

Use of Big Data Analytics in Understanding the Spread of Infectious Diseases

Mariana Oliveira da Silva¹, Ricardo Mendes Carvalho²

¹ Institute of Public Health, Federal University of Minas Gerais, Belo Horizonte, Brazil

² Department of Epidemiology, University of São Paulo, São Paulo, Brazil

ABSTRACT

The rapidly developing field of big data analytics is the only one that can support the exponential collection, processing, and accumulation of big data in the healthcare industry. Better patient and disease classification, treatment outcome predicting, and enhanced services modelling are just a few of the many applications for which the latter provides the ability to make sense of, understand, and use large data. The research study employed Smart PLS software to create results based on primary data analysis, which included correlation coefficient analysis and descriptive statistics. The use of big data analytics in the study of infectious diseases has resulted in numerous modifications to the information accumulation models. These are studied by comparing the traditional and contemporary theories of data accumulation. The modelling of transmission-aiding infection control policies and plans, as well as the emergency response evaluations required in the case of a national or worldwide outbreak, are rapidly recognizing the critical role that big data analytics plays. In addition, the possible ethical ramifications of extending big data analytics in the near future to include information from mobile phones and wearables as a component of frameworks for infectious diseases are examined. Taking everything into consideration, it is becoming more and more crucial to constructively and transparently incorporate ethical problems in this rapidly evolving profession in order to establish a moral foundation for societal acceptance and the responsible advancement of technological advancements.

Keywords: Big Data Analytics (BDA), Spread (SS), Infections Diseases (ID), Smart PLS Algorithm

INTRODUCTION

In big data analytics, all those methods, gadgets, and things where can be used for collection, processing, and derivation of perception from different sets of data whose volume is very high and whose velocity is also very fast. These sets of data are derived from different sources like we can consider example web, mobile, email, social media, and all those smart devices that are based on the application of networks. In the process in which different customs are uncovered different procedures are unveiled and then connected in the form of various forms of data that are found in raw form to aid in the formulation of all those decisions that are found based on data-informed [1]. Big data analytics is a way that leads towards the reduction of charges and moves towards better development, generates such products that are related to the attention of customers, and provides very good services. Data analytics is a way of giving clues towards all those perceptions that are all about improvement of various functions that are happening in society. Data analytics has the caliber of conversation of all that data that is found in rough and random form into that form that is all about activity in which action can be seen. A cast variety of gadgets, recent technologies, and procedures are utilized to discover recent trends and find the solutions to various problems by finding a collection of data that is based on the solution of all problems [2]. Data analytics is such a useful thing that can give a uniform shape to the business, bring improvements in all decision-making procedures, and provide nourishment to the business. The main purpose of big data analytics is the examination of data that is found in vast amounts. The purpose of the analysis is to unveil all those procedures that are in hidden form. To see all hidden associations and perceptions also come under the category of big data analytics. With the help of recent technology, it is very easy to observe your required data and conclude the solutions to your problems very quickly. It is a struggle that seems very slow and less effective but its working is very high and it can provide all those solutions that are based on conventional business intelligence [3]. Let's try to

understand the term big data analytics with the help of the following example. The attitude of consumers can be tracked and the habits of shopping to provide an apotheosis that is all about recommendations of products related to all those customers which are individuals. Investigation of methods of payment and then comparing them against all the activities which are all about customer activities that happened in the past to recognize the people who are fraud in the recent times [4]. After performing big data analytics successfully, it can guide the workers of healthcare to get all those perceptions which are helpful in the following ways. Like the recognition of all those diseases which are all about the prediction of spreading of infectious diseases. It will prove helpful for the hospitals to do all such preparations that are all about dynamics like the surety of more beds of patients and all that equipment that are all about personal protection related to workers working with the healthcare department. Big data analytics proves very helpful in providing clues to all organizations that are all about healthcare to observe the data of large numbers of patients. In that data, electronic health records are also involved [5]. Certain other things are also included in it which are data about genetics and data of that's wearable device. It is also helpful in the recognition of all symptoms that are all about disease prediction of expected results of about health. All the gadgets that are utilized in big data analytics have the caliber of providing a very authentic representation that is all about its progress. A specific branch of big data is geospatial which utilizes geographic and spatial information to provide perception and the beginning of ideas in mind about the spreading of diseases. In measures that are all about prevention, diagnosis, and detection play a significant role [6]. Various models that prove motivation influencers are arranged in such a way that they can observe the history of patients and represent all abnormalities that prove a representation of all that's stages and conditions that appear at the start of any disease. Predictive analytics is that thing which proves a sign for all healthcare workers to analyze the recent data which is all

