Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 6): pp. 5312-5321
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.06.514

Big Data Analytics In Pandemic Like Covid-19: Importance, Challenges And Techniques For Early Detection And Prevention

Khalid Bashir Dar , Sheikh Mudasir Hussain Zargar , Dr. Niraj Singhal

Abstract

Big data analytics is a process of examining, cleaning, transforming, and modeling large sets of data to discover useful information and support decision-making. In healthcare, it can be used to improve patient outcomes by identifying patterns and trends in patient data, reducing unnecessary costs and identifying new drug targets. The use of big data analytics can play a critical role in improving the quality, accessibility, and affordability of healthcare for people around the world. In terms of pandemics like COVID-19, big data analytics can play a significant role in early detection and prevention of pandemics like COVID-19. By analyzing large amounts of data from various sources such as social media, electronic health records, and surveillance systems, patterns and trends can be identified that may indicate the emergence of a new disease. This information can then be used to quickly identify and track outbreaks, as well as to develop targeted prevention and control measures. Big data analytics can help detect and prevent by analyzing data on factors that contribute to the spread of a disease, using various algorithms like machine learning, network analysis algorithms, GLEAM algorithm, NLP algorithm etc. and using data visualization tools to present the results.

Key words: NLP, Big data, GLEAM, COVID-19.

1. INTRODUCTION

Big data analytics refers to the process of examining, cleaning, transforming, and modeling large sets of data to discover useful information, draw conclusions, and support decision-making. This process typically involves the use of advanced analytical tools, such as machine learning and statistical algorithms, to analyze and extract insights from very large, complex data sets. The goal of big data analytics is to help organizations gain insights and knowledge from their data that can be used to improve their operations, products, and services, and make more informed decisions.

Big data has the potential to revolutionize healthcare by providing new insights into disease prevention, diagnosis, and treatment [01]. Some specific ways in which big data can be used in healthcare by analyzing large amounts of patient data, healthcare providers can identify patterns and trends that can help them make more informed decisions about patient care.[02] This can lead to earlier diagnoses, more effective treatments, and improved patient

outcomes. By identifying inefficiencies in the healthcare system and identifying high-risk patients,[10] big data analytics can help healthcare providers reduce unnecessary costs.[03] Big data analytics can be used to identify new drug targets and improve the efficiency of clinical trials. Big data analytics can play a critical role in improving the quality, accessibility, and affordability of healthcare for people around the world.

2. Importance of big data analytics in health care during pandemic

Big data analytics plays a crucial role in early detection and prevention of pandemics like COVID-19.[20] The vast amount of medical data generated during a pandemic, including information on cases, deaths, hospitalizations, and testing, as in (figure 1.0) can be analyzed to identify trends and patterns that can help inform public health decisions.[14,19]

I. Real-time monitoring

One of the key advantages of big data analytics in the context of pandemics is its ability to provide real-time monitoring. By using big data and advanced analytics, health officials can monitor the spread of the virus in real-time, and quickly respond to changes in the situation, such as new outbreaks or mutations.[04] This can help to reduce the overall impact of the pandemic by identifying and containing outbreaks early on, before they have a chance to spread widely.[11] For example, data on travel patterns and social media activity can be used to identify and track outbreaks in real-time. In addition, data on hospitalizations and deaths can be used to identify high-risk populations and target interventions.

II. Predictive modeling

Another important use of big data analytics in the context of pandemics is predictive modeling. These models can be used to identify individuals who are at high risk of contracting the disease and to predict the spread of the disease in different geographic areas. [05] By analyzing large amounts of data from various sources, such as social media, news articles, and electronic health records, big data analytics can help identify which individuals and communities are most at risk of contracting a disease and predict the potential impact of a pandemic. This can help inform public health interventions and policies, such as identifying priority groups for vaccination or allocating resources to the areas that are most affected.

