ISSN 2350-109X
www.indianscholar.co.in

Indian Scholar

An International Multidisciplinary Research e-Journal

MEMBERS OF INTERNATIONAL SCHOOLS' ASSOCIATION

# DATA SCIENCE: A WEAPON AGAINST PANDEMICS

**Anuva Khanna**
EuroSchoolAiroli
anuvakhanna07@gmail.com

**Abstract**

Out of all health related adversities, an epidemic is the most unpredictable one. An epidemic is a pathogen that affects people of only a certain country. While a pandemic is a pathogen that affects people across the globe, In 2020, COVID 19, a novel airborne virus, emerged as the 7th amongst the 6 previously known coronaviruses. This pathogen uses RNA (Ribonucleic acid) instead of DNA(Deoxyribonucleic acid). The virus contains spikes, thus making penetrating the cells much easier. This causes the contracted person's symptoms to go unnoticed, making it harder to track. In the wake of the COVID 19 surge, data scientists have been using various tools like Artificial Intelligence, Big Data, and Machine Learning to predict the rate of spread of the virus. This paper talks about the factors taken into consideration and the methods used to predict the speed and spread of the novel coronavirus.

**Keywords**: *Data Science in Epidemic Prediction, Big Data Analytics in Epidemiology, Virus case count prediction, Virus transmission rate algorithms, Machine learning, AI, Big Data in Epidemics, Prediction in Epidemics*

## INTRODUCTION

Over the decades, various viruses have affected micro level healthcare systems and macro level healthcare systems. Since the early Stone Age, the world has seen several such pathogens arising. Some of the most harmful pathogens were: the Bubonic Plague (1346-1353) [1], the Spanish Flu (1918-1920) [2] with an estimated case count of 500 million, Influenza Virus A (H1N1) (1918) [3], Influenza Virus (H2N2) (1918-1919) [4], Human Immunodeficiency Syndrome (1981) [5] and the currently ongoing COVID 19 [6]. With the technological advancements in the past century, healthcare systems have strengthened and dealing with epidemics has become more effective. Data Science, a field that uses mathematics and statistics to analyse current data and to predict future data, has played a major role in the prediction of COVID 19 cases. Under this broad category come several other emerging technologies. Machine learning is one such example. It is a type or form of Artificial Intelligence that allows machines to imply future scenarios from data. This research is of utmost importance as it will affect the actions of the future. This is because the case count predictions for a particular location will decide whether the place will be in containment for that period of time or will be 'unlocked' due to fewer cases. This is a strategy that the governments of several nations have been using, to ensure that minimal risk is taken with the case increase while also ensuring that economic activity in the nation does not halt completely. Not only does this affect government action, it also affects micro-level decision making as the prediction proves fruitful for people to understand the risk level to a certain age group, also affecting economic and entrepreneurial decisions made on the microscopic level. For these reasons, the predictive part of Data Science's contribution to the pandemic is the most vital part. This research is not just restricted to predicting the number of cases in a certain locality. It can extend to predictive analysis for recovery from the virus. Such research can include factors like predicting the number of vaccines a city or district will need. Recent research predicted through algorithms in data science (Sicilia, 2020, 9) has been using one aspect of public interaction through the Internet like Google and other social media to understand the severity of cases in that particular area via the tonal implication of the searches. Google Flu Trends, a source of study that is now discontinued, was a tracking platform to check how many times questions about a certain persisting disease were asked on Google. This is one of the many types of data used to predict the case count in different locations using prediction and transmission rate algorithms. This is just one example of how different data sets are being used for predictions, thus proving how useful Data Science can be in a health crisis like this.

**Body: An Overview**

As a topic of current discussion, lots of research and conclusions have been carried out. Each paper has taken into account different variables that could help in prediction of the case count. All these variables play instrumental roles in the course of the virus' rapid spread, also giving a rough idea of how the virus will travel and where it could potentially cause the most harm. Listed below and explained in detail are the factors considered while predicting the virus' movements, effective recovery from the virus and prediction of future viruses for the world to be better prepared for, also showing how Data Science was used to collect this information and implicate it. [7]

**Prediction Factor A**

Due to the unpredictable nature of an epidemic, it is impossible to predict where the spread will escalate the quickest. For this reason, data scientists consider the movement of the population on a daily basis [8]. Factors like the lifestyle of a population can give useful insight and can help understand where the virus might spread easily. The first and most commonly used variable to form data sets in Data Science is the understanding of where hotspots could potentially be formed. This is ideally the first approach in the wake of a pandemic as the data available for this set is relatively easily available.

