v. 5, n. 1, 2025

International Journal of Health Science

BIG DATA IN THE PANDEMIC: CLUSTERING, PREDICTIVE MODELS AND COVID-19 CONTROL STRATEGIES

Acceptance date: 06/01/2025

Beatriz Vasconcelos Cipola
Student at the Universidade Nove de Julho, São Paulo/SP, Brazil

Giovana Silveira Dalmas
Student at the Universidade Nove de Julho, São Paulo/SP, Brazil

Isabely Guedes da Silva
Student at the Universidade Nove de Julho, São Paulo/SP, Brazil

Ingrid Lemes Alves
Student at the Universidade Nove de Julho, São Paulo/SP, Brazil

Millena Giulia Silva Barbosa
Student at the Universidade Nove de Julho, São Paulo/SP, Brazil

Letícia Delgado Borges
Professor at the Universidade Nove de Julho, São Paulo/SP, Brazil

Danilo Gagliardi
Teacher at the Universidade Nove de Julho, São Paulo/SP, Brazil

Abstract: Introduction: The COVID-19 pandemic, which began in 2019, has brought substantial challenges to health systems, requiring rapid and effective responses. In this context, the use of Big Data and Artificial Intelligence (AI) technologies has emerged as a promising tool for analyzing and controlling the spread of the virus. This study presents a systematic review on the applications of Big Data and AI in the fight against COVID-19, with a focus on predictive modeling, social mobility and optimization of health resources. **Objectives**: The main objective of this study is to assess the impact and applications of these technologies in the management of the pandemic, identifying their potential and limitations, with a view to proposing a framework to help respond to future public health crises. **Materials and Methods**: Following the PRISMA methodology, a search was carried out in the PubMed and Embase databases to identify relevant studies between 2020 and 2024 using the descriptors "(Big Data) AND (COVID--19[MeSH Terms])" and ('big data'/exp OR 'big data' OR (big AND data)) AND ('covid 19'/exp OR 'covid 19') respectively, with a title filter, resulting in 309 articles. Rayyan[4] *software* was used for *screening*, identifying 233 duplicates, giving a total of 182 articles. Strict inclusion and exclusion criteria were applied, resulting in a selection of 20 articles dealing with everything from modeling viral spread to analyzing data from mobility. **Results**: The analysis showed that predictive modeling with Big Data made it possible to identify patterns in the spread of the virus and predict the demand for health resources, such as hospital beds and ventilators. Studies have also shown that the use of mobility data has been key to understanding the impact of lockdown policies and identifying areas of greater risk. In terms of resource allocation, AI has been essential for optimizing the distribution of medical supplies, contributing to more effective management of available resources. However, the application of these technologies faces ethical and logistical challenges, such as data privacy and the need for a robust technological infrastructure. **Conclusion**: Big Data and AI have played a crucial role in the response to COVID-19, offering valuable insights and helping to formulate effective mitigation strategies. Despite the challenges, these technologies have proven indispensable for managing public health crises, highlighting the importance of continued investment in technological innovation. The lessons learned can guide the development of monitoring and response systems that are better prepared for future health emergencies.

INTRODUCTION

The COVID-19 pandemic, which began in late 2019, has rapidly transformed the global landscape, creating unprecedented challenges for health systems, economies and societies across the planet. From its first cases in Wuhan, China, to its global spread, the disease has revealed flaws and shortcomings in public health infrastructures and highlighted the urgent need for new approaches to pandemic crisis management. With rapid transmission rates and a high impact on health systems, the pandemic required immediate and adaptive responses that often could not be met by traditional infectious disease control methods.

Given this scenario, the use of emerging technologies, such as Big Data and artificial intelligence (AI), has proved to be one of the most promising alternatives for dealing with the crisis. These technologies, which were already being applied in various areas of science and public management, have become essential tools to help understand the pandemic and formulate effective responses. The ability to process large volumes of data in real time and to identify patterns invisible to human analysis has enabled researchers, heal-

International Journal of Health Science ISSN 2764-0159 DOI https://doi.org/10.22533/at.ed.1595125030115

th professionals and public managers to gain valuable insights into the spread of the virus, the demand for hospital resources and the areas most at risk of outbreaks. In short, Big Data and AI have established themselves as central players in the fight against COVID-19, providing a solid basis for strategic decisions at a time of extreme uncertainty (VILLANUSTRE et al., 2021; PHAM et al., 2020).

Drawing on a wide range of data sources, such as health records, mobility information, genomic data and social network monitoring, these technologies have enabled the evolution of the pandemic to be followed in real time and the future of the disease to be projected with unprecedented precision. This has not only helped in the response to the immediate crisis, but has also opened up new perspectives for the management of future public health emergencies, highlighting the importance of technological innovation for global security.

BIG DATA AND AI IN THE FIGHT AGAINST COVID-19

Several studies confirm the effectiveness of using Big Data and AI in the response to the pandemic. For example, Villanustre et al. (2021) demonstrated how the HPCC Systems platform was used to track and model the spread of COVID-19 in different regions of the world. This system enabled a multi-level analysis of the disease, using epidemiological data and causal models to predict the need for hospital resources, such as beds and ventilators, and to support strategic public health decision-making. Big Data analysis was essential for identifying hidden patterns of contagion and predicting future waves of the pandemic, allowing for a more coordinated and efficient response by governments and health systems.

Another significant example is the study by Pham et al. (2020), which carried out a broad review of AI and Big Data applications in the fight against the pandemic. The authors highlighted how these technologies have been used to support the prediction of the spread of the virus, optimize contact tracing efforts and even speed up vaccine development. The use of machine learning algorithms and AI techniques, such as neural networks, has made it possible to create predictive models that have helped to identify factors critical to the spread of the virus and to formulate effective mitigation strategies. In many cases, these tools have helped to reduce the pressure on health systems, allowing resources to be allocated more effectively and minimizing the impact of the pandemic on vulnerable populations.

