International Journal of Trend in Research and Development, Volume 9(5), ISSN: 2394-9333
www.ijtrd.com

Analysis on Data Science Technologies to Monitor Real Time Disease Outbreak, Forecasting and Spotting Real Time Trends for Governments, Health Organisations and Society

Vikas Kasina
Senior Data Scientist, Enterprise Data Science, Elevance Health, Hyderabad, Telangana, India

Abstract: The early diagnosis of public health hazards necessitates the monitoring of infectious diseases. A wide range of human and environmental factors influence the emergence of novel diseases. These include population density, travel, and trade. One of the most exciting developments in molecular diagnostics is the rising number of new technologies being developed. Risk assessment and outbreak detection can be improved through web-based surveillance systems and epidemic intelligence methodologies employed by all major public health institutes. Since its discovery on Chinese soil in December of this year, the COVID-19 virus has made its way to every continent but Antarctica. The World Health Organization has identified SARS-CoV-2-caused Coronavirus Disease 19 (COVID-19) as a global health emergency (WHO). If the rate of transmission is higher than that of SARS or the usual flu, then the fight against this disease must continue. To further understand this conflict, this essay focuses on the role of data science. A wide range of fields, such as epidemiology, drug discovery, and molecular design, benefit from the application of data science in conjunction with statistical analysis, computer science, and computational biology. Models based on data and mathematics have been developed for COVID-19, including correlations and forecasts. Data science methods are used to analyse huge COVID-19 epidemiological datasets in this study. COVID-19 confirmed cases can be better understood with the help of this solution. Our data science approach has been shown to be effective in gleaning meaningful information from the massive COVID-19 dataset.

I. INTRODUCTION

Among these include medical imaging, genetic research, the development of new drugs, customer service, and the field of predictive medicine. This includes medical imaging, genetics and genomics, drug development, customer service and predictive medicine, all of which benefit from today's data science technology. As a result of COVID-19, this has come to light. The use of data analytics to track disease outbreaks in real time, forecast, and spot patterns in real time has thus far benefited governments, health organisations, and society at large [1]. Wet lab investigations and imaging and sensor-based (from wearable sensors) devices can now be used to collect data (radiomics, for example). Data management, data visualisation, and statistical machine learning are some of the subfields of data science. Data may be organised, sorted, processed and analysed in real time using various methodologies in each area. COVID-19's response can currently be managed by chemical engineers utilising appropriate data analytic approaches.

Those in charge of preventing and combating disease receive timely access to data generated as part of routine public health monitoring (Thacker and Berkelman 1988). Public health monitoring is used to analyse the health and behaviour of populations served by the country's health, financial, and donor organisations. After an intervention has been done, it is important to keep an eye on the results to see if they have been successful. When leaders and managers have timely and relevant facts at their fingertips, they can make better decisions.

Increasingly, health and finance officials in developing countries and donor organisations are realising that data from good monitoring systems may be utilised to distribute resources and evaluate programmes. Surveillance plays a key role in safeguarding individual nations and the global population from epidemics like HIV and severe acute respiratory syndrome (SARS). As a result of the new USAID surveillance strategy that emphasises data use to promote public health interventions, countries like Brazil and Argentina were able to quickly develop their monitoring and response capability thanks to World Bank funding (USAID 2005). As noted in the WHO's recommendations for implementing the laws, WHO member nations are also required by the 2004 drafted amended International Health Regulations to have key personnel and core capacities in surveillance.

Medical and public health practitioners have benefited greatly from the widespread use of data science approaches [1–4]. Data science has become increasingly crucial during the ongoing coronavirus disease 2019 (COVID-19) pandemic because of these breakthroughs, the enthusiasm of academics and practitioners, and the acute demand for data-driven insights [5].

SARS-CoV-2 (SARS-CoV-2) is the virus that causes COVID-19, and it will have spread throughout the world by 2021, killing around 3.4 million people. Rapid and reliable data sources for individuals and the entire population have been emphasised as a vital requirement for disease surveillance and control in the wake of this epidemic [6]. Because it has drawn the attention of not only doctors and public health professionals, but also experts in other data and computational sciences fields that have been more peripheral in previous epidemics (like SARS, Ebola, HIV, and MERS), the COVID-19 pandemic has received a great deal of attention [7,8].