ISSN 2350-109X
www.indianscholar.co.in
Indian Scholar
An International Multidisciplinary Research e-Journal

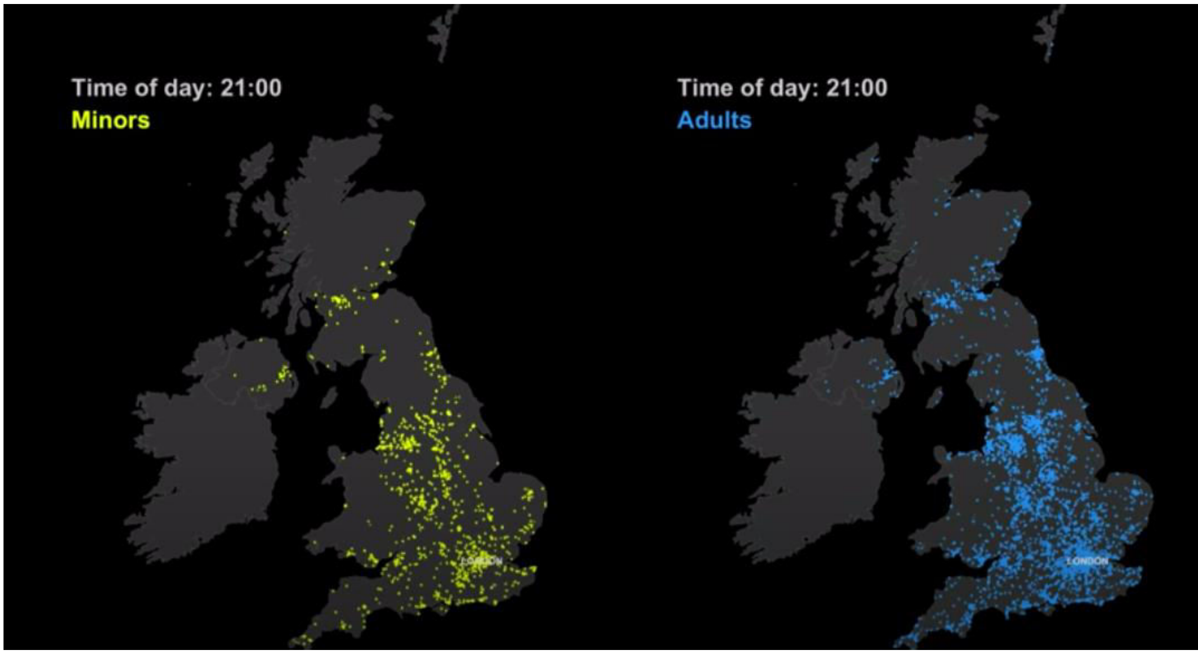


*Fig 1 Caption: Satellite tracking of adults and minors [9]*

As shown above, *Fig 1*, depicts the real time satellite tracking in the normal life of people in the location recorded. This data shows the number of commuters in different places across the country. Based upon this data, it becomes easier to predict where cases could rise quickly after analysing where most people travel to and fro from. Though in this case, the sets show different possible 'hotspots' as it examines different age groups. However, this problem can be solved. These sets could become one cohesive set after combining the commuter number for both minors and adults, thus giving an estimate of the population travelling across the country and the places where travel is more frequent. This could result in earlier allocation of preventive measures like imposing night curfews, weekend lockdowns and restriction of the number of people attending social events, thus slowing the spread of the virus.

**Prediction Factor B**

The second predictive factor comes into play in a later stage where a number of people have already been infected and are now susceptible to carrying the virus. This factor is actually a number of different factors in synchronisation. Though this dataset is used much later into the pandemic, it is more effective than Predictive Factor A since it is actually dealing with the real time statistics of those who have been infected, those who have recovered, the death count and the population of the area, based on the level that the data scientists are dealing with. This method is more efficient in analysing smaller ecosystems. Such analysis on a national level would be more inconvenient as it requires more data than other methods.