In addition, He et al. (2020) explored the application of Big Data to assess the environmental impact of the pandemic, analyzing the relationship between air pollution and COVID-19 cases. The study highlighted the importance of analyzing large volumes of environmental data, such as the concentration of pollutants, to better understand how external factors can influence the spread of the virus and affect public health more broadly. Using spatial and temporal data, it was possible to develop models that not only assessed the environmental risks associated with COVID-19, but also provided valuable insights for future pandemic control strategies, highlighting the interconnection between public health and the environment.

MOBILITY AND SOCIAL DYNAMICS DURING THE PANDEMIC

Another relevant aspect of the application of Big Data during the pandemic was the study of mobility and social behavior. With the lockdown measures implemented in various parts of the world, there have been significant changes in the movement patterns of populations, which has directly impacted the spread of the virus. Pan and He (2022) investigated how the pandemic has affected the mobility

International Journal of Health Science ISSN 2764-0159 DOI https://doi.org/10.22533/at.ed.1595125030115

of different social groups in the Greater Bay Area, China, using mobile phone data. The research identified sharp declines in mobility, especially among the most vulnerable groups, such as low-income populations and informal workers. This data was crucial for the formulation of policies to support these populations, such as the distribution of emergency aid and the implementation of social protection measures.

Aragón et al. (2023) also explored the use of Big Data to analyze mobility in times of crisis. Using mobility data from European populations, the authors investigated the relationship between travel patterns and the spread of the virus, identifying local "hotspots" of contagion. The study highlighted the importance of understanding mobility dynamics in order to implement more effective and targeted interventions. Lieberman et al. (2023), in a similar study in South Africa, reinforced the relevance of using Big Data to identify areas at greater risk of contagion and help formulate control strategies that are more tailored to the local characteristics of the pandemic.

These studies demonstrate how analyzing large volumes of data in real time can provide valuable insights for managing public health crises. By mapping mobility and social behaviors, it was possible to identify areas most at risk of contagion and thus target control efforts more effectively. In addition, analyzing mobility data was also essential for understanding how the pandemic has affected different social groups and for formulating public policies to mitigate the economic and social impacts of the crisis, especially among the most vulnerable.

CHALLENGES AND OPPORTUNITIES

Despite the significant contributions made by the use of Big Data and AI during the pandemic, some challenges remain. One of the main ones is the issue of data privacy and security. The collection and processing of large volumes of health, mobility and social behavior data raises important ethical questions about the protection of individuals' privacy. Studies, such as that by Rahman et al. (2021), highlight the need to develop more robust policies and regulations to ensure that data is used responsibly and transparently, without compromising individual rights.

Another challenge is integrating these technologies into traditional healthcare systems. Although Big Data and AI have demonstrated their value during the pandemic, many health systems still face difficulties in implementing these technologies effectively and on a large scale. Lack of adequate technological infrastructure, insufficient training of healthcare professionals and resistance to change are some of the obstacles that need to be overcome if the potential of these technologies is to be fully realized. However, studies such as that by Kiganda and Akcayol (2023) show that, with adequate investment and an integrated approach, it is possible to overcome these challenges and strengthen the capacity of health systems to respond to future crises.

JUSTIFICATION

The COVID-19 pandemic has highlighted the importance of Big Data and AI in managing public health crises. These technologies have not only made it possible to monitor the spread of the virus in real time, but have also provided valuable input for the formulation of public policies to mitigate the impacts of the pandemic. Although there are still challenges to overcome, especially in relation to privacy and the integration of these technologies

into traditional health systems, the benefits of using Big Data and AI in the fight against COVID-19 are undeniable. As we move into a post-pandemic future, it is crucial to continue investing in these technologies and perfecting their applications, ensuring that we are better prepared to face future public health emergencies.

OBJECTIVES

This systematic review seeks to assess the role of these technologies in the response to COVID-19, identifying the most promising applications and the challenges to be overcome. By synthesizing the available evidence, this research will contribute to the development of a framework for the effective use of big data and artificial intelligence in future public health emergencies. The results of this review could inform the creation of early warning systems, the optimization of resource allocation and the improvement of risk communication.

MATERIALS AND METHODS

This systematic review followed the guidelines established by the PRISMA method (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) to guarantee the transparency and reproducibility of the results. PRISMA is widely used for systematic reviews and allows for the rigorous selection and critical analysis of studies, ensuring that only relevant, high-quality articles are included in the analysis process.

SEARCH STRATEGY

The search for articles was conducted in two electronic databases: PubMed and Embase. The search was conducted between March and August 2024, using key terms related to the COVID-19 pandemic, Big Data, artificial intelligence and predictive modeling. Specific terms such as "COVID-19" and "Big Data" were combined with Boolean operators to refine the results, respectively (Big Data) AND (COVID-19[MeSH Terms]) and ('big data'/exp OR 'big data' OR (big AND data)) AND ('covid 19'/exp OR 'covid 19'). The search strategy was adapted as necessary for each database, aiming to cover a comprehensive range of studies related to the use of Big Data and AI in the management of the pandemic.

INCLUSION AND EXCLUSION CRITERIA

The inclusion criteria were:

1. Studies published in English or Portuguese between January 2020 and August 2024;
2. Research that addresses the use of Big Data or AI for monitoring, modeling, predicting or mitigating the COVID-19 pandemic;
3. Articles presenting case analyses, reviews or predictive modeling applied to health systems or public policies.

Studies were excluded if:

1. Whether they were conference abstracts, editorials or short communications;
2. They didn't apply AI or Big Data tools in their analysis;
3. They were not available in full text or accessible for critical review;
4. They do not correspond to Qualis of at least A1, A2, A3, A4, B1, B2 and B3, or if they do not have Qualis, they should have an impact factor greater than 2.0.

SELECTION PROCESS

The selection process followed the phases established by PRISMA: identification, screening, eligibility and inclusion. In the identification phase, 309 articles were extracted from the databases mentioned. A total of 127 duplicates were identified, resulting in 182 studies for the screening phase. In this phase, four independent reviewers assessed the titles and

abstracts to ensure that the articles met the inclusion criteria. In the event of disagreement, a fifth reviewer was consulted. After this process, 131 were excluded.