III. Contact tracing efforts

Big data analytics can also be used to support contact tracing efforts. By analyzing large amounts of data from various sources, such as mobile phone records, big data analytics can help identify and track individuals who have been in contact with someone who has tested positive for COVID-19. This can help to quickly identify and isolate individuals who may be at risk of contracting the virus, and prevent further spread. So analyze large volumes of data from a variety of sources. [21] This can include data from social media, search engines, and other online platforms, as well as data from traditional sources such as healthcare records and government

statistics.[06] Analyzing this data, big data analytics can provide insights into the spread of a disease, such as identifying patterns in the movement of infected individuals, or identifying areas where the disease is spreading rapidly.

For example, in the early days of the COVID-19 pandemic, big data analytics was used to track the spread of the virus by analyzing social media posts and search engine queries. This helped to identify hotspots for the disease, such as Wuhan, China, which was one of the earliest epicenters of the pandemic. Similarly, big data analytics was used to track the spread of the disease in other countries, such as Italy and the United States, and to identify areas where the virus was spreading rapidly.

IV. Resource allocation.

Big data analytics can also play an important role in resource allocation. By analyzing large amounts of data, big data analytics can help identify areas where resources, such as personal protective equipment, ventilators, and medical staff, are needed most urgently. This can help to optimize the distribution of these resources and ensure that they are being used in the most effective way possible. Big data analytics can also be used to track the distribution and effectiveness of vaccines. By analyzing data on vaccine distribution and uptake, big data analytics can help identify priority groups that need to be vaccinated first and monitor the effectiveness of the vaccine in different populations.[07,12] This can help ensure that the vaccine is being distributed equitably and that it is having the desired impact on public health.

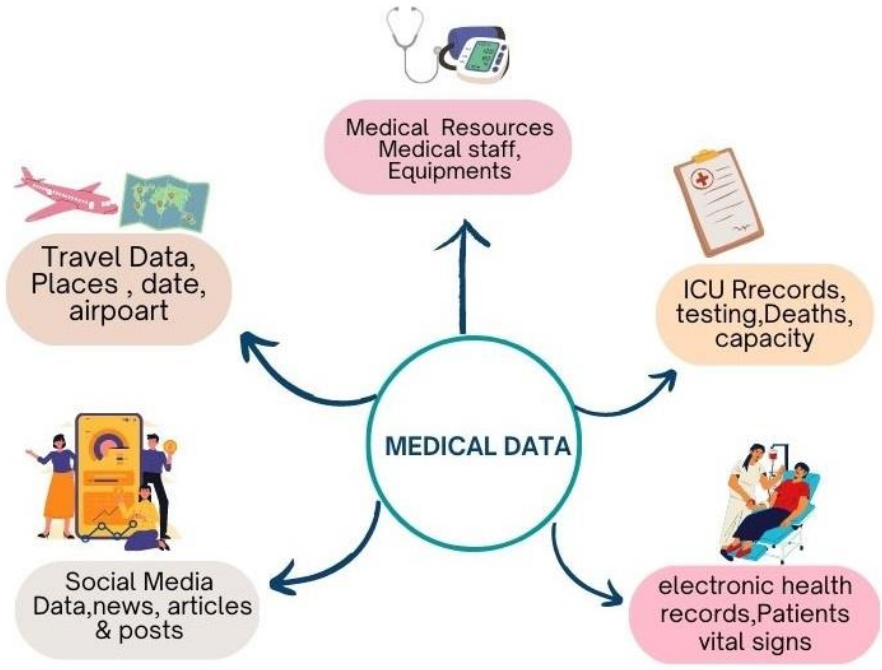

Figure 1.0 Sources of Data

3. Challenges in big data analytic during pandemic

i. Significant amount of data and sophisticated modeling techniques

However, despite these advantages, the ability of big data analytics to predict a pandemic like COVID-19 is limited. One of the main challenges is that predicting a pandemic requires a significant amount of data and sophisticated modeling techniques. [08] For example, predicting the spread of a disease requires data on the number of infected individuals, the rate at which the disease is spreading, and the characteristics of the population at risk. [13]

