The Review of Contemporary Scientific and Academic Studies
An International Multidisciplinary Online Journal
www.thercsas.com

ISSN: 2583-1380 Vol. 3 | Issue No. 6 | June 2023 Impact Factor: 4.736 (SJIF)

Lessons from COVID: The Use of Leveraging Data Science and Effective Preparation for the Next Pandemic

Shashank Agarwal (shashanka757@gmail.com)
Consumer Value Stores (CVS) Health, Woonsocket, Rhode Island, United States

 Crossref/DOI: https://doi.org/10.55454/rcsas.3.06.2023.009

Abstract: *The present paper discusses how to learn from COVID-19 outbreak can help us take a proactive stance toward epidemic readiness by using state of the art developments in data science and artificial intelligence. It argues the importance of using an integrated data approach along with deployment of machine learning algorithms to facilitate the timely identification, diagnosis and treatment of epidemic diseases. Finally, key challenges such as data security, data privacy and data quality are addressed.*

Keywords: Artificial Intelligence, COVID-19, Machine Learning, Pandemic

Article History: Received: 20 June 2023; Accepted: 25 June 2023; Published/Available Online: 30 June 2023;

Introduction

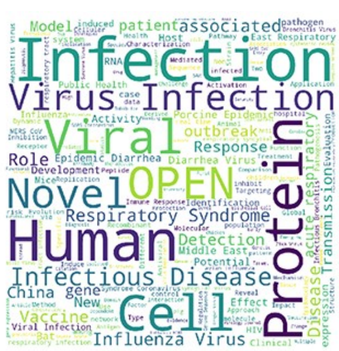
The COVID-19 pandemic has significantly burdened the healthcare infrastructure of the United States, resulting in over 6.15 million hospitalizations and 1.13 million fatalities (as of May 6th, 2023), according to the Centers for Disease Control and Prevention (CDC) (CDC COVID Data Tracker). Epidemiologists posit that the probability of future occurrences of zoonotic diseases akin to Covid-19 is elevated owing to various factors such as climate change, industrialized animal farming practices, wildlife trade, augmented human-animal contact, and rapid advancement in global travel (Sarfraz et al. 2022).

The COVID-19 pandemic has highlighted the importance of investing in digital technology and analytics ecosystems as a proactive approach to epidemic management. Unlike the current reactive approach, this approach can aid in effective response and preparation for future pandemics (Naudé 2020). Data analytics facilitates the detection of patterns, trends, and early warning signals that can assist in early detection and rapid response by utilizing copious amounts of data from diverse sources, including epidemiological records, social media, and healthcare systems (Lünich and Kieslich 2022).

Data Science's Potential Applications

Early Detection: By integrating diverse data sources, such as social media, health records, and environmental data, data scientists can expeditiously devise algorithms and machine learning models to detect potential outbreaks.

The utilization of **natural language processing** (NLP) algorithms to monitor preliminary unofficial news reports, Twitter accounts, Facebook posts, and other sources worldwide, flagging whenever they mention high-priority diseases, such as coronavirus, or more endemic ones, such as HIV or tuberculosis or mention of specific keywords like epidemic virus, infectious disease, outbreak, influenza etc. can help spot early signs of an epidemic disease (such as Covid-19) (Jojoa et al. 2022).

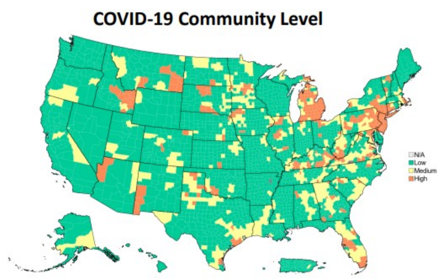


Source: NLP Word-Clout for Covid19 based on COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University

The Review of Contemporary Scientific and Academic Studies
An International Multidisciplinary Online Journal
www.thercsas.com

ISSN: 2583-1380 Vol. 3 | Issue No. 6 | June 2023 Impact Factor: 4.736 (SJIF)

Boosted regression trees, a type of **machine learning** approach, can utilize remote sensing data, including satellite data, vegetation, and weather data, in conjunction with information on the potential locations of animal hosts and early reports of outbreaks to forecast future disease outbreak hotspots and provide early warning signs of epidemics (Ferchichi et al. 2022).