In the eligibility phase, the 51 articles were evaluated in full text. At this stage, articles that did not present clear methodologies or that failed to demonstrate the applicability of Big Data or AI in the context of COVID-19 were excluded. After this filtering and full reading, 20 articles were included in the final review.

DATA EXTRACTION AND SYNTHESIS

Data from the included studies was extracted and organized into categories such as the use of Big Data, AI tools, epidemiological forecasts and impact on public health. The information was analyzed qualitatively, highlighting the main contributions and challenges reported. The data synthesis considered the variability between approaches and discussed the impacts of these technologies in different geographical and social contexts.

This method ensured that the review included relevant, high-quality studies, providing a critical analysis of the role of Big Data and AI in tackling the COVID-19 pandemic.

RESULTS

Analysis of the 20 articles selected for this systematic review revealed various applications of Big Data and artificial intelligence (AI) in tackling the COVID-19 pandemic. The main areas of study involved predictive modeling of the spread of the virus, analysis of population mobility, optimization of health resources and evaluation of containment policies. In addition, challenges were identified and opportunities in the use of these technologies, with a focus on privacy issues, data infrastructure and the integration of health systems.

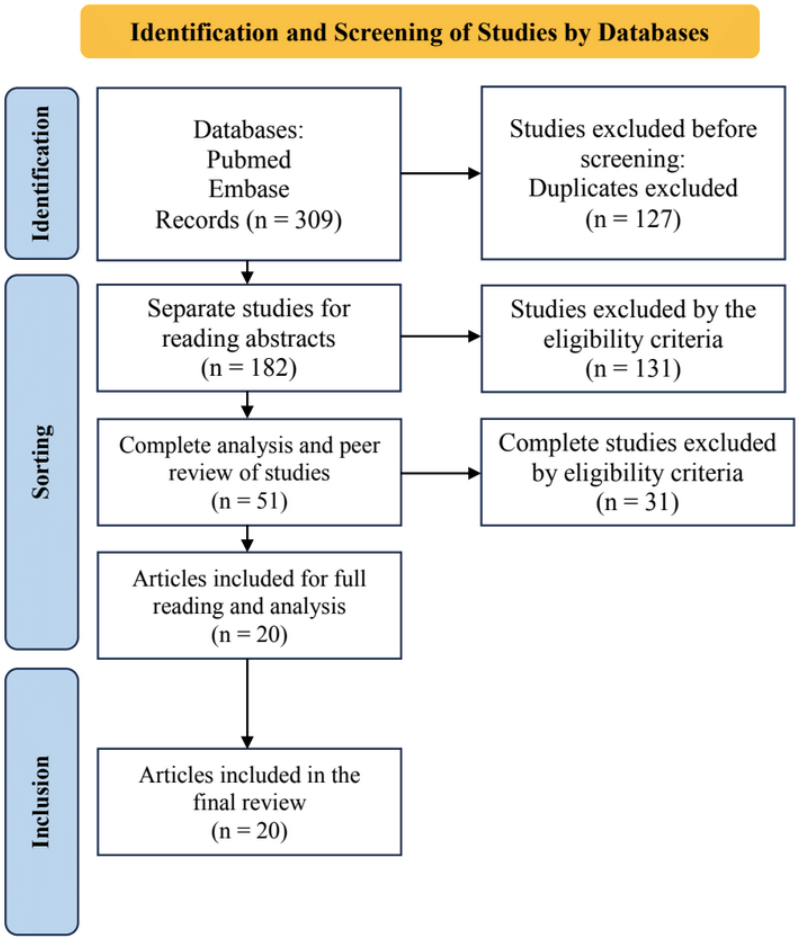
MODELING AND FORECASTING THE SPREAD OF COVID-19

One of the most discussed applications in the literature has been predictive modeling to track the spread of the virus and project infection peaks. Villanustre et al. (2021) presented one of the pioneering studies using the HPCC Systems platform to model COVID-19 cases based on real-time epidemiological data. Their approach used Big Data to predict the evolution of the pandemic and provide crucial insights into the demand for health resources, enabling a more efficient response from the authorities.

Kiganda and Akcayol (2023) also used deep learning techniques to predict the spread of COVID-19, highlighting the efficiency of these tools in modeling large volumes of data. The application of deep learning algorithms allowed for more accurate predictions about the progress of the pandemic in specific regions, enabling better allocation of resources, such as hospital beds and ventilators.

Zhou et al. (2022) demonstrated the effectiveness of predictive modeling by applying spatio-temporal simulation methods to assess the risk of virus transmission in public areas of Guangzhou, China. This study showed that modeling with Big Data can be crucial for targeting more precise local interventions, such as selective lockdowns, rather than generalized policies.

The study by Wu et al. (2020) in China also stood out by using massive data to track the evolution of the virus in real time. Combining public health data with AI technologies allowed authorities to monitor the spread of COVID-19 more effectively, as well as optimizing the response to containment policies.



This flowchart visualizes the path taken by the studies, from their identification in the databases to their selection for the systematic review, guaranteeing the quality and relevance of the articles included in the final analysis.

MOBILITY AND SOCIAL IMPACT ANALYSIS

The analysis of population mobility during the pandemic was another topic widely explored in the studies. Pan and He (2022) analyzed mobility patterns in the Greater China Bay Area using mobile phone data. The study revealed that the mobility restrictions imposed by the government significantly reduced the number of commuters, especially among disadvantaged social groups. The use of Big Data to monitor mobility behavior was essential to understanding how different groups were impacted by the containment measures.

Aragón et al. (2023) complement this analysis by investigating mobility in European populations during the pandemic. Using Big Data data, the authors identified a correlation between mobility levels and local COVID-19 outbreaks. The study highlighted that mobility data allowed public health authorities to implement more targeted policies to contain the spread of the virus, avoiding widespread lockdowns that could have severe economic and social impacts.

Lieberman et al. (2023) analyzed COVID-19 "hotspots" in South Africa using Big Data tools to identify areas most at risk. The study demonstrated that real-time data analysis can quickly identify hotspots, allowing for more agile and effective interventions.

