

Role of Big Data in Monitoring and Tracking COVID 19.

Soumya Koshy¹, Akshara Aji², Cini Joseph³, Naveen Thomas Joseph⁴

(¹Assistant Professor, Kristu Jyoti College of Management and Technology, Kottayam, Kerala, India)

(²MCA, Kristu Jyoti College of Management and Technology, Kottayam, Kerala, India)

(³Assistant Professor, KristuJyoti College of Management and Technology, Kottayam, Kerala, India)

(⁴MCA, Kristu Jyoti College of Management and Technology, Kottayam, Kerala, India)

Abstract.

We are now in a world reshaped by the Covid-19 pandemic. This pandemic had hit every aspect of our day-to-day life. Early and rapid case identification is crucial during a pandemic for the isolation of cases and appropriate contacts in order to reduce onward spread and understand key risks and modes of transmission. Accurate and up-to-date statistics from governments are not always accessible. The number of persons affected by Covid-19 is increasing rapidly and the death rate is also going high. We have seen first and second wave of the pandemic. The conditions are more severe in the second wave than in the first wave in many countries like India, South Korea, New Zealand, UK and so on. The virus had undergone various genetic mutations, which is more deadly and rapidly spreads from persons to persons.

This sudden outbreak of the pandemic had created vital source of information and knowledge. And there is also an immediate need to store the large amount of data about the cases, using data storage technologies. Big data technology is used to store the massive amount of data of the patients and also minimize the risk of spreading the virus. These data help in fighting the pandemic, as it helps in research and study about the virus, vaccines and also to take precautions from getting affected. It also helps in understanding the nature of virus in detail. Totally, the big data technology provides information about disease transmission, details about virus, details of confirmed patients, recovered patients, quarantined persons, details of hospitals etc.

There is an urgent need for the identification of effective therapeutics for Covid-19, in order to put a full stop to the Covid-19 disease. Big data is characterized by high volume, velocity and variety of data. Vaccine research using big data technology avoid cost, time-consuming and labor-intensive steps. Health scientists use every tool to crunch the number of patients and find the ways to create the drugs the world needs and big data helps in achieving it. Big data tools help in aggregating and synthesizing information to generate, validate and optimize novel drug candidates. Also, it helps in identifying patterns in virus data, make appropriate predictions, and potentially uncover the right drugs for testing. Computers are better to test and provide different chemical combinations and interactions between virus and drugs. Let's have a look on each topic deeply.

Keywords.

Big data, COVID-19, Datasets, Dashboards, Database, Machine learning, Vaccine

I. Introduction

Big data refers to collection of data that is so huge and complex that none of the data management tools are able to store it or process it efficiently. Big data plays a key role in reconfiguring academic, social, industry, business, knowledge and so on. Big data and technology are going hand in hand in managing the pandemic crisis. There is vast amount of data collected in real time by private sector actors that could be helpful to monitor the impact of the pandemic on mobility, migration and on migrants' lives. That is data from mobile phone operators, social media and other internet-based platforms. These data have been used successfully in past health crisis such as the Ebola outbreak, swine flu and so on.

The rapid outbreak and spread of COVID-19 pandemic had made the life of people miserable and taken away life of many people. So, to tackle this problem many doctors, researchers, brilliant data scientists, epidemiologists, public health experts and government authorities were brought together to find a solution. In response to Covid-19 some governments have partnered with main telecommunications operators in their countries to analyze anonymized data and see how mobility restrictions or social distancing measures are

affecting the spread of the virus. Every day they come up with new news about Covid-19. They are able to do so, only because of the use and regular analysis of such big data systems. There are several risk factors like age, locations, pre-existing health conditions, weather and climate, personal hygiene, number and level of contact with infected persons which can be used for preparing a better model about the spread of disease.

Several graph-based datasets and dashboard models can be modeled to provide critical information to slow down the spread of disease. Dashboards are extensively being used in the pandemic to collect real-time public health data, including confirmed cases, deaths, testing figures, to keep the public informed and support health authorities. We can also analyze how masking, social distancing, hand hygiene, lockdown had reduced the spread. Data can help authorities identify the most vulnerable communities. Data is also being used to develop therapies and vaccines. As the number of patients are increasing day by day it is difficult to collect and analyze each record. A solution to the huge data records is the use of big data and its applications. These data can be used for many predictions.