II. LITERATURE REVIEW

Using a model developed by the authors of Reference [9], it is possible to distinguish between COVID19 and four other viral chest infections. To gather data and keep tabs on the patient's

health, it makes use of a wide range of body sensors. These include sensors for measuring temperature and blood pressure as well as heart rate, breathing, and blood glucose levels. Patient symptoms can be identified by using artificial intelligence-enabled expert systems built into a cloud database, which can then be utilised to determine the best treatment method for those who are infected with or think they have COVID-19. However, it's not clear exactly what information about the patient's health the hospital's workers will be given. Additional research on mathematical models for COVID-19 detection and forecasting was assessed. It was found that detecting COVID-19 cases may be made more accurate by utilising AI, big data, and nature-inspired computing (NIC), according to the results from a recent survey.

Medical devices that can diagnose and track COVID-19 symptoms have been created by the authors of Reference [10]. The system makes use of headphones and a cell phone to identify respiratory problems. When the mobile app is used to acquire and save the audio data for further analysis, the MATLAB application is used to identify COVID-19-related respiratory symptoms. A report sums up the findings of the investigation.

Researchers [11] have also built a method to remotely monitor COVID19 patients who have been released. Each patient who registers up for the app receives a pulse oximeter and a thermometer so that they may monitor their symptoms, oxygen saturation, and temperature. An investigation by a team of nurses reveals that the patient's vital signs are abnormal. Based on the results of the evaluation, the patient may need to be readmitted to the Emergency Department once more (ED). When a patient is discharged from the hospital, the programme reduces emergency department visits and provides scalable remote monitoring capabilities.

Pre-symptomatic identification of COVID-19 can be achieved using smartwatches, according to a recent study [12]. Infected COVID-19 patients' physiological and activity data was analysed using smartwatches. According to the researchers, a two-tiered warning system may have caught more than half of the instances of COVID-19 before symptoms appeared in the patients. In addition, they found that early diagnosis of respiratory infections can be aided by the use of wearable devices that assess mobility and health.

In a study cited in reference [13], researchers looked at the symptoms of healthcare workers who had tested positive for COVID-19 (HCWs). This was followed by a COVID-19 PCR test to determine which symptoms were connected with each instance. As compared to those who tested negative for COVID-19, those who tested positive were more likely to suffer from fever, aches, and nausea, whereas those who tested negative were more likely to suffer from nasal congestion and a sore throat.

Investigators in the New York metropolitan region set out to examine the clinical characteristics and outcomes of 5700 COVID-19-infected patients [14]. The study, on the other hand, only included patients who were not severely sick, and the duration of follow-up was limited.

A website and Android app that can distinguish between a COVID-19 cough sound and other respiratory noises might be built using crowdsourced data from about 7000 unique users (more than 200 of whom reported a recent positive test for COVID-19). Data from cough sounds were classified using classifiers such as Logistic Regression, Gradient Boosting Trees (GBT), and Support Vector Machines (SVMs). They also classify users based on their health, such as whether or not they have asthma, smoke, or are otherwise in good health. Users are instructed to cough three or five times, and then asked to repeat the exercise every two days until their general health status has been updated. Using their method to screen for COVID-19, they were able to identify the cough from other lung disease coughs. More than 82% of patients with COVID-19 positivity were located using AUC (AUC). There needs to be a lot more field study done in order to better distinguish the COVID-19 cough sound vs the sounds of other respiratory processes.

III. MANAGING COVID – 19 PANDEMIC USING DATA SCIENCE AND TECHNOLOGY

To better understand how coronavirus genetics, origins, transmission, and incubation are related to both climate stability and change, researchers use data analysis. Figure 1 depicts the prospective applications of artificial intelligence that have been successfully achieved. Diagnostics, risk assessment, and patient monitoring have all benefited from the usage of data technology. Non-pharmaceutical remedies and vaccines and medicines can also be developed using data analysis. Several technology start-ups are partnering with clinicians, researchers, and government organisations to address the COVID-19 pandemic.

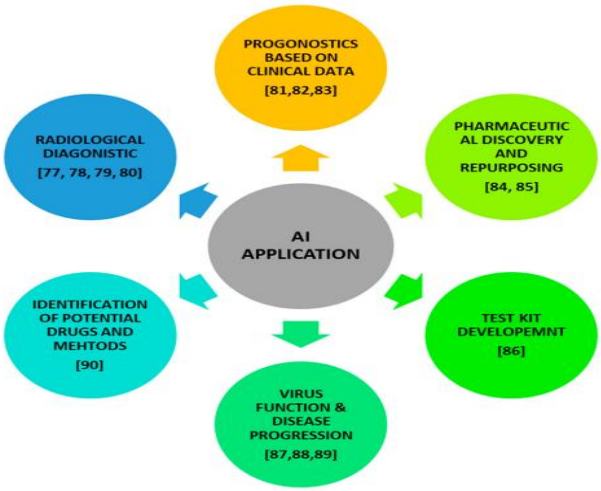

Figure 1. Combating COVID-19 via artificial intelligence. The corresponding reference numbers are shown in brackets with numerals.