In this predictive method, several factors are taken into account:

1. The death count of that location
2. The number of infected people
3. The number of contacts for the infected and susceptible people

ISSN 2350-109X
www.indianscholar.co.in
Indian Scholar
An International Multidisciplinary Research e-Journal

4. The number of recovered people
5. The population of that location (including minors)

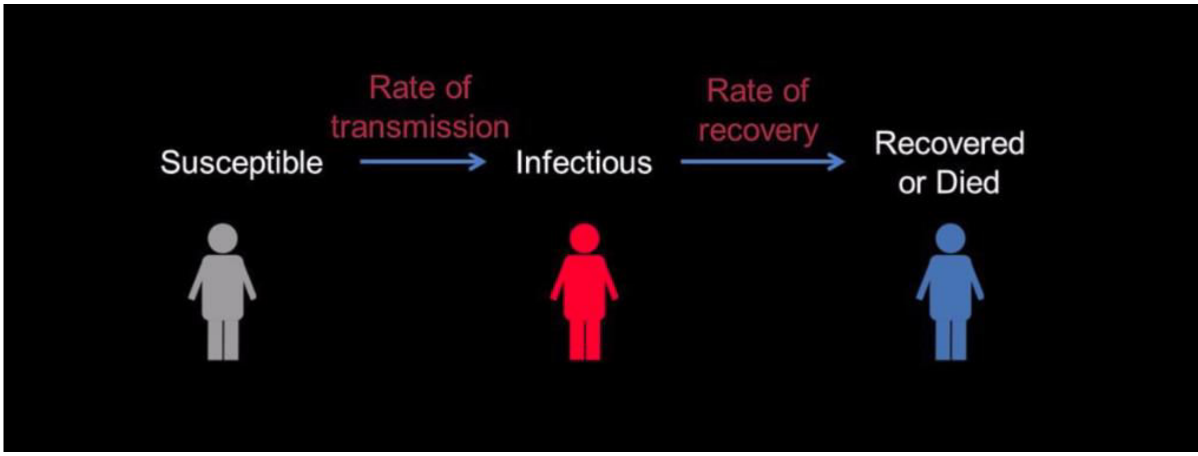


*Fig 2: Caption: Factors involved in prediction [10]*

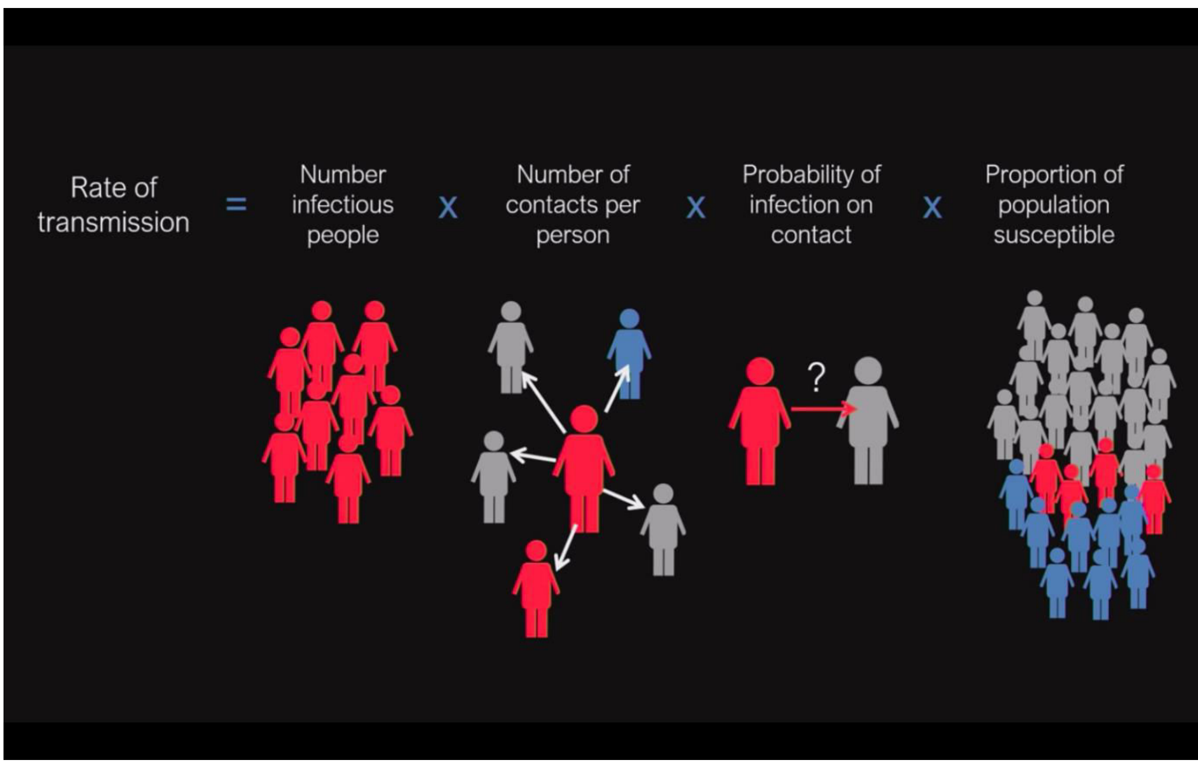


*Fig 3: Caption: Formula to predict the rate of transmission, to predict case count[11]*

As shown in *Fig 3*, the transmission rate can be predicted by multiplying the number of infectious people by the number of contacts by the probability of the infection on contact by the susceptible population proportion. Another method around this is subtracting the death count, recovery count and case count from the total population.

*Total population= Number of people who died + Number of people who recovered + Number of people who have contracted the virus + Remaining population vulnerable to the virus*

This gives a very accurate estimate on the number of people who are still vulnerable to the virus, as compared to the population on a whole, thus giving a ratio of vulnerable to the total population. This, once multiplied by the rate of transmission, can be used to predict cases in that location.

This method is effective with small populations as it gives an accurate prediction and can help the governing bodies to decide on which steps to take next. But other more widely available predictive factors can be used to reach a conclusion for large scale communities and ecosystems, that would otherwise require more data if this predictive factor was applied. Based on the predictions, the leading body of the community can decide on which activities can be cut down on. The body can also decide to which extent must these activities be held off and for long they must be held off. Also, the decision of which activities should go on to avoid the economy from coming to a complete standstill can be made.