	% of Counties	% of Pop.
Low	67.3%	56.9%
Medium	22.0%	23.4%
High	10.6%	19.7%

Source: CDC Covid-19 Community Level Map

Early Diagnosis: Integrating electronic health records (EHRs) which encompass physician-documented patient visits, clinical notes, laboratory tests and fitness tracker data can facilitate the detection of aberrations in daily routines. Such anomalies may include the irregular cessation of exercise by typically active individuals or the prompt identification of symptoms associated with specific zoonotic diseases, such as Covid-19 (e.g., sore throat, shortness of breath, cough, and fever) (Adulyasak et al. 2023). By leveraging machine learning techniques, early diagnoses of such diseases in patients can be achieved. The utilization of transfer learning, a technique in which an artificial intelligence model that has already been trained to perform a specific task can be rapidly adapted to perform a similar task, has the potential to be highly advantageous in the domain of diagnosing epidemic diseases in patients (Desai et al. 2020). Transfer learning can be utilized in medical imaging analysis, whereby pre-existing models can be fine-tuned by incorporating medical images associated with particular diseases. Utilizing a vast dataset of non-epidemic medical images enables the model to acquire knowledge and proficiency in identifying patterns and features that could signify the initial indications of an epidemic disease (Surianarayanan and Chelliah 2021).

This methodology is valuable when annotated data are scarce for a particular illness outbreak. The model can acquire knowledge from a more extensive spectrum of data through transfer learning, thereby enhancing its capacity to identify and diagnose epidemic diseases during their nascent phases.

Faster Treatment: Data science plays a crucial role in facilitating the development of treatments for zoonotic diseases through the acceleration of drug discovery and the identification of potential drug candidates (Singh and Singhal 2022). There are several ways in which it can contribute:

- Drug repurposing involves utilizing machine learning algorithms to analyze extensive datasets of existing drug information, such as clinical trials, research papers, and drug databases, intending to identify approved drugs that could be repurposed for treating zoonotic diseases. Repurposing existing drugs can decrease both the development timeline and associated costs substantially.
- AI techniques, namely molecular docking and virtual screening can replicate the interactions between target proteins of zoonotic diseases and potential drug candidates. This process is known as virtual screening. This facilitates the recognition of molecules that can bind with the target proteins, thereby impeding their functionality and conceivably serving as a remedy for the ailment.
- The employment of artificial intelligence algorithms, specifically deep learning and generative models can facilitate the production of innovative drug candidates through De Novo drug design. Artificial intelligence models can produce novel molecules with desired properties and refine them for safety and effectiveness by leveraging extensive datasets of chemical structures and their related properties.

ISSN: 2583-1380 Vol. 3 | Issue No. 6 | June 2023 Impact Factor: 4.736 (SJIF)

Challenges: The field of data science exhibits significant potential in its capacity to prepare for future pandemics; however, it is imperative to acknowledge and address the distinct challenges and intricacies accompanying this endeavor.

Data Quality and Accuracy: Artificial intelligence models depend highly on a diverse and high-quality dataset for training. Acquiring exhaustive and dependable information regarding nascent zoonotic illnesses, particularly in the initial phases of a pandemic, can pose a formidable challenge. The accuracy of AI-driven methodologies can be improved by adequate or prejudiced data, leading to accurate predictions (Carnevale et al. 2021).

The issue is further compounded by various factors, including but not limited to the presence of unreliable news sources, the tendency of governmental authorities to downplay the situation, concerns regarding data privacy, and the limited availability of training data.

Data Security and Privacy: The existence of HIPAA privacy regulations in the United States, which restrict the accessibility of healthcare data to the public, coupled with the dearth of contemporaneous data on hospital admissions and diagnoses, presents an added difficulty. Diverse regions across the globe are subject to varying privacy regulations concerning the dissemination of medical information, thereby exacerbating the challenge of charting the epidemic's trajectory across nations (Brotman and Kotloff 2021).

Although current data processing methods, such as differential privacy and collaborative learning, may present a possible resolution, the technology is in its early developmental phase.

Existence of Available Data in Silos: The COVID-19 pandemic has underscored the significance of centralization and integration across diverse data sources, including electronic health records, social media, and environmental data (Tacconelli et al. 2022). Such integration could have facilitated the prompt identification of COVID-19 symptoms and enabled expeditious studies of effective treatments. The utilization of de-identified patient data at a national level, coupled with the implementation of artificial intelligence and machine learning algorithms, can accelerate the detection of patterns that surpass the limitations of individual researchers working with restricted patient samples (Jalali et al. 2022).

The healthcare data currently exists in remote repositories, necessitating legislative action, and technological investment by governmental entities to establish a unified database that can function as a singular source of verifiable information (Megahed and Abdel-Kader 2022).