ii. Variety of data

Using big data for pandemic detection is the sheer volume and variety of data that must be analyzed. This includes data from a wide range of sources, such as electronic health records, social media, and government websites. [14] In addition, the data may be in different formats and may need to be cleaned and preprocessed before it can be analyzed. [15]

iii. Complexity of the disease

Another challenge is that even with large amounts of data and sophisticated models, the accuracy of predictions can be limited. This is because predicting a pandemic is a complex task that is influenced by a wide range of factors, including the characteristics of the virus, the behavior of infected individuals, and the response of public health officials.[09] Additionally, the complexity of the disease and the interactions between different factors can make it difficult to predict the spread of the disease with a high degree of accuracy.[16]

iv. Ability to analyze data in near real-time

Another challenge is the need for real-time analysis. In a pandemic situation, decisions must be made quickly to contain the spread of the disease and protect public health. This requires the ability to analyze data in near real-time and provide timely insights to decision-makers. [04,22]

4. Algorithms and techniques that can be used for big data analytics to detect and prevent pandemics

Several algorithms and techniques that can be used for big data analytics to detect and prevent pandemics like COVID-19.

One approach is to use machine learning algorithms to analyze large amounts of data on various factors that may contribute to the spread of a disease, such as population density, travel patterns, and social interactions. This can help identify potential hotspots for outbreaks and predict how a disease may spread in a given region. [17, 18]

1. **Network Analysis for Pandemics**

Another approach is to use network analysis algorithms to study the connections and interactions between individuals, such as who they come into contact with and how often. This can help identify key individuals or groups who may be more likely to spread a disease and target interventions accordingly. Network analysis is a technique used to analyze complex systems made up of interconnected elements, such as people, organizations, or even biological cells. This approach can be used to analyze large datasets and identify patterns or relationships that would be difficult to detect using other methods.

In the context of a pandemic, network analysis can be used to track the spread of a disease through a population. By analyzing data on how people interact with each other, researchers can identify clusters of individuals who are at high risk of infection and target them for interventions, such as vaccinations or quarantines. [23] One example of this approach is the use of contact tracing, which involves tracking the people that an infected individual has been in close contact with in order to identify others who may have been exposed to the disease. By analyzing the data collected through contact tracing, researchers can identify patterns in how the disease is spreading and target interventions to the areas of highest risk. Another example is the use of social network analysis to identify influential individuals or groups in a population that could be targeted with public health messages. For example, if a group of people who are highly connected in a social network are also at high risk of infection, public health officials could target them with targeted messaging to encourage them to take preventative measures to protect themselves and others. In addition, network analysis can also be used to track the spread of misinformation about a pandemic, which can be a major challenge for public health officials. By analyzing data on how information is shared on social media, researchers can identify sources of misinformation and target interventions to correct false information.

Overall, network analysis is a powerful tool for understanding and controlling the spread of a pandemic. By analyzing large datasets and identifying patterns in how a disease spreads, researchers and public health officials can target interventions to the areas of highest risk and minimize the impact of the outbreak.[24]

2. **EpiEstim Algorithm Pandemic Control**

One specific algorithm that has been used to analyze COVID-19 data is the EpiEstim algorithm, which uses a Bayesian approach to estimate the basic reproduction number (R0) of a disease, which is a measure of how contagious it is. This can help predict how quickly a disease will spread and inform public health interventions. EpiEstim is a statistical algorithm that can be used for big data analytics in the context of pandemic detection and prevention. The algorithm uses data on the spread of a disease, such as the number of reported cases and deaths, to estimate key epidemiological parameters, such as the reproduction number (R_0) and the effective reproduction number (R_t). [25] These parameters provide important information about how fast a disease is spreading and how effective current control measures are in slowing its spread. EpiEstim is based on a Bayesian estimation approach,

which involves using prior information and data to update our understanding of the disease outbreak. The algorithm uses a likelihood function to estimate the probability of the observed data given a set of model parameters, and a prior distribution to incorporate prior information about the parameters.