II. Methodology

Monitoring Patients Results – As the number of patients affected by coronavirus is increasing each day, so as the data about them is also increasing. In-order to store huge amount of data, the health care community is seeking the options such as big data and machine learning applications to monitor the patient details. The doctors can analyze the patient data, give first aid, allocate resources, study disease pattern and provide proper care. For example, the data about senior citizens with diabetes, obesity, blood pressure or persons taking several drugs for certain diseases can be collected. This data helps in giving distinct diagnosis and care to different people.

Databases and data visualization platforms have helped hospitals track and predict the availability of resources, identifying patterns and helps in clinical decision-making. These datasets and software can give recommendations in treatment such as hospitalizing, giving first aid etc.

Tracking the spread of COVID-19 – Many data scientists and public health experts keenly observe the spread of Covid-19, especially the locations, timing of infection, hotspots, etc. In past the data was collected using pen and paper. It results in huge piles of files and papers. If we want to search a file or a particular data, we had to go through many files. Also, the collection and storing of these records safely takes much time, money and human effort. Tracking the spread of Covid-19 using such traditional methods may take heavy investments in labor and time. Wasting time and effort on such old systems may make the situation worse.

By using big data tools like Apache Hadoop-to handle batch processing, Apache Spark-to handle real time data, Apache Cassandra-for parallel and real time data processing, MongoDB-NoSQL database for high volume data storage, virus tracking dashboards can be created. These dashboards can provide many vital information from many countries such as confirmed number of cases, death ratio, severe spread locations, available resources, precautions taken etc. Also, these models can predict the peak infection rates and its impacts. To avoid mass spread of this pandemic, public health authorities can collect data using big data algorithms from data of health records, airline ticketing data, news reports, government press release, social media posts and all data records entered manually. Datasets can be used to compare the countries which imposed complete lockdown with the restricted lockdown and analyze the variation in the number of Covid-19 cases. The list of covid positive cases, disease spreading areas, disease spreading age groups, can be extracted from the collected datasets. The datasets contain all information about a positive case like name, age, location, contact level (direct primary contact or secondary contact), source of disease, positive confirmed date, recovery date which can be used for various studies and observations.

These data are capable of predicting community spread locations, hotspots and warn the healthcare authorities to take the needed precautions. Through contact tracing we can find the relationship between infected parties and isolate them from further contacts.

Contact Tracing using mobile apps – Scientists estimate that on infected person passes the virus onto around two to three people on average. Many mobile apps like ‘Aarogya Setu’ app developed by Indian government, ‘Close Contact Detector’ developed by Chinese government, ‘GuideSafe’ developed by Alabama, ‘Covidwise’ by Virginia, etc. have been developed by several countries to track and trace contact

with covid positive patients. The mobile app users can trace their contacts and check whether they are exposed to the virus or not. The app shows the number of positive cases near the app user. In this way it provides personal tracking, early warning to the public. These apps can also give self-assessment service to the users to identify their health status and also symptoms. These apps can provide valuable data for coronavirus control strategy. People who are quarantined can use app to communicate with the local government case officers and report their symptoms. The app is not mandatory for everyone to use, and people can opt out.

Tracking hospital details – In a pandemic situation, knowing at a glance how many hospital beds are available in each hospitals and which hospital is best equipped to deal with incoming casualties or illness. Keeping tabs using the big data tools such as Apache Hadoop, MongoDB gives the number and condition of patients as they are admitted in hospital is important and crucial. Each hospital can keep its own datasets and records which helps in future references. For example, the details like the number of number patients admitted in the hospital, patient details-name, age, health condition, diagnosis given, admitted date and discharge date, consulted doctor can be recorded. Details including positive cases, recovered cases, death ratio, treatment given can be used for further verifications and predictions.

Helps in vaccine research – The impact of Covid-19 on the world is growing day by day. An effective vaccine is certainly needed to make an end to the COVID pandemic. The data versatility is being put into test and track the variants of COVID-19 virus and find an effective vaccine. Nowadays more scientists are relying on machine learning to predict and find an interaction between virus protein and existing drugs. For example, mainly four combinations are under study namely Remdesivir (in cure), Ribavain (in combination with others for cure), Umifenovir (as a prophylactic and cure) and Sofosbuvir(as a cure). Another prediction is Tenofovir alafenamide, which is a novel drug prodrug developed in order to improve renal activity. Another successful example of rationally designed web-based vaccine is the vaccine against Neisseria Meningitidis, commercially available with the trade name of Bexsero. It takes many years to produce a new vaccine, but now it had become somewhat easy to find the relationship between drugs and virus using the datasets to tests several combinations. Computers are more capable than human researchers in finding several chemical combinations. It also helps to develop many vaccines and also find the best drug to treat COVID-19.

