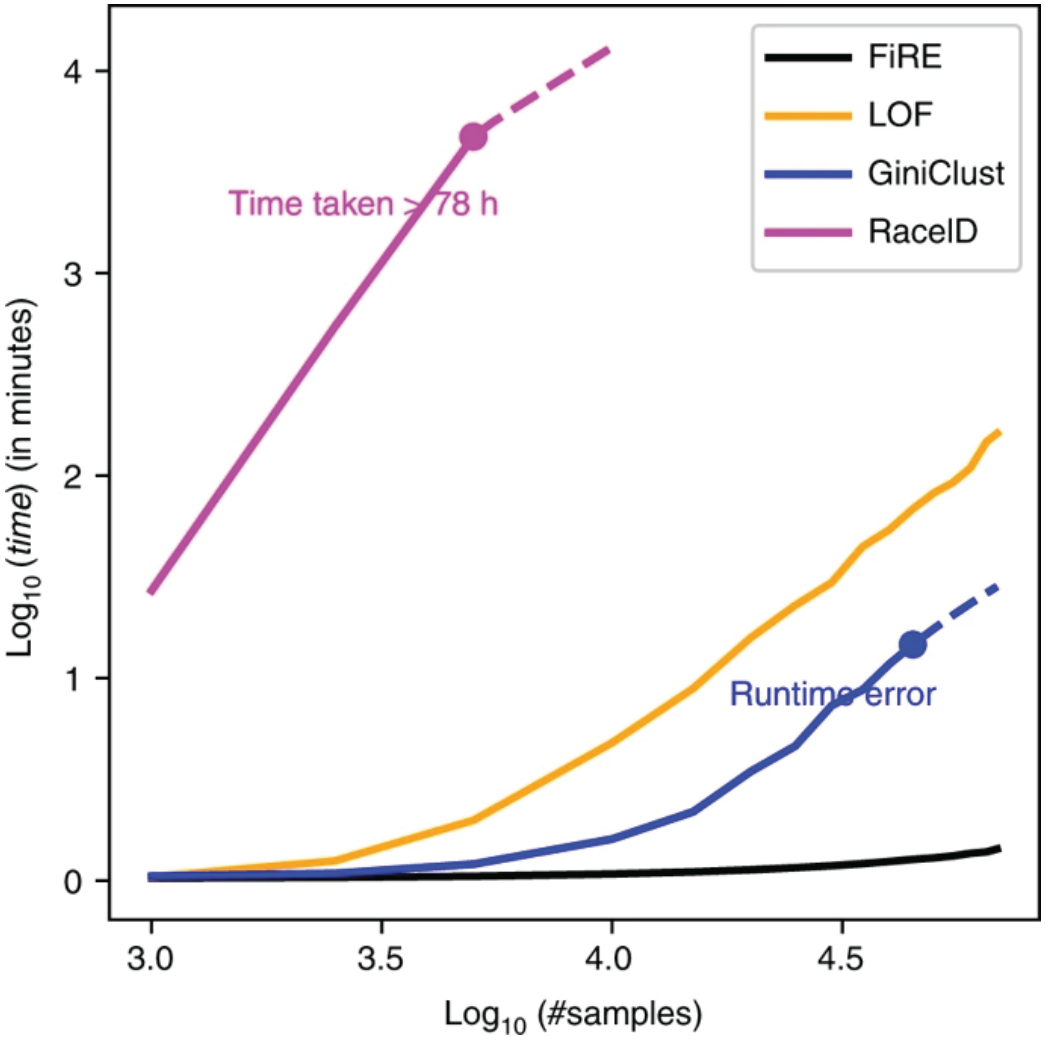


Image source
https://bit.ly/2HoZwQB

cell types include circulating tumor cells, cancer stem cells, circulating endothelial cells, endothelial progenitor cells, antigen-specific T cells, invariant natural killer T cells, etc. Despite low abundance, rare cell populations play an important role in determining the pathogenesis of cancer, mediating immune responses, angiogenesis in cancer and other diseases, etc.

Researchers from IIITD under Dr. Debarka Sengupta designed Finder of Rare Entities (FiRE), a conspicuously fast algorithm to estimate the density around each subjected multidimensional data point.
Design of FiRE is inspired by the observation that rareness estimation of a particular data point is the flip side of measuring the density around it.
One way to imagine it is that it basically creates buckets, the cell originating from a large cluster shares its bucket with many other cells, whereas a rare cell shares its bucket with only a few

FiRE assigns a continuous score to each cell, such that outlier cells and cells originating from the minor cell populations are assigned higher values in comparison to cells representing major subpopulations. A continuous score gives users the freedom to decide the degree of the rareness of the cells, to be further investigated.

FiRE attempts to fill that gap, with a number of pragmatic design considerations. Most notable among these is its ability to avoid clustering as an intermediate step. A typical clustering technique is not only time consuming, but also incapable of comprehensively charting the minor cell types in a complex tissue on a single go.

FiRE can recover artificially planted rare cells
FiRE is sensitive to cell type identity
FiRE is scalable and fast

FiRE is fast, accurate and cheap. For example, FiRE took ~31 s to analyze a scRNA-seq dataset containing ~68 k expression profiles (for reference this would have taken >78h for RaceID). Such unrivaled speed, combined with the ability to pinpoint the truly rare expression profiles, makes the algorithm future proof.

ON AIR WITH BioBytes

Dr. Debarka Sengupta

Dr. Debarka Sengupta is an Assistant Professor of the Departments of Computational Biology and Computer Science at IIIT-Delhi. Debarka did his doctoral research at the Machine Intelligence Unit of Indian Statistical Institute. After graduation in 2013, he pursued his postdoctoral research at the Genome Institute of Singapore where he got exposed to the then-emerging field of single-cell genomics. His group pioneered single-cell research in India and published several breakthrough findings, including the discovery of a rare subtype of pars tuberalis lineage in mouse brain (Jindal et al., Nature Communications, 2018). His current research focuses on early cancer detection using liquid biopsy techniques and functional interpretation of human variants. Debarka leads the data science research at Circle of Life Healthcare Pvt. Ltd., a health analytics company. He is a recipient of the prestigious INSPIRE Faculty Award. He serves on the editorial boards of PLOS One and Scientific Reports (A Nature group journal).

Get the full interview on :
youtube.in/biobytes
Support us by subscribing to our channel

A production of :

BioBytes

The Computational Biology Club of IIIT Delhi

INDRAPRASTHA INSTITUTE *of* INFORMATION TECHNOLOGY DELHI