about history which gives this permission to all professionals related to healthcare to search for all those opportunities which are all about the decisions that need to be very efficient and effective, can predict the very common customs. Predictive analytics also has the caliber to manage the widespread disease problem. Big data analytics is a field that is evolving in the present time and it is authentic enough that it can accumulate, process and make accurate results in the field of healthcare[7]. It also can understand, rationalize, and usage of big data to provide its service for different objectives. The objectives can be from service modeling in an improved version and the prediction of consequences of treatment[8]. In the field where infectious diseases are dealt with, big data analytics made several variations in the accumulation methods of information. Big data analytics is such a thing that it is becoming a necessary element in the transmission of modeling[9]. These measures provide help in the control of infection and making effective policies. It also deals with all those observations that are all about emergency response which are demanded during outbreaks that happen at local or international level. Hence big data analytics is a very useful technique[10].

Research Objective

The main purpose of this article is to analyze the role of big data analytics in developing an understanding of the spreading of infectious diseases, their symptoms, and the expected outcomes of these diseases.

LITERATURE REVIEW

Researchers claim that a wide range of microorganisms cause infectious diseases in humans. The origin of infectious disease could be either from a direct or indirect source. the severity caused to due infectious diseases leads to pandemic conditions. In some individuals, the symptoms of infections could be mild and are regarded as asymptomatic. while in certain individuals the infection could be serious and lead to life-threatening problems. to tackle the problems related to the spread of infectious diseases the use of AI in the health sector is increasing at a rapid rate[11]. studies suggest that disease dynamics is identified using a disease dynamic model This model uses AI to increase the accuracy of results provided by the model. The dynamic models are improved in the modern technology world using a mathematical data model. The mathematical model efficiently calculated the mathematical values from the given complex data-based information[12] Studies claim that the lives of people were disturbed in the year 2021 when the COVID pandemic took place. covid pandemic had a horrible impact on the lives of people. similarly, the Ebola virus outbreak in African states is one of the major outbreaks of infectious disease[13]. The field of biomedical has advanced as a result of advancements in IoT The data analysis process is carried out in biomedical fields through the use of IoT. also, the approach of data analysis with deep learning algorithms provides a workable solution for overcoming the spread of infectious diseases around the world [14] Also, the treatment process related to patients and the health management process is easily managed using BDA. By analyzing the benefits of using BDA in the health sector the use of this technology-based system has increased in the medical sector [15]. studies suggest that the pandemic of COVID was overcome by following the hygiene protocols. Social distancing was the most important protocol followed by people when the COVID-19 pandemic overtook the world in 2021 [16]. studies claim that overpopulation in the world has increased the

