IIT-B Hackathon 2020

Fighting the Global Wuhan Virus Crisis

AI/ML and Data Science as a Weapon

Healthcare
Data Science

Can you build an original, comprehensive solution to help handle the crisis?
We are looking to predict, visualize and act to contain the 2019 n-Corona Virus,
through the power of data science.

CitiusTech

10,000 Infected Globally in 2 weeks!

The news emerging from Wuhan, a city in central China, about the outbreak of n-Corona Virus (2019 nCOV) is causing immense panic and anxiety across the world. Infections have spread across 20 cities in China and to countries like Taiwan, Thailand, Malaysia, Japan, South Korea, USA and even India, within 2 weeks. The Chinese government has since quarantined the Wuhan city of 11 million people, banned travel from four more cities and dispatched more than 30000 medical professionals to battle the epidemic.

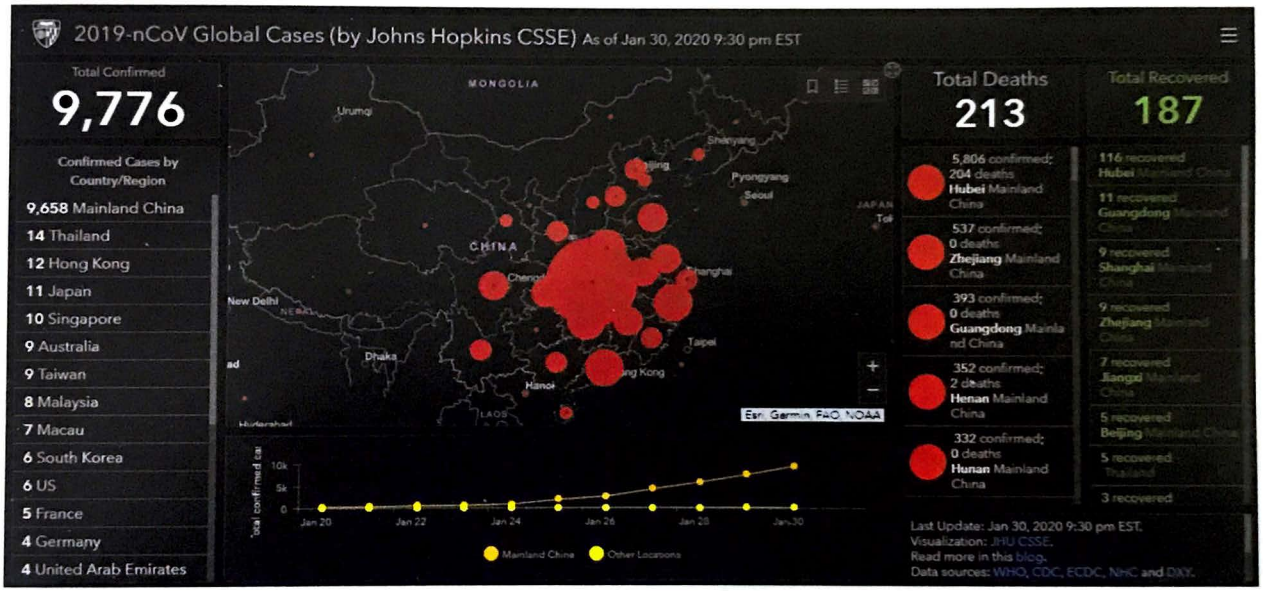


Figure-1: Total cases and deaths from recent Wuhan virus outbreak
(The latest case count as per real-time global map. (Johns Hopkins))

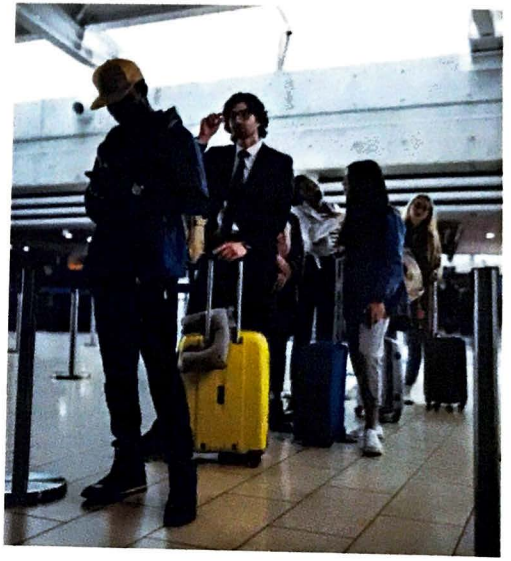
Can Data Science Give Us the Edge?