OPTIMIZATION OF HEALTH RESOURCES

The COVID-19 pandemic has put unprecedented pressure on health systems around the world. In this context, Big Data and AI have played a key role in optimizing the allocation of medical and hospital resources. He et al. (2020) used Big Data to assess the health risks associated with exposure to PM2.5 during the pandemic. The data was used to predict the impacts of environmental factors on the evolution of COVID-19, helping to allocate medical resources in areas with a higher risk of respiratory complications.

The study by Wang et al. (2021) looked at the use of spatio-temporal data to monitor the pressure on health systems. Real-time data analysis allowed for the prediction of outbreaks and the consequent allocation of medical resources in regions with the greatest need. This approach was crucial in preventing the collapse of health systems in several regions, especially during peaks of infection.

Wu et al. (2020) also highlighted the use of Big Data to optimize the allocation of healthcare resources in China, allowing the government to respond more effectively to the growing demand for medical equipment and hospital beds. Data analysis helped identify critical areas and prioritize the shipment of medical supplies.

IMPACT AND SUCCESS IN DIFFERENT COUNTRIES

The review also identified different successful strategies in countries such as Saudi Arabia and China. Kuppuswamy et al. (2022) analyzed how Saudi Arabia used AI to monitor the spread of the virus and predict future outbreaks. The use of tools such as Weka Machine Learning helped the Saudi government make strategic decisions, such as closing borders and implementing selective quarantines. This use of AI and Big Data has contributed significantly to controlling the pandemic in the country.

In the case of China, Dong et al. (2021) reviewed the use of AI technologies to track the spread of the virus and monitor compliance with public health policies. The application of these technologies has helped China to control the spread of the virus and implement rapid and effective containment measures.

CHALLENGES AND LIMITATIONS

Despite the advances, studies have also identified challenges related to the use of Big Data and AI during the pandemic. Rahman et al. (2021) highlighted the difficulties in clinical understanding of the virus and the need to improve the application of these technologies in planning for future crises. The study suggests that combining clinical data with AI modeling can increase the ability of health systems to predict and deal with future pandemics.

Another important challenge was the issue of data privacy and security. The large-scale collection of personal data, such as health and mobility information, has raised concerns about the protection of sensitive data. Several studies, including that by Tosi et al. (2020), have highlighted the need to develop stricter privacy policies to ensure that the data collected is used responsibly, respecting individual rights.

FUTURE OPPORTUNITIES

The pandemic has also provided opportunities to improve the use of Big Data and AI in public health. Kuppuswamy et al. (2022) argue that the lessons learned in Saudi Arabia can be applied to future disease outbreaks, with the development of more robust AI-based public health monitoring systems. The study suggests creating integrated data platforms that can provide a coordinated, real-time response to public health crises.

Dong et al. (2021) also point to the growing use of AI in vaccine development and disease diagnosis, predicting that the application of these technologies will become even more common in future health crises.

PREDICTIVE MODELING AND PUBLIC POLICY

Predictive modeling has been one of the most promising applications of Big Data and AI during the pandemic. Tosi et al. (2020) analyzed the use of these tools in Italy and the Lombardy region, showing that predictive models helped to accurately predict the behavior of the pandemic and guide government decisions. These predictive models also served as a basis for implementing public policies, such as lockdowns and public health measures, ensuring that resources were allocated efficiently.

CONCLUSION OF RESULTS

The studies analyzed demonstrate the positive impact of the use of Big Data and AI in the management of the COVID-19 pandemic. From predictive modeling of the spread of the virus to optimization of health resources and population mobility analysis, these technologies have played a crucial role in the global response to the pandemic. However, challenges such as data privacy and technological infrastructure need to be addressed to maximize the benefits of these tools in future public health crises.

The lessons learned during the pandemic, as discussed by several authors, highlight the importance of continuing to invest in the development and application of Big Data and AI technologies. The creation of integrated systems, more robust data protection policies and improved predictive capabilities can transform the way public health responds to future crises. Thus, the use of these technologies offers a promising path towards tackling global emergencies with greater efficiency and precision.

DISCUSSIONS

Based on the results obtained, it is possible to outline a discussion that highlights both the similarities and differences in the approaches and results of the studies analyzed.

DISCUSSION OF COMMON RESULTS

The most common results observed in the literature reveal a strong consensus on the effectiveness of predictive modeling as an essential tool for understanding and combating the spread of the virus. Studies such as those by Villanustre et al. (2021) and Kiganda and Akcayol (2023) show that the use of Big Data and deep learning techniques has provided more accurate predictions about the spread of COVID-19, allowing for better allocation of health resources. This finding is corroborated by several authors, who stress the importance of predictive modeling to guide public health interventions and the allocation of resources in areas at greater risk.

Another area of consensus is the analysis of population mobility, where most studies, such as those by Pan and He (2022) and Aragón et al. (2023), show that monitoring mobility was crucial to understanding how restrictions affected different social groups and to implementing more effective containment policies. Real-time data analysis has made it possible to identify areas most at risk and respond quickly to changes in mobility dynamics.

In addition, the use of Big Data to optimize health resources is a recurring theme. Studies such as those by He et al. (2020) and Wang et al. (2021) show how the analysis of spatio-temporal data has helped to predict outbreaks and allocate medical resources more efficiently, avoiding overloading health systems during the peaks of the pandemic.