chances of disease spread Any new pandemic that outbreaks spread very quickly because of overpopulation of people. to process of urbanization has resulted in accumulating the overpopulations of world. But instead of benefiting the population urbanization has increased the risk of disease spread [17]. studies predict that information related to any disease condition is transferred to people using an ontology approach. CIDO is a community-based program that works to provide knowledge to the people about the pandemic of covid 19. the process of data analysis, as well as data interaction related to the pandemic, is performed By CIDO to provide information to people to protect themselves from the deadly covid 19 virus[18]. studies have claimed that the increases in the chances of outbreak of pandemic conditions have increased over the last few years. the spread of pathogens in the population in a short period has created severe health problems for people. to test disease people the complex laboratory process is usually time-consuming. to replace these time-consuming processes the use of POCT tests are made.. these tests use the IoT approach to rapidly diagnose the presence of disease-causing pathogens in a person[19]. studies reveal that technological advancement in industries has resulted in the development of Industry 4.0. the health industries have also become Industry 4.0 because of technological advancement in these industries. the healthcare services provided by Industry 4.0 improve the clinical process of providing treatment against diseases[20] Studies explain that the process of early disease detection has become easier through the help of IoT use in medical sectors. the early disease outbreak warning system utilizes block chain technology along with IoT to provide the most accurate and early prediction regarding the sudden outbreak of any pandemic[21]. studies explain that two approaches are commonly used for controlling the spread of any infectious diseases. The first approach is based on the use of modern technology to tackle the problems related to disease onset. the second approach is based on the use of human driven concept for tackling the pandemic condition. maintaining social distancing during COVID 19 pandemic was one of the human driven approach that controlled the spread of virus to a great extent[22] Studies suggest that in public health-related sectors, the combination of two advanced technologies has revolutionized the working of the health sector. NGS and bioinformatics are the two main technological approaches combined to tackle the problems related to the spread of severe infections. The combined technological approaches work to identify the organ, cause, and dynamics of infectious disease prevalent during an outbreak of disease[23]. scholars predict that the decision-making process is a very critical step to control the spread of the virus by providing the public with treatment of the disease. The decision-making process in the medical and health sectors is made using machine learning algorithms. these algorithms are developed to support the facts regarding any medical decision[24]. scholars made brief remarks regarding the spread of virus outbreaks through urban transportation Urban transportation is the reason behind the faster spread of disease conditions. The exposure of the urban population to disease is explained by a mathematical model. The models predict that the disease exposure is related to the dynamics of urban transformation or movement[25]. studies reveal that data related to medical sectors are always very complex. to simplify medical data the data analyst of medical sectors uses modern technology-based systems that convert data into understandable form. the large unstructured data becomes organized and well patterned using modern data analyzing systems. The process of developing preventive medicine against infectious

disease has become easier using advanced data analysts in medical sectors [26]. studies reveal that one big strategy to manage the spread the disease is to analyze the data related to the disease statistic. For this purpose, the data collection process takes place at large scale for massive data collection. this massive data is then analyze and the information taken out form the data helps in tackling the disease spread and handling problems[27]. studies shows that understanding of the dynamic patterns of infectious disease spread is done using mathematical model. The disease propagation trough different population is studied suing the mathematical prediction models[28]. moreover ,the main concern to public health is the sudden outbreak of pandemic. the antimicrobial resistance in humans is caused due to changes in various climatic factors. this change in climatic conditions results in the development of novel pathogen. the novel pathogen becomes resistant to medication and thus results in epidemic and then pandemic condition[29]. Studies suggest that using the AI in health areas often helps these sectors to extreme health crisis related problems during any pandemic related circumstances[30].

Implications

The term “big data analytics” stands for analysis of data at the macro level in which a large amount of data related to some large population is taken to understand any given aspect. As we know the Study of a small amount of data is quite easy but when any aspect related to the whole population has to be studied, so there is a need for a large amount of data for examination and this is called big data analytics. Big data analytics has been used in various fields of life including the study of the spread of infectious diseases. Following are the important implications of the use of Big Data Analytics in Understanding the Spread of Infectious Diseases.

Tracking of Outbreaks of Diseases, Prediction Before Infection, and Identification of High-Risk Areas

As we know the basic reason for every disease is any kind of specific pathogen. The mode of transmission of these pathogens is also particular for causing any specific disease. If the mode of transmission of these pathogens is studied well by using big data analytics, this understanding can be used for tracking the outbreaks of diseases. If these patterns and trends of diseases are explored well, they can help in reducing the risk of infectious diseases. The other important implication of big data analytics for understanding the spread of infectious diseases is that it can help in aspect of prediction of infection. We know that some diseases spread in the form of infection in the whole community. If prediction is done timely by using Big Data Analytics, it can be used to prevent such infections in any population. By using a few machine algorithms, prediction can be easily done. After the prediction of infectious diseases, the next step is formulating policies to take preventive measures to minimize the risk of the onset of disease.