The last two decades have witnessed multiple instances of spread of pathogens that have caused international health security concerns. These include the SARS (Severe Acute Respiratory Syndrome) and MERS-CoV (Middle East Respiratory Syndrome Coronavirus), influenza, Nipah, henipaviral diseases, and Lassa fever.

The risk of new and unheralded infectious diseases and epidemics continue to pose enormous challenges to populations, healthcare systems and governments. With today's globalized economy involving extensive international travel and trade, the risk of a global pandemic has increased exponentially, with largescale socioeconomic consequences.

Measures like quarantines, vaccinations, food and travel advisories are effective only when administration can act early with the spatio-temporal information of transmission risk.

Mathematical modelling and machine learning techniques that leverage disparate data sources such as health data, census, social media feeds, mobility information in the backdrop of outbreak reports can be used to characterize and predict the spatio-temporal rates of pathogen transmission.



International travel is a primary cause of a global epidemic

Such analyses can help policymakers to evaluate the threat to public health, determine the resources required such as healthcare workforce, quantity of vaccines, hospital beds to reduce disease burden, and guide disease surveillance efforts and the deployment of interventions.

The Problem Statement

With the objective of understanding and minimizing epidemic spread, we are asking you to develop a user-friendly tool that can effectively and accurately predict and visualize the risk of 2019 n-Corona Virus spread posed by outbreak events in real-time, by integrating disparate data sources.

Your tools needs to cater to one or more of the following stakeholders:

- Payers (insurance companies)
- Providers (hospitals, clinics)
- Health policy and intervention strategists

Prediction Coverage

- You need to predict expected new cases for China and USA, for two specific dates - 31 Jan 2020 and 1 Feb 2020
- You need to provide the percentage error in the prediction of new cases for China and USA using the test data that will be provided
- If you provide predicted cases for other countries as well, it may carry additional credit points.
- You need to provide visual summary of the outputs from your tool and the data that is used for modeling

Data Sets & Techniques

You are free to use any public data sets, tools and modeling techniques that you think would be relevant for the problem defined.

Test Data

Number of new nCoV cases recorded in China and USA on 31 Jan 2020 and 1 Feb 2020 will be used to evaluate accuracy of the created tool.

Daily status report from WHO website will be the source of our data

Timelines and Goals

In the **first 6 hours** you are required to make a 2-minute video to talk about who is your intended end user and what relevant information you plan to utilize and what insights you will bring in with the solution that you develop.

At the **end of 27 hours** you need to be ready with your deliverables:

- The complete solution / tool / product
- A 4-minute video about the solution / tool / product that you developed.

All artifacts need to be saved on the virtual machines provided to you. Details of the artifacts and how to use them will be provided separately.

Evaluation Criteria

The initial video (after 6 hours) will be used to evaluate concept and relevance. We may use these videos to share feedback with the teams.

Final entries will be evaluated on the following criteria, listed in Table-1

Criterion	Credits
Originality and Innovativeness (will happen in two stages)	10
Accuracy of prediction model on the test data provided	10
Story telling	10
Insights and recommendations	10
Relevance and utility for targeted stake holders	10

Table-1: Evaluation criteria for each team

Getting Started

How do respiratory diseases spread?

In any given location, the risk of disease spread from a local or distant source depends on several factors (see Figure 2):

1) the initial number of cases
2) the transmissibility of the pathogen,
3) the connectivity between areas,
4) the susceptibility of the populations at risk,
5) the quality of the local health systems.

Multiple free and openly accessible data sources are available that can provide information to quantify these drivers.

The tool we have asked you to build will integrate several of these resources (data streams) to quantify spatial heterogeneity in the risk of disease spread.

For a given country, the tool will predict the number of cases at any given time from the outbreak of the disease.