Title	Authors	Year	Key words	Main results	Conclusion
Analyzing COVID-19's impact on the travel mobility of various social groups in China's Greater Bay Area via mobile phone big data.	Yu Pan and Sylvia Y. He	2022	Activity space; COVID-19; Greater Bay Area (GBA); Mobile phone data; Social inequity; Travel frequency.	Urban GDP and green areas were associated with a reduction in PM2.5, while roads, cabs, buses and factories increased pollution, with factories being the main factor. The study recommends limiting industrial emissions, promoting clean transportation and encouraging the use of sustainable means of transport.	Discretionary activities have not yet recovered and remain below the pre-epidemic level, and that disadvantaged social groups had limited access to superior preventive measures to avoid infection. The study provides policies related to the recovery of travel mobility.
An architecture for COVID-19 analysis and detection using big data, AI, and data architectures.	Ahmed Mohammed Alghamdi, Waleed A Al Shehri, Almalki, Jameel Najlaa Jannah, Faisal S Alsubaei	2024	Big Data; COVID-19; Twitter; Compliance; Trust	The architecture of the proposed model analyzed different data sources for preliminary diagnosis, detected the affected area and located the abnormalities. In addition, the blockchain approach supported the decentralization of the central repository so that it is accessible to all interested parties. The model proposed in this study described the architecture in four layers. The aim of the proposed architecture is to use the latest technologies to offer a reliable solution during the pandemic; the proposed architecture was sufficient to cover all current problems, including data security. The layers are unique and individually responsible for handling the steps required for data acquisition, storage, analysis and reporting using blockchain principles in a decentralized P2P network.	The proposed architecture uses a four-layer structure to ensure decentralized data acquisition, storage, analysis and security, offering a robust solution to the challenges of the pandemic.
Application of Big Data and Artificial Intelligence in COVID-19 Prevention, Diagnosis, Treatment and Management Decisions in China.	Dong J, Wu H, Zhou D, Li K, Zhang Y, Ji H, Tong Z, Lou S, Liu Z.	2021	Artificial intelligence; Big data; COVID-19; Deep learning; Epidemic prevention and control.	Big Data and AI have contributed to identifying infections, predicting disease progression, developing vaccines and recommending treatment strategies. The study concluded that these technologies play a crucial role in sudden public health events and in crisis management decisions.	The application of big data and AI technology can contribute to prevention, diagnosis, treatment and management decision-making in relation to sudden public health events in the future.
Application of Big Data Technology for COVID-19 Prevention and Control in China: Lessons and Recommendations.	Jun Wu, Jian Wang, Stephen Nicholas, Elizabeth Maitland, and Qiuyan Fan	2020	big data, COVID-19, disease prevention and control	Big data has been essential for tracking people, monitoring the virus and managing medical resources. The study recommends greater integration of data sources, better sharing of information and tackling privacy issues for effective management in future pandemics.	China has used big data technology to prevent and control COVID-19 in a timely manner. To prevent and control infectious diseases, countries must collect, clean and integrate data from a wide variety of sources; use big data technology to analyze a large set of data; create platforms for data analysis and sharing; and address privacy issues in the collection and use of this data

Application of healthcare big data in active case finding of COVID-19 in Yinzhou district of Ningbo. Infectious Diseases of Poverty.	Sun Y.X., Lyu J., Shen P., Zhang J.Y., Lu P., Huang W.Z., Lin H.B., Shui L.M., Li L.M.	2020	Active case finding, big data, COVID-19, Paradigm	The monitoring system with Big Data worked well, but the detection rate of confirmed cases needs to be improved. Epidemiological analysis provided evidence to improve surveillance.	The Yinzhou district successfully implemented an innovative big data-driven COVID-19 case detection model, which combined online screening of suspected cases with offline verification and management. This model, based on data sharing between health and non-health departments and sectors, was key to identifying and controlling sources of infection during the pandemic. In addition, the effort laid a solid foundation for continuous surveillance and early warning of infectious disease outbreaks, highlighting the essential role of big data-based surveillance in preventing and controlling future public health crises.
Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts.	Quoc-Viet Pham, Dinh C. Nguyen, Thien Huynh-The, Won-Joo Hwang and Pubudu N. Pathirana	2020	Artificial intelligence (AI); Big Data; COVID-19; Coronavirus; Epidemic Outbreak; Deep Learning; Data Analytics; Machine Learning;	It has provided researchers and communities with new insights into the ways in which AI and big data improve the COVID-19 situation and has spurred further studies to stop the COVID-19 outbreak.	The article offers a comprehensive overview of the use of AI and Big Data in the response to the pandemic, highlights the applications and challenges, and provides recommendations for improving the control of COVID-19.
Big Data Analytics + Virtual Clinical Semantic Network (vCSN): An Approach to Addressing the Increasing Clinical Nuances and Organ Involvement of COVID-19.	Rahman, Fuad; Meyer, Rick; Kriak, John; Goldblatt, Sidney; Slepian, Marvin J	2021	Big Data; Clinical Analytics	The innovative approach proposed using AI and patient data to predict future outbreaks and enabled better planning and preparation for new waves of infections.	They presented a new approach to modeling COVID-19 using patient data from related diseases, combining clinical understanding with artificial intelligence modeling.
Big data- and artificial intelligence-based hot-spot analysis of COVID-19: Gauteng, South Africa, as a case study.	Benjamin Lieberman,, Jude Dzevela Kong, Roy Gusinow,Ali Asgary,icola Luigi Bragazzi,-Joshua Choma, Salah-Eddine Dahbi, Kentaro Hayashi, Deepak Kar,Mary Kawonga,M-duduzi Mbada, Kgomotso Monnakgotla,James Orbinski, Xifeng RuanFinn Stevensn, Jianhong Wu, and Bruce Mellado	2023	Artificial intelligence; Big data; COVID-19; Control intervention; Gauteng Department of Health; Critical point; Risk-adjusted strategy; South Africa.	The study provided a methodology for identifying clusters of COVID-19 spread, using big data to quantify the severity and progression of hotspots. This data can be used to target specific interventions and control actions during epidemic outbreaks.	Hot-spot analysis is an effective tool for quickly identifying high-risk clusters in outbreaks, helping to adopt public health measures. It complements traditional epidemiological research and supports data--based decisions in epidemic situations, such as vaccination campaigns.