The other important implication of big data analytics for understanding the spread of infectious diseases is that it can help in the identification of those areas that are at high risk of getting infection. We all know that some particular conditions give rise to the optimum growth of these pathogens in the body. Some specific areas provide suitable conditions for the spread of infectious diseases. The information related to different areas is usually collected by using big data analytics then the analysis of this data is done to know about those

areas that are at high risk of getting infectious diseases. So, there are some strict preventive measures are taken in these areas which will help in reducing the risk of spread of infection in other areas as well. The migration should be minimal from infected areas because human beings may act as vectors for spreading infectious diseases from one place to another. All of these aspects are only possible because of Big Data Analytics used in understanding the spread of contagious diseases. So, all of these aspects may be considered as implications of using big data analytics.

Monitoring of Dynamics of Transmission of Pathogens, Spreading Awareness to People

As we have discussed earlier some specific modes of transmission of pathogens are decisive factors for the spread of infectious diseases. Big Data analytics can be used to thoroughly study these dynamics associated with pathogen spread. It is important to closely monitor and manage the variables that have the potential to transmit infectious diseases. Gaining control over pathogen-related transmission dynamics will also enable us to regulate the spread of infectious diseases. Raising public awareness is one of the big data analytics' other key implications for comprehending the transmission of infectious diseases. As is well known, infectious disease cases have had a greater impact on the general populace. Additionally, it is well known that educating the public about pathogen transmission mechanisms can effectively reduce the spread of infectious diseases. It is essential to be aware of any specific disease since knowledge contributes to disease prevention. It is imperative that educational tools and approaches be developed so that every member of a community is aware about infections, their modes of transmission, consequences, and other related information. The introduction of social media has made raising awareness a relatively simple task these days. We can raise awareness using social media platforms so that everyone in the population can hear about it. We can raise public awareness about the spread of infectious diseases by using social media to disseminate information about it, thanks to the assistance of big data analytics. These consequences of big data analytics demonstrate how crucial big data analytics is to comprehending the transmission of infectious illnesses.

Genomic Analysis of Pathogen, Environmental Analysis, and Sharing of Data

The pathogenicity of any pathogen is related to the specific genome of that pathogen. This is because the receptive site in the host is based on the genome of the pathogen. The exact site of the pathogen bounds to the particular receptive site of the host depending upon the genomes of pathogens.

The other aspect is that we have seen that medication resistance has been increasing day by day in pathogens. Because of this resistance, the infections caused by these pathogens are difficult to treat. All of these aspects are because of the changing genome of pathogens day by day. If the correct genome of the pathogen can be studied and analyzed with the help of Big Data Analytics, it can be helpful in prescribing medication or treatment for that pathogen. In this way, we can say that big data analytics is helpful in genomic Analysis of pathogens as well. As we have enumerated already there are a variety of conditions in the environment that lead to the spread of infectious diseases such as temperature, humidity, air, pressure, and others. A better analysis of these environmental conditions can also be done by using Big Data Analytics. All the data related to infectious

diseases can also be shared easily by using big data analytics. All of these implications of the use of big data analytics for understanding

the spread of contagious diseases convince us that big data analytics may prove much more effective shortly.