**Predictive Factor C**

Predictive Factor C is not used for prediction of positive cases in a certain area. This factor uses another aspect of Data Science helping in battling epidemics, it aids the immediate response sector like hospitals, testing centres and other healthcare based institutions that are helping in battling the pathogen. Despite this, Factor C uses Data Science at its heart to get the information needed to carry out the required action and allocation. This factor uses predictive graphs to help understand the medical supplies needed to battle the virus. As an epidemic is a sudden jerk to the healthcare ecosystem, it is difficult to predict the medical supplies that will be needed to treat patients who have contracted the virus. Furthermore, it is difficult to predict cases, which will indirectly affect the prediction of the stockpile of medical equipment, including medicines, needed to treat those patients.

As the number of cases increased, the World Health Organisation or WHO, came up with a 'Hospital Readiness Checklist for COVID 19'. [12]

ISSN 2350-109X
www.indianscholar.co.in
Indian Scholar
An International Multidisciplinary Research e-Journal
MEMBERS OF INTERNATIONAL SCHOOLS' ASSOCIATION

World Health Organization

RAPID HOSPITAL READINESS CHECKLIST FOR COVID-19

The WHO 2019 novel coronavirus (2019-nCoV): strategic preparedness and response plan for novel coronavirus disease (COVID-19) outlines the public health measures that need to be considered by countries to prevent, prepare for and respond to the COVID-19 pandemic. The plan supports countries to rapidly identify relevant actions from their national action plans for health security and pandemic influenza preparedness, which are pertinent to managing the COVID-19 pandemic and can be adapted using the knowledge that has been gained about the COVID-19 virus. Using these plans will help guide and align all national and international partners to support national governments in managing the pandemic. The nine pillars of the Strategic Plan for COVID-19 refer to different aspects of the pandemic and allow countries to develop capacities to respond to the crisis, including by taking measures to strengthen their health systems. In the context of COVID-19, specialized services offered by health care facilities, especially those available in hospitals, will be required to serve the affected population.

On 30 January 2020, the Director-General of the World Health Organization, declared the COVID-19 outbreak to be a global public health emergency of international concern under the International Health Regulations. Following the spread of COVID-19 cases in many countries across continents, COVID-19 was characterized as a pandemic on 11 March 2020 by the Director-General, upon the advice of the International Health Regulations Emergency Committee.