Figure 2. Type and source of medical data

A start-up in Canada Blue dot artificial intelligence was one of the first to detect, track, and forecast an outbreak in Hubei province, as well as the first to predict the virus's first eight cities of infection. The COVID-19's whereabouts and activities can be tracked with the help of their real-time information. According to Alibaba, its AI-powered diagnostic technologies are 96% accurate in diagnosing coronavirus from CT scans of patients. When it comes to respiratory disorders such as coronavirus, Google Deep Mind AI is assisting in the identification of the structure of the protein involved. Benevolent AI is assisting in the identification of prospective therapeutic candidates. Terra drones are being used to transport medical supplies, while robots (such as those from Blue Ocean Robotics and Pudu Technology) are being utilised to clean and sterilise medical equipment, minimising human contact. The current pandemic crisis needs the effective management of the epidemic through the application of data technology and artificial intelligence.

A. Data Type and Source

There is a wealth of data available for use in the medical and health care sectors. Medical data can be classified into six categories based on the type of data and the source from which it is obtained, as seen in Figure 2. The analysis of this data will aid in the prediction of future occurrences, the comprehension of the existing situation, and the making of a number of important decisions. Medical data can be gathered from a variety of sources, including sensors included in wearable/mobile devices or medical devices, online questionnaires or mobile apps, hospital records, as well as data from open databases or social media websites.

B. Government Level

Big data analysis on social media can assist in identifying false information about diseases, alerting people to it, and preventing it from spreading. Using data on international air travel, researchers can keep tabs on how far the pandemic has spread and devise preventive measures before it is too late. Additionally, using big data science techniques such as deep learning, advanced machine learning techniques such as deep learning, mathematical and statistical models such as the autoregressive integrated moving average, optimization techniques such as particle swarm optimization, and simulation models such as SEIR, it is possible to accurately predict outbreaks such as COVID-19 (Susceptible, Exposed, Infected, and Recovered states). Authorities can forecast outbreaks, control epidemics, and assess the effects of interventions and control measures that have already been implemented or are now being planned using these models. This new data, when combined with what we already know about COVID-19, could be used to better characterise the outbreak's dynamic characteristics, allowing us to anticipate and prepare the healthcare system for any eventualities.

IV. CASE STUDY ON REAL-LIFE COVID-19 DATA

Data collection, integration, and pre-processing are all key components of successfully completing a project. We demonstrated the efficacy of our data science approach by utilising a range of epidemiological data sets, including those from Statistics Canada2. The Public Health Agency of Canada collated and consolidated data from provincial and territorial public health bodies to develop this dataset (PHAC).

COVID-19 cases in the Canadian population were recorded in the dataset as recently as November 12th, 2020. A specific episode week is provided in 190,108 occurrences. For the next several weeks, despite the fact that the first Canadian case occurred in Week 3, a relatively low number of new cases per day were reported. In (Episode) Week 8 (February 23-29), a total of 107 instances from Weeks 3-8 (February 23-29) were gathered in order to preserve the anonymity of these early cases while simultaneously accumulating statistically significant bulk. As of Week 9, the data is accurate due to the way each episode was recorded.

A. Security and Privacy

Authorities and patients alike are worried about the security and privacy of healthcare data, and medical data is only transferred in specified circumstances and for specific specialists/researchers and for defined reasons, according to the American Medical Association. This necessitates a clear definition of how medical data can be accessed without jeopardising patients' privacy or being used for undesirable reasons, especially in the face of severe conditions like COVID-19 and the development of hazardous epidemics.

B. Hospital Level

Remote patient monitoring data can be used to estimate the number of patients in a certain location in order to plan for any anticipated growth in the number of patients beyond the hospital's capacity. It is also becoming increasingly challenging to employ standard representation methods like tables because of the exponential growth of health data. Using artificial intelligence in conjunction with data analytics technologies can assist solve this problem, and the Savana system is an example.

Many diseases in general, as well as COVID-19 in particular, can be diagnosed and prognosticated using AI and ML algorithms. Patients can also be monitored from a distance

using a remote COVID-19 patients triaging device. Medical technologies that don't require invasive procedures and sensors that can be integrated into smart devices and watches make remote monitoring easier. AI and ML algorithms will use the data collected by sensors for diagnosis and prognosis. These applications will allow doctors to monitor COVID-19 patients with a moderate ailment remotely because of the high number of patients and the risk of infection.