Descriptive Statistic

Table 1: Result of Descriptive Statistic

Name	No.	Mean	Median	Scale min	Scale max	Standard deviation	Excess kurtosis	Skewness	Cramér-von Mises p value
ID1	1	1.520	1.000	1.000	3.000	0.608	-0.366	0.747	0.000
ID2	2	1.500	1.000	1.000	3.000	0.574	-0.523	0.653	0.000
ID3	3	1.480	1.000	1.000	3.000	0.574	-0.414	0.735	0.000
ID4	4	1.460	1.000	1.000	3.000	0.573	-0.283	0.819	0.000
ID5	5	1.440	1.000	1.000	3.000	0.637	0.330	1.188	0.000
ID6	6	1.560	2.000	1.000	3.000	0.605	-0.522	0.599	0.000
ID7	7	1.620	2.000	1.000	3.000	0.629	-0.590	0.522	0.000
ID8	8	1.500	1.000	1.000	3.000	0.671	-0.116	1.025	0.000
ID9	9	1.580	2.000	1.000	3.000	0.635	-0.507	0.654	0.000
ID10	10	1.520	1.000	1.000	3.000	0.608	-0.366	0.747	0.000

As can be seen from the above results of table 1, in a descriptive statistical study, the skewness value of each indicator, which included both dependent and independent variables, reflected the mean values, the median rate, and the standard deviation. The ID1, 2, 3, 5, 6, 7, 8, 9, and 10 criteria are used to determine big data analytics to understand the spread of infectious illnesses. For any element, the average values 1.520, 1.500, 1.480, 1.440, 1.560, 1.620, and 1.520 are all considered positive average values. With a minimum value of 1.000 and a maximum value of 3.000, the total result shows that the standard deviation rates are 60%, 57%, 63%, 60%, and 67% of the mean values. Every indicator has a 100% significant level, as shown by Table 1's overall probability value of 0.000. Concerning big data analytics and the understanding of infectious disease transmission, the table-1 shows the favorable skewness rates of each variable at 74%, 65%, 73%, 81%, and 65%, respectively. Big data collected from social media, the internet, electronic health records, and other digital sources could provide information about infectious disease outbreaks and hazards faster and more thoroughly than traditional inquiry techniques. But some obstacles must be surmounted.

The gold standard for infectious disease surveillance is regarded as traditional methods, which are usually based on laboratory testing and other epidemiological data obtained by public health groups. The authors do point out that it typically lacks the spatial accuracy required for accurate monitoring, can have temporal gaps, and is costly to construct. Furthermore, in low-income nations, the cost could be very high. On the other hand, real-time large data streams derived from web searches may follow disease activity locally, but they also come with inherent biases. The experts speculate that hybrid technologies—

which blend traditional monitoring with enormous data sets—might prove to be a better approach in the future, complementing current techniques rather than replacing them.

Ensuring Data Policy

The authors claim that although big data presents a "tantalizing opportunity" to enhance public health surveillance, its application in this domain lags behind other domains like climatology and marketing by decades. For example, anonymized electronic health data can be used to track the progression of infectious diseases, the uptake of vaccines, and the occurrence of adverse drug responses. The authors claim that one reason for the sluggish adoption of the data for surveillance has been ethical worries over patient privacy. Furthermore, there are few scholarly studies showing how this kind of data compares to more conventional surveillance techniques.

Recognizing Negative Drug Reactions

In addition to improving infectious disease surveillance, nontraditional data streams from the internet and social media have the ability to supplement current systems for reporting adverse drug reactions (ADRs). Although consumers hardly ever utilize the official ADR reporting systems, they do look up information about pharmaceuticals online and post about potential side effects on social media and in online health forums. Although there are obstacles, mining and analyzing social media posts and internet search data may be able to identify ADR signs faster than conventional physician-based reporting systems. A prevalent ethical inquiry about the utilization of these unconventional sources is the appropriateness of making privately owned data available for public health investigations.