Discussion on early warning, prevention and control of emerging infectious diseases from a macroscopic perspective based on big data and effective distance model: enlightenment of COVID-19 epidemic data in China.	Wang Zhenkun, Chen Zhishui, Du Aihua, Wang Congyi, Liu Hong, Wang Ziwei, Hu Jifa	2020	COVID--19;Effective distance;Population floating;Infectious diseases;Early warning, prevention and control	The time of arrival of the epidemic and the cumulative number of confirmed COVID-19 cases showed a significant linear relationship in China, with significant regression coefficients (P<0.001). At the provincial level, the effective distance explained around 71% of the variation in arrival time and around 90% of the variation in cumulative confirmed cases. At the municipal level, this distance explained approximately 66% of the variation in arrival time and 85% in the magnitude of confirmed cases	Big Data data from LBS and the effective distance model can be used to estimate the trajectory, time and extent of the spread of the epidemic, providing a useful reference for early warning, prevention and control of emerging infectious diseases.
Epidemiological characteristics of COVID-19 monitoring cases in Yinzhou district based on health big data platform.	Sun Y.X., Shen P., Zhang J.Y., Lu P., Chai P.F., Mou H., Huang W.Z., Lin H.B., Shui L.M.	2020	COVID-19, epidemiological characteristics, health big data, monitoring	Implementation of an innovative model for the early detection of COVID-19, combining online screening with offline verification, establishing a basis for continuous surveillance of infectious diseases.	The COVID-19 monitoring system based on the big data health platform was working well, but the detection rate of confirmed cases needs to be improved.
Forecasting the Spread of COVID-19 Using Deep Learning and Big Data Analytics Methods.	Cylas Kiganda and Muhammet Ali Akcayol	2023	Deep learning, COVID-19, Artificial neural networks, Long short-term memory, Autoregressive integrated moving average, Prophet	The LSTM model performed best, identifying Mali, Angola, Egypt, Somalia and Gabon as the most vulnerable countries in the five African regions. The study made forecasts for the next 61 days, helping to manage resources to contain the pandemic.	The study contributes to predicting the spread of COVID-19 on the African continent, using statistical and deep learning models. The LSTM model performed best, identifying Mali, Angola, Egypt, Somalia and Gabon as the countries most vulnerable to an increase in cases. These predictions can guide regional preventive measures and pandemic mitigation policies.
How Data Analytics and Big Data Can Help Scientists in Managing COVID-19 Diffusion: Modeling Study to Predict the COVID-19 Diffusion in Italy and the Lombardy Region.	Tosi D., Campi A.	2020	Big data; COVID-19; Data analysis	The model predicted the spread of the virus and estimated the total number of positive cases, anticipating the end of the first wave in June 2020.	This article has shown that big data and data analysis can help medical experts and epidemiologists to readily design accurate and generalized models to predict the different evolutionary phases of COVID-19 in other countries and regions, and for the second and third possible epidemic waves.
Measuring human mobility in times of trouble: an investigation of the mobility of European populations during COVID-19 using big data.	GUARDABASCIO, Barbara; BROGI, Federico; BENASSI, Federico	2023	Big data - Human mobility - Mobility restriction index - COVID-19 - Exogenous shocks - Fast demography	The methodology identifies clusters of exponential case growth, helping to expose localized transmission hotspots, which can guide the implementation of more precise non-pharmaceutical interventions, based on the specific dynamics of the virus in highly mobile urban areas.	The study highlights the importance of using innovative data sources to understand changes in human mobility, especially in times of crisis. The MRI index proved effective in measuring variations in mobility, sensitive to external factors such as pandemics. Mobility, intrinsic to human nature, can only be restricted for short periods. The results indicate that, in the medium and long term, policies to contain mobility tend to fail. In addition, changes in mobility affect not only population dynamics, but also trade and consumer behavior.

Modeling and tracking Covid-19 cases using Big Data analytics on HPCC system platform.	Flavio Villanustre Arjuna Chala, Roger Dey. Lili Xu, Jesse Shaw LexisNexis, Borko Furht Taghi Khoshgoftaar	2021	Big Data; Covid-19; HPCC system; Modeling Corona spread; SARS-Cov-2; Spreading indicators	Creation of a public website with daily and cumulative indicators, helping to manage the pandemic and control its spread	The HPCC Systems Covid-19 Tracker system provides a detailed analysis at different levels (global to local), with risk indicators and important statistics that can help with decision-making and managing the pandemic. Through this model, it is possible to help contain the disease, provide relevant information to society and contribute to reducing panic. In addition, the project is open source, allowing public access and data transparency.
Risk assessment for precise intervention of COVID-19 epidemic based on available big data and spatio-temporal simulation method: Empirical evidence from different public places in Guangzhou, China.	Shuli Zhou , Suhong Zhou, Zhong Zheng, Junwen Lu, Tie Song	2022	Big data; COVID-19; Gravity mode; Precise intervention measures; Risk assessment; Spatio-temporal spreading process	It identified four spatio-temporal transmission patterns, dependent on urban structure and location. The research can help policymakers carry out rapid risk assessments and implement targeted interventions before epidemic outbreaks.	The integration of agent-based models and SEIR accurately simulated the spread of the epidemic in various public places. The results identified four transmission patterns, influenced by urban structure and location, highlighting the need for region-specific interventions. This approach offers policymakers a tool for rapid risk assessments and targeted preventive measures in future outbreaks.
Spatiotemporal Big Data for PM2.5 Exposure and Health Risk Assessment during COVID-19.	He, H.; Shen, Y.; Jiang, C.; Li, T.; Guo, M.; Yao, L.	2023	spatiotemporal big data; empirical orthogonal function (EOF); geographic weighted regression (GWR); population distribution; COVID-19	The study showed that during the COVID-19 pandemic, the concentration of PM2.5 followed a U-shaped pattern, with an initial drop and an increase throughout the year. There was a significant reduction in pollution in the central and northern regions of China, while the health risks were greater in the center and east, due to the higher population density.	The conclusion of the study highlights that population exposure and health risks related to PM2.5 have decreased during the COVID-19 pandemic, due to mobility restrictions and reduced industrial activity. The big data analysis showed that PM2.5 concentrations are highest in central and northern China, but have decreased in 2020. Factors such as GDP and green areas were associated with a reduction in PM2.5, while roads, cabs, buses and factories increased pollution. The study recommends limiting industrial emissions, promoting clean energy transportation and encouraging the use of sustainable means of transportation to reduce health risks.
The COVID-19: Saudi Government's Comprehensive Guide to Safety Measures Analysis with Big Data.	Dr. Prakash Kuppuswamy, Dr. Saeed Q. Al--Khalidi Al-Maliki, Mohammad Khamruddin, Syed Ameen Saadullaha Hussaini Quadri, Ahamed Ali Shaik Meeran	2022	COVID-19, Pandemic preparedness, Preventive measures, SARS- -CoV-2 Management, Kingdom of Saudi Arabia.	The study highlights Saudi Arabia's efficiency in combating COVID-19, using Artificial Intelligence to analyze and compare its strategies with those of other countries. The analysis uses WHO data and machine learning tools such as Weka to evaluate the different approaches, highlighting the crucial role of the government, health authorities and the population in responding quickly and effectively to the pandemic.	Saudi Arabia has stood out in the fight against COVID-19, and the study explores the use of Artificial Intelligence to analyze its strategies in comparison with other countries, using WHO data. It highlights the role of the government, health authorities and the population in responding to the pandemic.