The current and rapidly evolving nature of the COVID-19 pandemic requires hospitals to have in place all essential preparedness measures. These measures may vary depending on the designated role of each hospital and the way in which hospitals are linked to the country's overall plan for managing the pandemic while continuing to provide essential services to other patients who require care. These roles are likely to include (a) testing people for COVID-19 and managing early investigations to identify confirmed cases who require hospital care, (b) providing treatment for COVID-19 cases, (c) continuing to provide routine essential health services, (d) preventing patients from acquiring COVID-19 while in hospital , and (e) communicating information on COVID-19 as part of the country's and hospitals' risk communication strategy in coordination with the central response system and communities, with the aim of containing and mitigating the pandemic.

This WHO rapid hospital readiness checklist for COVID-19 was developed based on WHO's COVID-19 strategic response and preparedness plan, hospital preparedness for epidemics and interim versions of similar checklists from the WHO Regional Office for the Americas/Pan American Health Organization, the WHO Regional Office for Europe and WHO headquarters. The purpose of the Checklist is to help hospital managers prepare for COVID-19 patient management by optimizing each hospital's capacities. It has been designed to be user-friendly, taking into consideration the human resources and the assessment time required to conduct and complete the checklist in its entirety.

*Fig 3: Caption: A Tabular Summary of the WHO Hospital Readiness Checklist[13]*

The World Health Organisation has also established detailed guidelines and a certain standard set per criterion for every hospital to go by through all stages of the pandemic [14]
Over the course of the pandemic however, several hospitals, particularly over the course of the passing of the Delta variant in India in early January, have struggled with severe shortage of supplies [15]
This situation could have been avoided had hospitals gotten some idea on the estimated amount of supplies needed for that particular period. A recent survey shows how unprepared hospitals were for the second wave, which in turn slowed down the recovery rate [15]
Despite the fact that the epidemic is a completely unpredictable healthcare scenario, to a certain extent, an estimate of the supplies could've been taken. Using data from previous cases in the area, the hospitals could've been assigned a certain amount of emergency stock. Based on carrying factors like the hospital's popularity amongst the crowd, the location and how convenient it is to travel to, hospital budget, rooming capacity, staff number, staff qualification, type of hospital (specialisation). number of positive tested patients checked in and ranking in terms of facilities available, hospitals should be allocated a minimal amount of resources to use for treating patients. In this manner, hospitals of all areas and all specialisations will, to a minimal extent, get access to basic resources and equipment needed to treat and help patients recover from the virus.

While this factor is time consuming and will need more data as the population increases, it is one of the fewer and more fool proof ways to understand the hospitals' readiness in terms of supply of vaccines and important dosages and medications for the surge in cases and their equippedness in terms of hospital staff and qualification. Despite the detailed analysis that will come out of such a survey, this factor can only be restricted to less densely populated areas as such an in-depth analysis will be difficult when the area is more widespread as the place will have more hospitals.

However, with the World Health Organisation's Checklist, it becomes easier. The implementation of government rules could help understand the preparedness of hospitals. If the government (Central or State), gives out rules that every public and private hospital has to submit the checklist. If made a compulsion, this rule can help immensely. If this is followed by allocating a separate Cabinet to affairs of recovery from the virus, effective measures can be taken more quickly, thus reducing the potential risk. The creation of such a cabinet will streamline the process of allocation of resources. This particular body could analyse reports submitted by hospitals from that region. After this, it could submit a collective estimate of the supplies of injections, masks, sanitisers, Personal Protective Kits, vaccines and oxygen cylinders that the region needs.

This could not only help in the current situation, it could also help streamline the healthcare ecosystem. This division of work will help people on a micro level, thus affecting the population on a macro level. This, in turn, will help nations prepare themselves for the next epidemic or major global health crisis in the foreseeable future. This particular method will take time. However, it is one of the long term benefits and should not be overlooked. With more people in such a Cabinet and more data scientists and surveyors, this particular method could become streamlined and could eventually become a part of our healthcare system.

**Predictive Factor D**

Predictive Factor D is a vastly different statistical and analytical factor. It involves Artificial Intelligence and Data Science working hand in hand.

This factor requires information from past pandemics and major health crises like the Bubonic Plague (1346 – 1352), Influenza Virus A (H1N1) (1918), Influenza Virus (H2N2) (1918-1919), Human Immunodeficiency Syndrome (HIV)(1981) and other pandemics to understand the connection between the causes of these viruses, particularly zoonotic diseases, as those are the most uncertain amongst other types of viruses.