Conclusion

To summarize, the COVID-19 outbreak has emphasized the significance of taking a proactive stance toward epidemic readiness, and the integration of data science and artificial intelligence can serve as a crucial factor in this regard. By integrating diverse data sources, including social media, health records, and environmental data, data scientists can create algorithms that facilitate the timely identification and diagnosis of epidemic diseases.

In addition, the implementation of artificial intelligence (AI) can aid in creating therapeutic interventions through the repurposing of drugs, virtual screening techniques, and de novo drug design strategies. Notwithstanding, it is imperative to tackle obstacles such as data quality, privacy apprehensions, and interpretability. Through surmounting these obstacles and promoting cooperation, the field of data science has the potential to make a substantial contribution to the readiness for future pandemics, thereby preserving human lives and mitigating the effects on worldwide public health.

Conflict of Interest:

The authors declare neither present nor potential conflict of interest with the study reported here.

Data Availability:

The datasets used and analyzed during the study are available from the corresponding author upon reasonable request.

ISSN: 2583-1380 Vol. 3 | Issue No. 6 | June 2023 Impact Factor: 4.736 (SJIF)

References

Adulyasak Y, Benomar O, Chaouachi A, et al (2023) Using AI to detect panic buying and improve products distribution amid pandemic. AI Soc 1–30. https://doi.org/10.1007/S00146-023-01654-9/TABLES/15

Brotman JJ, Kotloff RM (2021) Providing Outpatient Telehealth Services in the United States: Before and During Coronavirus Disease 2019. Chest 159:1548–1558. https://doi.org/10.1016/J.CHEST.2020.11.020

Carnevale A, Tangari EA, Iannone A, Sartini E (2021) Will Big Data and personalized medicine do the gender dimension justice? AI Soc 38:829–841. https://doi.org/10.1007/S00146-021-01234-9/FIGURES/1

Desai SB, Pareek A, Lungren MP (2020) Deep learning and its role in COVID-19 medical imaging. Intell Med 3–4:100013. https://doi.org/10.1016/J.IBMED.2020.100013

Ferchichi A, Abbes A Ben, Barra V, Farah IR (2022) Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review. Ecol Inform 68:101552. https://doi.org/10.1016/J.ECOINF.2022.101552

Jalali SMJ, Ahmadian M, Ahmadian S, et al (2022) X-ray image based COVID-19 detection using evolutionary deep learning approach. Expert Syst Appl 201:116942. https://doi.org/10.1016/J.ESWA.2022.116942

Jojoa M, Eftekhar P, Nowrouzi-Kia B, Garcia-Zapirain B (2022) Natural language processing analysis applied to COVID-19 open-text opinions using a distilBERT model for sentiment categorization. AI Soc 1–8. https://doi.org/10.1007/S00146-022-01594-W/TABLES/9

Lünich M, Kieslich K (2022) Exploring the roles of trust and social group preference on the legitimacy of algorithmic decision-making vs. human decision-making for allocating COVID-19 vaccinations. AI Soc 1–19. https://doi.org/10.1007/S00146-022-01412-3/TABLES/8

Megahed NA, Abdel-Kader RF (2022) Smart Cities after COVID-19: Building a conceptual framework through a multidisciplinary perspective. Sci African 17:e01374. https://doi.org/10.1016/J.SCIAF.2022.E01374

Naudé W (2020) Artificial intelligence vs COVID-19: limitations, constraints and pitfalls. AI Soc 35:761–765. https://doi.org/10.1007/S00146-020-00978-0/METRICS

Sarfraz M, Mohsin M, Naseem S (2022) A blessing in disguise: new insights on the effect of COVID-19 on the carbon emission, climate change, and sustainable environment. Environ Sci Pollut Res 29:29651–29662. https://doi.org/10.1007/S11356-021-17507-1/TABLES/7

Singh J, Singhal D (2022) Potential antiviral therapies for coronavirus disease 2019 (COVID-19). Data Sci COVID-19 Vol 2 Soc Med Perspect 729–748. https://doi.org/10.1016/B978-0-323-90769-9.00036-0

Surianarayanan C, Chelliah PR (2021) Leveraging Artificial Intelligence (AI) Capabilities for COVID-19 Containment. New Gener Comput 39:717–741. https://doi.org/10.1007/S00354-021-00128-0/FIGURES/7

Tacconelli E, Gorska A, Carrara E, et al (2022) Challenges of data sharing in European Covid-19 projects: A learning opportunity for advancing pandemic preparedness and response. Lancet Reg Heal - Eur 21:100467. https://doi.org/10.1016/J.LANEPE.2022.100467

CDC COVID Data Tracker: Home. https://covid.cdc.gov/covid-data-tracker/#datatracker-home. Accessed 24 May 2023