V. RESULTS AND ANALYSIS

Table I supports the findings of the previous section. In addition to these findings, the study discovered that those between the ages of 20 and 40, as well as those over the age of 80, have higher COVID-19 rates than the national norm. When determining the percentage, one simply divides the number of people in a particular category (such as a given gender or age range) by the total number of people in that category. Males in their 20s, for example, make up 0.28 percent of this population group, which includes 19,049 cases. Seniors in their 80s and older are the most vulnerable, accounting for 1.31 percent of the COVID-19 population in their respective age groups (cf. the national norm of 0.53 percent of the national population).

At addition to looking at the total number of cases, our method also looks at how many of the 16 possible combinations of patients are currently in the hospital. Table II demonstrates that (a) the total number of patients admitted to the hospital increases as the patient's age increases. In conjunction with Table I, we can see that (b) hospitalizations continue to increase despite a decline in the number of cases between the 1920s and 1970s. As a result, the vast majority of young people infected with COVID-19 do not require hospitalisation. Individuals suffering from COVID-19 have a higher likelihood of requiring hospitalisation as they grow older. (c) Males in their 30s and older are more likely than women to be admitted to the ICU.

Table 1. On November 12, 2020, The Following Distribution Of Cumulative Covid-19 Cases (And Percentages With Respect To The Population Of The Corresponding Gender, Age Group -Combination) Was Determined

	Male		Female		Age Group
	#cases	*wrt corr. pop'n*	*#cases*	*wrt corr. pop'n*	*wrt corr. pop'n*
0-19	11,594	0.28%	11,374	0.29%	**0.28%**
20s	19,049	0.72%	19,316	0.78%	**0.75%**
30s	15,497	0.58%	16,151	0.62%	**0.60%**
40s	13,651	0.57%	15,851	0.65%	**0.61%**
50s	13,040	0.50%	14,935	0.57%	**0.54%**
60s	9,007	0.39%	8,584	0.36%	**0.37%**
70s	5,743	0.40%	5,790	0.37%	**0.38%**
80+	7,004	1.04%	14,755	1.49%	**1.31%**
Total	**94,585**	**0.50%**	**106,756**	**0.56%**	**0.53%**

Table Ii. As Of November 12, 2020, The Total Number Of Hospitalizations Is As Follows:

	Male		Female		Age Group
	ICU admitted	*Non-ICU hospitalized*	*ICU admitted*	*Non-ICU hospitalized*	*Total hospitalized*
0-19	11	79	11	95	**196**
20s	38	147	54	193	**432**
30s	74	288	62	292	**716**
40s	159	438	99	372	**1,068**
50s	421	777	211	557	**1,966**
60s	548	957	269	690	**2,464**
70s	489	1,150	268	1,042	**2,949**
80+	204	1,660	194	2,346	**4,404**
Total	**1,944**	**5,496**	**1,168**	**5,587**	**14,195**

Table 3: As Of November 12, 2020, The Percentage Of Hospitalization For Covid-19 Cases Of The Corresponding Gender, Age Group, Or Combination

	Male		Female		Age Group
	ICU admitted	*Non-ICU hospitalized*	*ICU admitted*	*Non-ICU hospitalized*	*Total hospitalized*
0-19	0.09%	0.68%	0.10%	0.84%	**0.85%**
20s	0.20%	0.77%	0.28%	1.00%	**1.13%**
30s	0.48%	1.86%	0.38%	1.81%	**2.26%**
40s	1.16%	3.21%	0.62%	2.35%	**3.62%**
50s	3.23%	5.96%	1.41%	3.73%	**7.03%**
60s	6.08%	10.63%	3.13%	8.04%	**14.01%**
70s	8.51%	20.02%	4.63%	18.00%	**25.57%**
80+	2.91%	23.70%	1.31%	15.90%	**20.24%**
Total	**2.06%**	**5.81%**	**1.09%**	**5.23%**	**7.05%**

Table III summarises the hospitalisation rates for COVID-19 patients by gender and age group. 38 male COVID-19 patients in their twenties constitute 0.77 percent of the 19,049 male COVID-19 patients in their twenties admitted to the ICU, as shown in Table II. As shown in Table III, hospitalisation rates for seniors 60 years and over range between 14.01 and 25.57 per cent (relative to the national mean of 7.05 per cent) and exceedsthem. Both ICU admission (8.51 percent of COVID-19 cases in men in their seventies) and hospitalisation (8.511 percent + 20.021 percent = 28.53 percent) are considerably higher in men in their seventies. A higher percentage of non-ICU hospitalizations is seen among (c) older guys (23.70 per cent).