Once the algorithm has estimated the key epidemiological parameters, it can be used to make predictions about the future spread of the disease and to evaluate the effectiveness of different control measures. For example, the algorithm can be used to estimate the impact of social distancing measures on the reproduction number, or to predict the number of cases and deaths in the future. One of the advantages of EpiEstim is that it can be applied to a wide range of data sources, including both official case counts and data from digital surveillance systems, such as social media and search engines. This allows for real-time monitoring of the spread of disease and early detection of outbreaks, which can be critical for pandemic control efforts.

Overall, EpiEstim is a useful tool for big data analytics in the context of pandemic detection and prevention, as it can provide important insights into the spread of disease and the effectiveness of control measures. However, it is important to note that the algorithm is only as good as the data that is used as input and the assumptions made.

3. **GLEAM: Detecting & Preventing Pandemics.**

Another algorithm that can be used is the GLEAM algorithm, which is a global epidemiological model that simulates the spread of a disease based on data on population demographics, travel patterns, and other factors. This can help predict how a disease may spread in different regions and inform public health responses.

GLEAM (Global Epidemic and Mobility Model) is an algorithm that uses big data analytics to simulate the spread of infectious diseases, such as pandemics. It combines data on human mobility (e.g. air travel, migration patterns) with information on disease transmission to predict the spread of an outbreak and identify potential hotspots. One of the key features of GLEAM is its ability to incorporate real-time data, such as air travel patterns, to generate dynamic predictions of disease spread. [26] This can help health officials and policymakers make more informed decisions about how to respond to an outbreak. GLEAM can also be used to evaluate the effectiveness of different control measures, such as travel restrictions or vaccination campaigns. By simulating the spread of the disease under different scenarios, GLEAM can help policymakers determine the most effective strategies for preventing or slowing the spread of a pandemic.

Overall, GLEAM is a powerful tool that can be used to detect and prevent pandemics by providing real-time predictions of disease spread and identifying potential hotspots. However, it is important to note that the accuracy of the predictions will depend on the quality and completeness of the data used as input.

4. NLP algorithm approach to detect and prevent pandemic

Big data analytics using natural language processing (NLP) algorithms can be used to detect and prevent pandemics by analyzing large amounts of text data, such as news articles, social media posts, and electronic medical records.[27]

One approach is to use NLP algorithms to identify key words and phrases that are associated with a particular disease or outbreak.[28] For example, if there is an outbreak of a new strain of influenza, NLP algorithms could be used to search for mentions of "influenza," "flu-like symptoms," and "outbreak" in news articles and social media posts. By analyzing the frequency and location of these keywords, public health officials can get a sense of the extent and severity of the outbreak.

Another approach is to use NLP algorithms to analyze electronic medical records in order to identify patients who may be at risk of contracting a particular disease. For example, if an outbreak of a new strain of influenza is identified, NLP algorithms can be used to search through electronic medical records for patients who have recently been diagnosed with influenza or who have recently shown symptoms of influenza.[29] This information can then be used to identify patients who are at high risk of contracting the new strain of influenza and to take steps to prevent further spread of the disease.

Big data analytics using NLP algorithms can also be used to track the spread of the disease, identifying the geographic location of the outbreak, and predicting the potential scale of the outbreak. This can help public health officials to take appropriate actions, including implementing measures such as quarantining and isolating infected individuals, and monitoring the spread of the disease. [30]