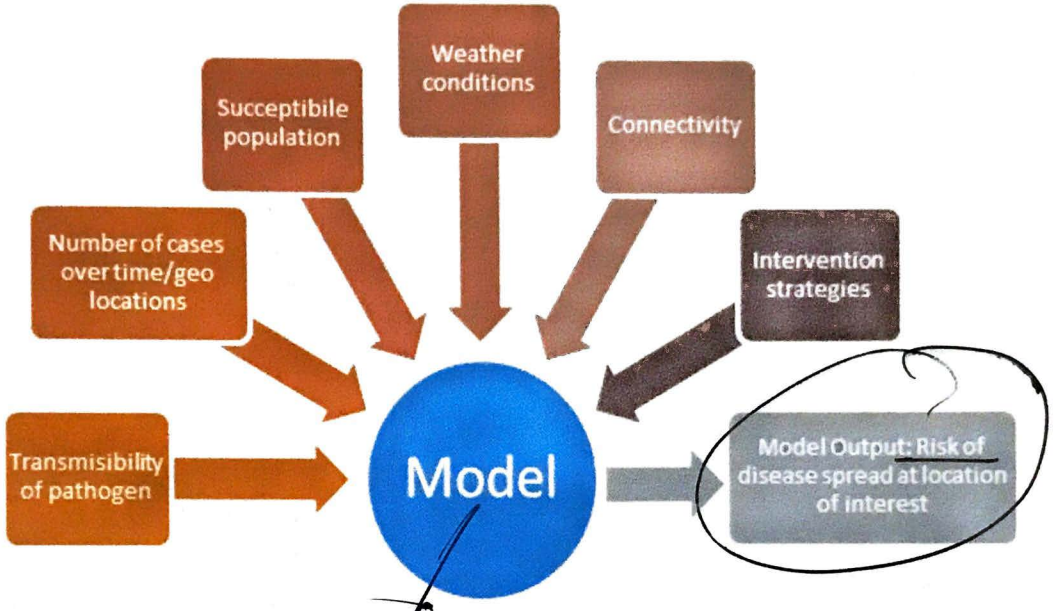


Figure 2: Schematic overview of the possible product including data streams

Suggested Videos

- Data driven forecasting
- Introduction to infectious disease model
- Long video: predicting future of infectious disease

Research Articles

- Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1)
- Transmission routes of respiratory viruses among humans
- Spatio-temporal hierarchical modeling of an infectious disease from (simulated) count data

Suggested Data Sources

- Outbreaks data: ProMED archives contain more than 70,000 records on outbreaks spanning 1994 to 2016 (www.promedmail.org).
- National Health Commission of the People's Republic of China (http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml)
- Demographic data: United Nation Population Division (UNPD, www.un.org) and from LandScan (Dobson et al. 2000).
- Flight data: EcoHealth Alliance's FLIRT (http://flirt.eha.io), a network analysis tool that enables detailed examination of flight networks
- Environmental factors: such as temperature or humidity, can be obtained from GSF model, NOAA
- Australian Infuenza Surveillance Report and Activity Updates, http://www.health.gov.au/internet/main/ publishing.nsf/Content/cda-surveil-ozfu-fucurr.htm
- Centers for Disease Control and Prevention. https://www.cdc.gov/
- R epidemic consortium: https://www.repidemicsconsortium.org/

About CitiusTech's Data Science Proficiency

CitiusTech's Data Science Proficiency is a multi-disciplinary team of data scientists, business analysts, statisticians, data architects and clinical informatics professionals. The proficiency offers a rich pool of Data Science, AI, Machine Learning and predictive analytics services and tools to help organizations improve care management practices, reduce operational and financial risks and make better decisions at the point of care.

Key features of the Data Science proficiency:

- Extensive knowledge of statistical mining, predictive modeling, deep learning, model life cycle management and Artificial Intelligence techniques
- Pre-built suite of analytics tools and components - NLP, chatbots, image analytics and Big Data analytics
- Expertise across a wide range of healthcare datasets - EHR, patient notes, claims, utilization data, socioeconomic data, consumer generated data, IoT, medical devices, imaging, lab data, etc.

Highlights

Data Science & Consulting

- Identifying and actioning opportunities around leveraging data science tools (e.g., care improvement, cost reduction, consumer engagement, healthcare data monetization)
- Expertise across AI and Machine Learning applications - NLP, chatbots, image analytics and Big Data analytics

Analytics Services

- Advanced analytics support, e.g., data profiling, model development, reporting and integrating analytics results with enterprise systems
- Expertise across disease modeling, statistical techniques and toolsets

Operationalizing Analytics

- Expertise across a wide range of healthcare datasets - EHR, claims, socioeconomic data, consumer generated data, IoT, medical devices, imaging, etc.
- Healthcare data engineers with rich expertise in healthcare data exchange standards and data formats