CONCLUSION

The amount of data gathered on the global epidemic caused by COVID-19 grows exponentially over time. Big data analytics and artificial intelligence capabilities are required to rapidly comprehend and contain the pandemic. A taxonomy framework for COVID-19 applications was developed in this study, which categorises COVID-19 applications into four categories: diagnostic, estimation or prediction of risk score, healthcare decision-making and pharmaceutical use. In this work, numerous data analysis tools were introduced and their key aspects were explained. In addition, we shared our thoughts on a number of potential roadblocks to COVID-19's effective use of data analytics tools. It is difficult for researchers to share data with each other because of privacy and security concerns, as well as the unwillingness of patients to share some medical information with researchers.

References

[1] Topol EJ. 2019 High-performance medicine: the convergence of human and artificial intelligence. Nat. Med. 25, 44–56. (doi:10.1038/s41591-018-0300-7)

[2] Khoury MJ, Ioannidis JPA. 2014 Big data meets public health. Science 346, 1054–1055. (doi:10.1126/science.aaa2709)

[3] Wong ZS, Zhou J, Zhang Q. 2019 Artificial intelligence for infectious disease big data analytics. Infect., Dis. Health 24, 44–48. (doi:10.1016/j.idh.2018.10.002)

[4] Mooney SJ, Pejaver V. 2018 Big data in public health: terminology, machine learning, and privacy. Annu. Rev. Public Health 39, 95–112. (doi:10.1146/annurevpublhealth-040617-014208)

[5] Who coronavirus (covid-19) dashboard. https://covid19.who.int/. (Accessed 15 May 2021).

[6] Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. 2020 The architecture of Sars-Cov-2 transcriptome. Cell 181, 914–921.e10. (doi:10.1016/j.cell.2020.04.011)

[7] Luengo-Oroz M et al. 2020 Artificial intelligence cooperation to support the global response to Covid-19. Nat. Mach. Intell. 2, 295–297. (doi:10.1038/s42256-020-0184-3)

[8] Abdel-Basst, M.; Mohamed, R.; Elhoseny, M. A Model for the Effective COVID-19 Identification in Uncertainty environment using Primary Symptoms and CT Scans. Heath Inform. J. 2020, 1–18. [CrossRef]

[9] Stojanovic, R.; Skraba, A.; Lutovac, B. A Headset Like Wearable Device to Track COVID-19 Symptoms. In Proceedings of the 2020 9th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 8–11 July 2020; pp. 1–4.

[10] Gordon, W.J.; Henderson, D.; DeSharone, A.; Fisher, H.N.; Judge, J.; Levine, D.M.; MacLean, L.; Sousa, D.; Su, M.Y.; Boxer, R. Remote Patient Monitoring Program for Hospital Discharged COVID-19 Patients. Appl. Clin. Inform. 2020, 11, 792–801. [CrossRef]

[11] Mishra, T.; Wang, M.; Metwally, A.A.; Bogu, G.K.; Brooks, A.W.; Bahmani, A.; Alavi, A.; Celli, A.; Higgs, E.; Dagan-Rosenfeld, O.; et al. Pre-Symptomatic Detection of COVID-19 from Smartwatch Data. Nat. Biomed. Eng. 2020, 4, 1208–1220. [CrossRef]

[12] Lan, F.-Y.; Filler, R.; Mathew, S.; Buley, J.; Iliaki, E.; Bruno-Murtha, L.A.; Osgood, R.; Christophi, C.A.; Fernandez-Montero, A.; Kales, S.N. COVID-19 Symptoms Predictive of Healthcare Workers' SARS-CoV-2 PCR Results. PLoS ONE 2020, 15, e0235460. [CrossRef] [PubMed]

[13] Richardson, S.; Hirsch, J.S.; Narasimhan, M.; Crawford, J.M.; McGinn, T.; Davidson, K.W.; The Northwell COVID-19 Research Consortium. Presenting Characteristics, Comorbidities, and Outcomes among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020, 323, 2052–2059. [CrossRef]

[14] Brown, C.; Chauhan, J.; Grammenos, A.; Han, J.; Hasthanasombat, A.; Spathis, D.; Xia, T.; Cicuta, P.; Mascolo, C. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 2020. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data (ACM), New York, NY, USA, 25–27 August 2020; pp. 3474–3484.

[15] 8. Latif S et al. 2020 Leveraging data science to combat COVID-19: a comprehensive review. IEEE Trans. Artif. Intell. 1, 85–103