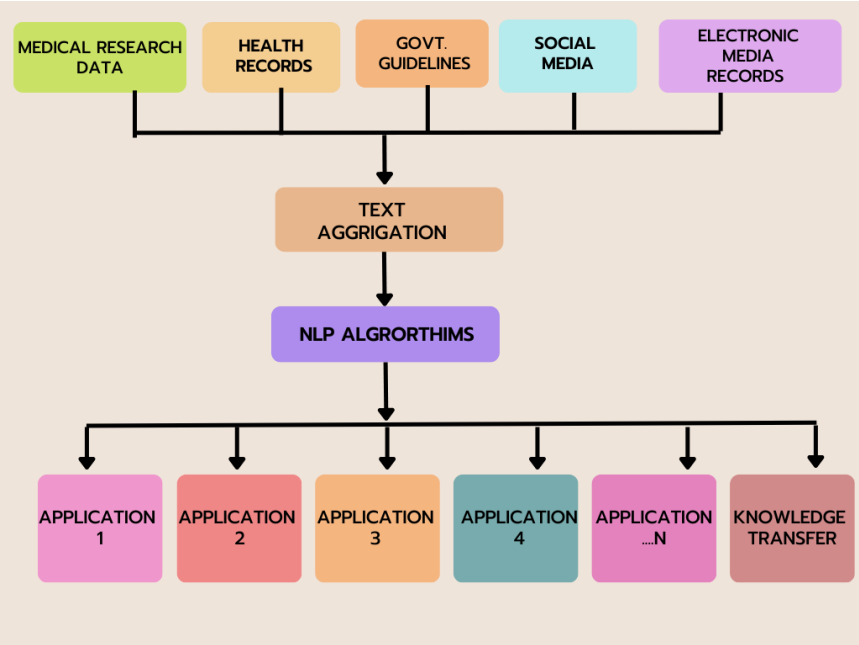

Figure 1.1 NLP for covid-19

In summary, big data analytics using NLP algorithms can be an effective tool for detecting and preventing pandemics by providing insights into the extent, severity, and spread of a disease, identifying high-risk individuals, and tracking the spread of the disease.

5. Data visualization

Data visualization tools can also be used to present the data and results of big data analytics in an easy-to-understand format for public health officials and decision-makers.

In summary, big data analytics can be used to detect and prevent pandemics like COVID-19 by analyzing large amounts of data on various factors that may contribute to the spread of a disease, using machine learning and network analysis algorithms, and using data visualization tools to present the results.

In conclusion, big data analytics plays a crucial role in early detection and prevention of pandemics like COVID-19. The vast amount of data generated during a pandemic can be analyzed to identify trends and patterns that can help inform public health decisions. However, the challenges of big data analytics in this context include the sheer volume and variety of data that must be analyzed, the need for real-time analysis, and the need to clean and preprocess the data. Big data analytics can provide a powerful tool for early detection and prevention of pandemics by allowing the identification and tracking of outbreaks, targeting interventions to high-risk populations, and monitoring the effectiveness of interventions.

REFERENCES

1. Bd-Haleem A, Javaid M, Khan IH, Vaishya R. Significant Applications of Big Data in COVID-19 Pandemic. Indian J Orthop [Internet]. 2020;54(4):526–8.
2. Benke K, Benke G. Artificial intelligence and big data in public health. Int J Environ Res Public Health. 2018;15(12).
3. Dolley S. Big data's role in precision public health. Front Public Heal. 2018;6(68):1–12. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. J Infect Dis. 2016;214(Suppl 4):375–9.
4. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time., Lancet Infect. Dis. 20 (2020) 533{534}.
5. PREDECTING EARLY----Z. Liu, P. Magal, O. Seydi, G. Webb, Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data, Mathematical Biosciences and Engineering 17 (4) (2020) 3040{ 3051. doi:10.3934/mbe.2020172
6. Nour Eldeen M Khalifa, Mohamed Hamed N Taha, Aboul Ella Hassanien, and Sarah Hamed N Taha. The detection of covid-19 in ct medical images: A deep learning approach. In Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach, pages 73–90. Springer, 2020.
7. WHO EMRO | Questions and answers | COVID-19 | Health topics.
8. Femke Mulder, Julie Ferguson, Peter Groenewegen, Kees Boersma and Jeroen Wolbers, "Questioning Big Data: Crowdsourcing Crisis Data Towards an Inclusive Humanitarian Response", Big Data & Society, 10 August 2016, available at: https://doi.org/10.1177/2053951716662054.
9. See UN Global Pulse, above note 8, pp. 24–34; Miguel Luengo-Oroz, "10 Big Data Science Challenges Facing Humanitarian Organizations", UNHCR Innovation Service, 22 November 2016, available at: www.unhcr.org/innovation/10-big-data-science-challenges-facing-humanitarian-organizations;
10. Cori L, Bianchi F, Cadum E, Anthonj C. Risk Perception and COVID-19. Int J Environ Res Public Health 2020 Apr 29;17(9):2020
11. Coronavirus disease (COVID-19) situation reports. World Health Organization. 2019. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
12. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. Library Catalog: www.who.int.
13. Patrick Meier, Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response, Routledge, New York, 2015, available at: https://doi.org/10.1201/b18023.
14. Kiran Adnan and Rehan Akbar, "An Analytical Study of Information Extraction from Unstructured and Multidimensional Big Data", Journal of Big Data, Vol. 6, Article No. 91, 17 October 2019, available at: https://doi.org/10.1186/s40537-019-0254-8.
15. Danah Boyd and Kate Crawford, "Critical Questions for Big Data", Information, Communication and Society, Vol. 15, No. 5, 2012, pp. 662–663.