The Bubonic Plague's outbreak was caused by a bite from a contaminated flea [16]. Swine Flu was caused by pigs [17]. The Avian Virus began due to a contaminated bird, most likely a chicken [18]. The novel SARS-CoV-2 began due to a contaminated bat.

These examples all have zoological and habitual connections. Such research has already begun by the BBC [19].

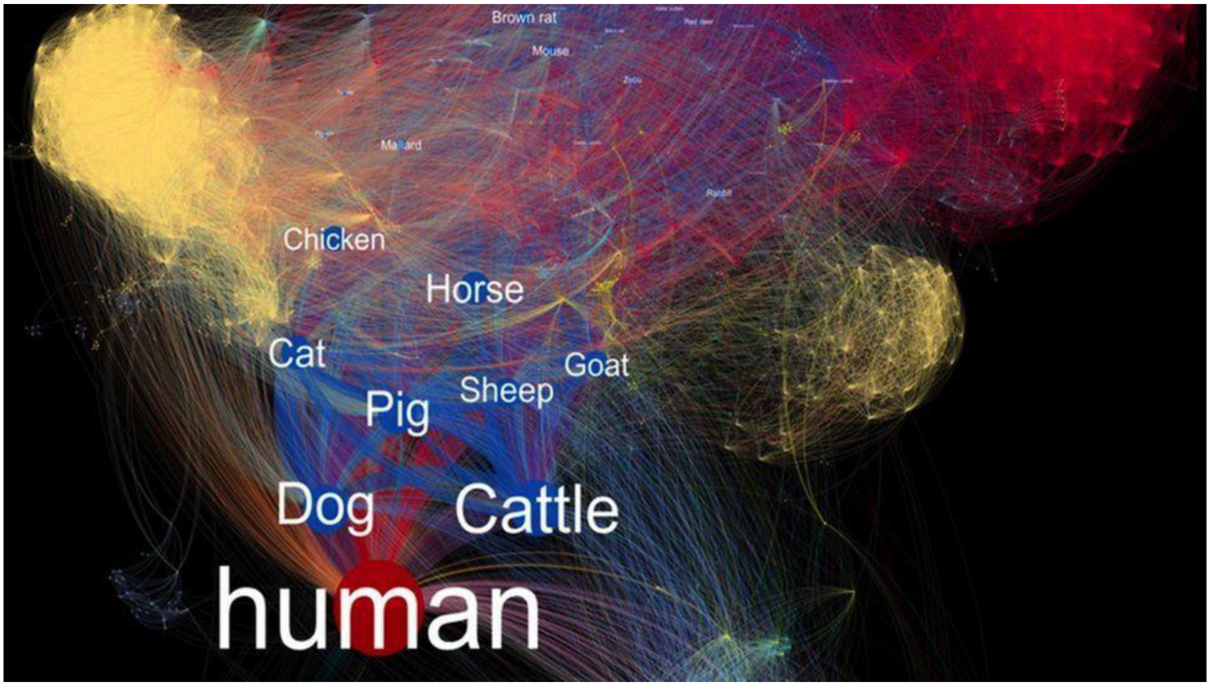


*Fig 4: Caption: A Visual Representation of the Potential Diseases between Different Species [20]*

This research involves Data Science to collect information on every epidemic and pandemic in history, along with the causes, symptoms, severity and cure. With this information, connections can be made. It will become easier to predict the next virus and its nature. This can be predicted by understanding the location of the disease's rise, the animal species, the behaviour and habitat of the species and how it has affected other species in relation. This will save a lot of unexpected trouble that will arise in the future when the next pandemic hits, as it is a known fact that the novel coronavirus will not be the last.

**Conclusion**

All of the evidence above thus proves the importance of Data Science in all parts of dealing with a pandemic: predicting the cases through the different waves and mutations, the imposition of lockdowns and curfews, immediate response for vaccination and prediction of the next epidemic. With Data Science, we have an advantage our ancestors did not have: technology. Using pure statistics along with Machine Learning and Artificial Intelligence, we can get through this pandemic and prepare ourselves for the next pandemic, now that we are truly aware of the impending danger a virus can bring.

**Acknowledgements**

A special vote of thanks to Ankit Ramchandani and his research paper on Machine Learning for Pandemics: https://ieeexplore.ieee.org/abstract/document/9179729

A gracious thank you to the Royal Society Publishing for their 'model' research paper that helped immensely.