Transmission and control pressure analysis of the COVID-19 epidemic situation using multisource spatio-temporal big data.	Fangxiong Wang, Ziqian Tan, Zaihui Yu, Siqi Yao, Changfeng Guo	2021	Not available	There is a significant correlation between migration from Wuhan to the Greater Bay Area and the severity of the epidemic. Migration depends on the geographical proximity, relations and economic development of the region of origin, while the severity of the epidemic depends mainly on the volume of migration and the severity of the epidemic in the region of origin. The risk of an epidemic is related not only to the severity of the epidemic in the region of origin, but also to the development of urban traffic and the degree of urban openness. After work resumed, the pressure to prevent and control the epidemic was concentrated mainly in Shenzhen and Guangzhou. The mass migration of the population makes it difficult to control the epidemic effectively.	Studying the relationship between migration volume, epidemic severity and epidemic risk is useful for analyzing types of transmission and predicting epidemic trends.
Trust predicts compliance with COVID-19 containment policies: Evidence from ten countries using big data.	Francesco Sarracino, Talita Greyling, Kelsey J O'Connor, Chiara Peroni, Stephanie Rossouw	2024	Big Data; COVID-19; Twitter; compliance; trust	Trust in people and national institutions predicts compliance with containment policies. The study shows that compliance changes over time, influenced by trust.	The results, obtained through various panel estimation techniques, show that compliance changes over time and that an increase (or decrease) in trust in others predicts an increase (or decrease) in compliance. This evidence reinforces the importance of cultivating trust to promote compliance with policies
Risk assessment for precise intervention of COVID-19 epidemic based on available big data and spatio-temporal simulation method: Empirical evidence from different public places in Guangzhou, China.	Shuli Zhou , Suhong Zhou, Zhong Zheng, Junwen Lu, Tie Song	2022	Big data; COVID-19; Gravity mode; Precise intervention measures; Risk assessment; Spatio-temporal spreading process	It identified four spatio-temporal transmission patterns, dependent on urban structure and location. The research can help policymakers carry out rapid risk assessments and implement targeted interventions before epidemic outbreaks.	The integration of agent-based models and SEIR accurately simulated the spread of the epidemic in various public places. The results identified four transmission patterns, influenced by urban structure and location, highlighting the need for region-specific interventions. This approach offers policymakers a tool for rapid risk assessments and targeted preventive measures in future outbreaks.

Table 1: Results of Studies on the Use of Big Data and Artificial Intelligence in the Management of the COVID-19 Pandemic.

This table summarizes the applications of Big Data and AI on various fronts in the fight against the pandemic, highlighting the main studies and their impact on the management of the public health crisis.

DISCUSSION OF DIFFERENT RESULTS

Despite the similarities, significant differences were also observed in the approaches and contexts of the studies. For example, while most studies emphasize the effectiveness of predictive modeling, the research by Zhou et al. (2022) focused on spatio-temporal simulations specific to public areas, highlighting the need for localized interventions rather than universal approaches. This emphasis on the granularity of interventions highlights a limitation in some studies that apply more generalized models.

Another point of divergence concerns the approach to the challenges of using Big Data and AI. While Rahman et al. (2021) highlight the need for a deeper clinical understanding of the virus, Tosi et al. (2020) emphasize concerns about data privacy and security. The tension between the need for data for effective responses and the protection of individuals' privacy is a theme that has not been equally addressed in all studies, reflecting different priorities and cultural contexts.

Finally, successful strategies in different countries, as observed by Kuppuswamy et al. (2022) in Saudi Arabia and Dong et al. (2021) in China, also illustrate the diversity in Big Data and AI applications. While Saudi Arabia used the technology mainly for monitoring and forecasting outbreaks, China implemented a more comprehensive system that combined tracking infections with overseeing compliance with public health policies. This variation in strategies highlights how the political and social context influences the adoption and success of public health technologies.

CONCLUSION

In summary, the results of the systematic review indicate that while there is widespread recognition of the importance of Big Data and AI in the response to the pandemic, approaches vary significantly in terms of methodologies and applications. The lessons learned and challenges identified provide valuable insights for future research and the implementation of public health technologies, emphasizing the need for a balance between innovation, effectiveness and privacy protection. Strengthening the technological infrastructure and drawing up robust policies can maximize the benefits of these tools in future public health crises.

REFERENCES

ALGHAMDI, A. M. et al. **An architecture for COVID-19 analysis and detection using big data, AI, and data architectures**. PLoS One, v. 19, n. 8, p. e0305483, 2024. doi: 10.1371/journal.pone.0305483. eCollection 2024. Disponível em: https://pubmed.ncbi.nlm.nih.gov/39088543/. Acesso em: 20 ago. 2024.

BENASSI, Federico; BROGI, Federico; GUARDABASCIO, Barbara. **Measuring human mobility in times of trouble: an investigation of the mobility of European populations during COVID-19 using big data.** Quality & Quantity, p. 1-19, 2023. doi: 10.1007/s11135-023-01678-9. Disponível em: https://www.semanticscholar.org/paper/Measuring-human-mobility-in-times-of-trouble%3A-an-of-Guardabascio-Brogi/1e7d5e4f0316e280f1a059f66241bde92bcc0b83. Acesso em: 20 ago. 2024.

DONG, J. et al. **Application of Big Data and Artificial Intelligence in COVID-19 Prevention, Diagnosis, Treatment and Management Decisions in China.** J Med Syst., v. 45, n. 9, p. 84, 2021. doi: 10.1007/s10916-021-01757-0. Disponível em: https://pubmed.ncbi.nlm.nih.gov/34302549/. Acesso em: 20 ago. 2024.