16. Iffat Idris, Benefits and Risks of Big Data Analytics in Fragile and Conflict Affected States, 17 May 2019,available at: https://tinyurl.com/5xmftqcy.
17.] Caroline Buckee, "Improving Epidemic Surveillance and Response: Big Data Is Dead, Long Live Big Data", The Lancet Digital Health, Vol. 2, No. 5, 17 March 2020, available at: https://doi.org/10.1016/S2589-7500 (20)30059-5.
18. C. JasonWang,MD, PhD Medicine, University of California ; Response to COVID-19 in Taiwan Big Data Analytics, New Technology, and Proactive Testing JAMA Published online March 3, 2020 American Medical Association.
19. Fagerberg, J., Landström, H., & Martin, B. R. (2012). Exploring the emerging knowledge base of 'the knowledge society'. Research Policy, 41(7), 1121-1131.
20. Grover, V., Chiang, R. H., Liang, T. P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. Journal of Management Information Systems, 35(2), 388-423.
21. Zicari, R. V. (2014). Big data: Challenges and opportunities. Big data computing, 564, 103-105.
22. Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, 263-286.
23. Ahmed, W.; Vidal-Alaball, J.; Downing, J.; Seguí, F.L. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J. Med. Internet Res. 2020, 22, e19458.
24. Eikenberry, S.E.; Mancuso, M.; Iboi, E.; Phan, T.; Eikenberry, K.; Kuang, Y.; Kostelich, E.; Gumel, A.B. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infect. Dis. Model. 2020, 5, 293–308.
25. Cori MA. Package "EpiEstim" Title Estimate Time Varying Reproduction Numbers From Epidemic Curves. (2019).
26. Vespignani, A. Predicting the behavior of techno-social systems. Science 325, 425{428, DOI: https: //doi.org/10.1126/science.1171990 (2009).
27. N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, S. Gelly, Parameter-efficient transfer learning for NLP,in International Conference on Machine Learning, PMLR, 2019, pp. 2790–2799
28. S. Ayanouz, B.A. Abdelhakim, M. Benhmed, A smart chatbot architecture based NLP and machine learning for health care assistance, in: Proceedings of the 3rd International Conference on Networking, Information Systems & Security, 2020, pp. 1–6.
29. Möller, T.; Reina, A.; Jayakumar, R.; Pietsch, M. COVID-QA: A Question Answering Dataset for COVID-19. In Proceedings of the ACL 2020Workshop on Natural Language Processing for COVID-19 (NLP-COVID), Seattle, DC, USA, 9 July 2020.
30. Tiago Almeida and Sérgio Matos. "Frugal neural reranking: Evaluation on the COVID-19 literature". In: Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020. Vol. 2. Online: Association for Computational Linguistics, Dec. 2020.