About data privacy

Data had become an inevitable part of technological world. Data privacy focuses on proper handling of data. Data privacy is centered around how data should be collected, stored, managed and shared with third parties as well as compliance with the applicable privacy laws. Challenges with the quality and consistency of data remain a concern. Lack of official standards and inconsistencies in government reporting statistics across countries make global comparison difficult. A huge amount of personal data we are releasing, because we want to fight Covid-19, and we have opened the door to allowing our data to be surveilled in many ways we never would have allowed in the past.

It is important to make sure that the personal data is used for right purpose and data ethics is maintained throughout the proceedings. There are growing privacy concerns about the ways governments use data to respond to the COVID-19 crisis. Even in this pandemic crisis, it is good to ensure that data privacy regulations are followed and data is used ethically.

III. Conclusion

In future big data will play important role in analyzing data about many virus, new disease, tracking human activity. Technologies like big data, machine learning, artificial intelligence can be effectively used to assist researchers in getting prepared for future pandemic. Big data is useful in guiding policy, but also it is important to check the privacy policies.

III. References

1. Susheel George Joseph, “The Usage of Machine Learning Evolutionary Algorithms in Medical Images Formed by Computer Tomography (CT) or X-Rays to Detect the Infections due to COVID 19”, PENSEE (penseereseach.com) ISSN: 0031-4773. Volume 51, Issue 4, Page No:1512-1518, April 2021. Available at: <https://app.box.com/s/n6nsv8myosb0wb16psvtb8conekpy ohj>

2. Susheel George Joseph, "A Machine Learning (ML) Modelling Approach in Monitoring and Controlling the Viral Pandemic- COVID 19", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 6, page no.1709-1717, June 2020: <http://www.jetir.org/papers/JETIR2006575.pdf>
3. S.Binny, "A survey concept on Deep Learning", International Journal of Scientific & Engineering Research (www.ijser.org), ISSN 2229-5518 , Volume 10, Issue 6, page 1570-1575, June-2019.
4. Susheel George Joseph, "A Machine Learning (ML) Modelling Approach in Monitoring and Controlling the Viral Pandemic- COVID 19", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 6, page no.1709-1717, June 2020: <http://www.jetir.org/papers/JETIR2006575.pdf>
5. Susheel George Joseph, Dr. Vijay Pal Singh, "Denoising of Images using Deep Convolutional Neural Networks (DCNN)", International Journal of Engineering Development and Research (IJEDR), ISSN:2321-9939, Volume.7, Issue 3, pp.826-832, September 2019, <http://www.ijedr.org/papers/IJEDR1903143.pdf>.
6. Emanuel EJ, et al. Fair allocation of scarce medical resources in the time of covid-19. N Engl J Med. 2020.
7. Livingston E, Bucher K. Coronavirus disease 2019 (COVID-19) in Italy. JAMA. 2020;323(14):1335.
8. Lee, I.-K.; Wang, C.-C.; Lin, M.-C.; Kung, C.-T.; Lan, K.-C.; Lee, C.-T. Effective Strategies to Prevent Coronavirus Disease-2019 (COVID-19) Outbreak in Hospital. J. Hosp. Infect. 2020, 105, 102–103.
9. Iacobucci, G. Covid-19: Emergency Departments Lack Proper Isolation Facilities, Senior Medic Warns. BMJ 2020, 368, m953.
10. Worldometers Coronavirus Cases. Available online: <https://www.worldometers.info/coronavirus/> (accessed on 30 December 2020).
11. Da'Ar, O.B.; Haji, M.; Jradi, H. Coronavirus Disease 2019 (COVID -19): Potential Implications for Weak Health Systems and Conflict Zones in the Middle East and North Africa region. Int. J. Health Plan. Manag. 2020, 35, 1240–1245.
12. Ajah, I.A.; Nweke, H.F. Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications. Big Data Cogn. Comput. 2019, 3, 32.