A thank you to the World Health Organisation for the 'hospital Readiness Checklist' that helped in more ways than one. Thank you to Adelaide Writing Centre for their guide to writing a research paper. Lastly, thank you to JMIR Publications for their ever so helpful research paper on Big Data Analytics' usage in the fight against COVID 19.

**REFERENCES**

1. Anonymous, 'The Bubonic Plague', Edited July 17, 2021
2. Centers for Disease Control and Prevention, 'The History of 1918 Flu Pandemic', March 21, 2018
3. Centers for Disease Control and Prevention, 'Types of Influenza Viruses', November 18, 2019
4. Anonymous, '1957-1958 Flu Pandemic', 4 July, 2021
5. Centers for Disease Control and Prevention, 'HIV', June 1, 2021
6. World Health Organisation, 'Coronavirus Disease', October 12, 2020
7. Benjamin Daziel, BabakPourbohloul and Stephen P. Ellner, The Royal Society, 'Human mobility patterns predict divergent epidemic dynamics among cities', September 7, 2013
8. Bernard Marr, Forbes Magazine, 'Coronavirus: How Artificial Intelligence, Data Science And Technology Is Used To Fight The Pandemic', March 13, 2020[7]: Benjamin Daziel, BabakPourbohloul and Stephen P. Ellner, The Royal Society, 'Human mobility patterns predict divergent epidemic dynamics among cities', September 7, 2013
9. Bernard Marr, Forbes Magazine, 'Coronavirus: How Artificial Intelligence, Data Science And Technology Is Used To Fight The Pandemic', March 13, 2020
10. Ted Archive, Adam Kucharski, 'How Data Can Predict the Next Pandemic', January 24, 2019
11. TEDx Talks, Rosalind Eggo, 'Epidemics and the End of Humankind', July 25, 2018
12. TEDx Talks, Rosalind Eggo, 'Epidemics and the End of Humankind', July 25, 2018
13. World Health Organisation, 'Rapid Hospital Readiness Checklist', November 25, 2020
14. World Health Organisation, 'Rapid Hospital Readiness Checklist', November 25, 2020
15. World Health Organisation, 'Rapid Hospital Readiness Checklist', November 2020 (Detailed PDF)
16. Ruth Pollard, Sudhi Sen, Bloomberg, 'It's Like a War', April 28, 2021
17. Centers for Disease Control and Prevention, 'Information on Swine/Variant Influenza', August 6, 2018
18. Centers for Disease Control and Prevention, 'Avian Influenza A Virus Type Infection in Humans', April 18, 2017
19. Victoria Gill, BBC News, 'AI used to predict the next coronavirus', February 16, 2021
20. Victoria Gill, BBC News, 'AI used to predict the next coronavirus', February 16, 2021[7]: Benjamin Daziel, BabakPourbohloul and Stephen P. Ellner, The Royal Society, 'Human mobility patterns predict divergent epidemic dynamics among cities', September 7, 2013
21. Bernard Marr, Forbes Magazine, 'Coronavirus: How Artificial Intelligence, Data Science And Technology Is Used To Fight The Pandemic', March 13, 2020
22. Ted Archive, Adam Kucharski, 'How Data Can Predict the Next Pandemic', January 24, 2019
23. TEDx Talks, Rosalind Eggo, 'Epidemics and the End of Humankind', July 25, 2018

24. TEDx Talks, Rosalind Eggo, 'Epidemics and the End of Humankind', July 25, 2018
25. World Health Organisation, 'Rapid Hospital Readiness Checklist', November 25, 2020
26. World Health Organisation, 'Rapid Hospital Readiness Checklist', November 25, 2020
27. World Health Organisation, 'Rapid Hospital Readiness Checklist', November 2020 (Detailed PDF)
28. Ruth Pollard, Sudhi Sen, Bloomberg, 'It's Like a War', April 28, 2021
29. World Health Organisation, 'Plague', October 31, 2017
30. Centers for Disease Control and Prevention, 'Information on Swine/Variant Influenza', August 6, 2018
31. Centers for Disease Control and Prevention, 'Avian Influenza A Virus Type Infection in Humans', April 18, 2017
32. Victoria Gill, BBC News, 'AI used to predict the next coronavirus', February 16, 2021
33. Victoria Gill, BBC News, 'AI used to predict the next coronavirus', February 16, 2021

10