HE, Hongbin et al. **Spatiotemporal Big Data for PM2.5 Exposure and Health Risk Assessment during COVID-19.** International Journal of Environmental Research and Public Health, v. 17, n. 20, p. 7664, 2020. DOI: 10.3390/ijerph17207664. Disponível em: https://www.mdpi.com/1660-4601/17/20/7664. Acesso em: 20 ago. 2024.

KIGANDA, C.; AKCAYOL, M. A. **Forecasting the Spread of COVID-19 Using Deep Learning and Big Data Analytics Methods**. SN Computer Science, v. 4, n. 4, p. 374, 2023. doi: 10.1007/s42979-023-01801-5. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10155670/. Acesso em: 20 ago. 2024.

KUPPUSWAMY, P. et al. **The COVID-19: Saudi Government's Comprehensive Guide to Safety Measures Analysis with Big Data**. Journal of Pharmaceutical Negative Results, 2022. doi: 10.47750/pnr.2022.13.S02.61. Disponível em: https://www.pnrjournal.com/index.php/home/article/view/1520/1285. Acesso em: 20 ago. 2024.

LIEBERMAN, B. et al. **Big data- and artificial intelligence-based hot-spot analysis of COVID-19: Gauteng, South Africa, as a case study.** BMC Medical Informatics and Decision Making, 2023. doi: 10.1186/s12911-023-02098-3. Disponível em: https://pubmed.ncbi.nlm.nih.gov/36703133/. Acesso em: 20 ago. 2024.

PAN, Y.; HE, S. Y. **Analyzing COVID-19's impact on the travel mobility of various social groups in China's Greater Bay Area via mobile phone big data.** Transportation Research Part A: Policy and Practice, maio 2022. doi: 10.1016/j.tra.2022.03.015. Disponível em: https://pubmed.ncbi.nlm.nih.gov/35317198/. Acesso em: 20 ago. 2024.

PHAM, Q.-V. et al. **Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts.** IEEE Access, v. 8, p. 130820-130839, 2020. doi: 10.1109/ACCESS.2020.3009328. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8545324/. Acesso em: 20 ago. 2024.

RAHMAN, F. et al. **Big Data Analytics + Virtual Clinical Semantic Network (vCSN): An Approach to Addressing the Increasing Clinical Nuances and Organ Involvement of COVID-19.** ASAIO Journal, v. 67, n. 1, p. 18-24, 2021. Disponível em: https://pubmed.ncbi.nlm.nih.gov/32796159/. Acesso em: 20 ago. 2024.

SARRACINO, F. et al. **Trust predicts compliance with COVID-19 containment policies: Evidence from ten countries using big data.** Economics & Human Biology, v. 54, p. 101412, 2024. doi: 10.1016/j.ehb.2024.101412. Epub 2024 Jul 20. Disponível em: https://pubmed.ncbi.nlm.nih.gov/39047673/. Acesso em: 20 ago. 2024.

SUN, Y. X. et al. **Application of healthcare big data in active case finding of COVID-19 in Yinzhou district of Ningbo.** Infectious Diseases of Poverty, v. 11, n. 1, p. 1611-1615, 2020. doi: 10.3760/cma.j.cn112338-20200608-00818. Disponível em: https://www.embase.com/records?subaction=viewrecord&id=L632333464. Acesso em: 20 ago. 2024.

SUN, Y. X. et al. **Epidemiological characteristics of COVID-19 monitoring cases in Yinzhou district based on health big data platform.** Infectious Diseases of Poverty, v. 11, n. 1, p. 1220-1224, 2020. DOI: 10.3760/cma.j.cn112338-20200409-00540. Disponível em: https://www.embase.com/records?subaction=viewrecord&id=L632748883. Acesso em: 20 ago. 2024.

TOSI, D. et al. **How Data Analytics and Big Data Can Help Scientists in Managing COVID-19 Diffusion: Modeling Study to Predict the COVID-19 Diffusion in Italy and the Lombardy Region.** JMIR Public Health and Surveillance, 2020. doi: 10.2196/21081. Disponível em: https://www.embase.com/records?subaction=viewrecord&id=L2010026348. Acesso em: 20 ago. 2024.

VILLANUSTRE, F. et al. **Modeling and tracking Covid-19 cases using Big Data analytics on HPCC system platform.** Journal of Big Data, v. 8, n. 1, p. 33, 2021. doi: 10.1186/s40537-021-00423-z. Epub 2021 Feb 15. Disponível em: https://pubmed.ncbi.nlm.nih.gov/33614394/. Acesso em: 20 ago. 2024.

WANG, F. et al. **Transmission and control pressure analysis of the COVID-19 epidemic situation using multisource spatio-temporal big data.** PLoS ONE, v. 16, n. 3: e0249145, 2021. doi: 10.1371/journal.pone.0249145. Disponível em: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0249145. Acesso em: 20 ago. 2024.

WANG, Z. K. et al. **Discussion on early warning, prevention and control of emerging infectious diseases from a macroscopic perspective based on big data and effective distance model: enlightenment of COVID-19 epidemic data in China.** Infectious Diseases of Poverty, v. 1, n. 1, p. 1989-1993, 2020. doi: 10.3760/cma.j.cn112338-20200306-00269. Disponível em: https://www.embase.com/records?subaction=viewrecord&id=L631656616. Acesso em: 20 ago. 2024.

WU, J. et al. **Application of Big Data Technology for COVID-19 Prevention and Control in China: Lessons and Recommendations.** Journal of Medical Internet Research, v. 22, n. 10: e21980, 2020. doi: 10.2196/21980. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7561444/. Acesso em: 20 ago. 2024.

YANG, S. D.; PARK, H. S. **Analysis of the Rate of Confirmed COVID-19 Cases in Seoul and Factors Affecting It Using Big Data Analysis.** Asia-Pacific Journal of Public Health, v. 34, n. 8, p. 824-831, 2022. Disponível em: 20 ago. 2024.