Correlation Coefficient Analysis

Table 2: Result of Correlation Coefficient Analysis

	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10
ID1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ID2	-0.057	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ID3	-0.085	-0.182	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ID4	-0.055	0.213	-0.002	1.000	0.000	0.000	0.000	0.000	0.000	0.000
ID5	0.235	-0.055	-0.031	0.048	1.000	0.000	0.000	0.000	0.000	0.000
ID6	-0.030	-0.000	-0.083	0.007	0.035	1.000	0.000	0.000	0.000	0.000
ID7	0.203	-0.028	-0.049	-0.014	0.118	-0.071	1.000	0.000	0.000	0.000
ID8	0.049	0.182	0.000	0.078	0.094	-0.148	0.119	1.000	0.000	0.000
ID9	-0.056	0.192	-0.050	-0.074	-0.038	-0.116	0.051	-0.023	1.000	0.000
ID10	0.242	0.057	-0.085	0.060	-0.126	-0.030	0.046	0.049	0.099	1.000

The above results of table 2 describes the correlation coefficient analysis result shows that some positive also that some negative link of

each variable related to the big data analytics in understanding the spread of infectious diseases.

Inadequate Analytical and Digital Abilities

Using big data analytics tools to detect, report on, contain, and recover from emerging infectious disease epidemics requires a strong foundation in digital skills and expertise. The inability of the surveillance systems in the SSA healthcare systems to evaluate, interpret, and report surveillance data due to a lack of digital skills and competence has been noted as one of its limitations. The adoption of advanced digital health technologies is expected to be hampered by the requirement to retrain and upskill healthcare personnel in big data analytics and digital health technologies in order to deal with the goal of detecting and reporting infectious disease surveillance data.

Smart PLS Algorithm Model

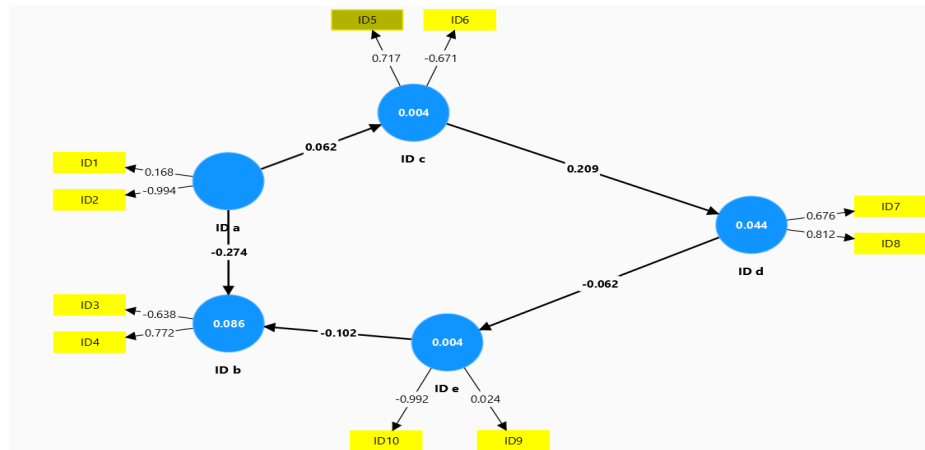


FIGURE 1: Smart PLS Algorithm Model

The model of figure 1 previously mentioned illustrates the intelligent PLS Algorithm model that measures big data analytics to comprehend the transmission of infectious diseases. The ID b indicates a negative yet significant correlation between IDa and ID b of -0.274.

Consequently, in order to utilize big data analytics and ensure that the surveillance systems in the healthcare systems managed by the SSA are current with global best practices, staff training is crucial. These three main contextual restrictions impede the development of an integrated disease surveillance and response system in Sub-Saharan Africa (SSA), and eventually throughout Africa. Therefore, comprehensive programs that address these and other difficulties that SSA healthcare systems face must be put into place. This is the case if the operationalization of the IDSR strategy successfully accomplishes the goal of establishing an evidence-based system for public health surveillance and monitoring.

The ID a indicates that there is a significant level of 99% and 2% between them, or -0.992 and 0.024. Similarly, the rate result for ID d describes a 67% and 81% positive and significant link between them, whereas ID c shows a 20% positive link.

Significant Analysis

Table 3: Result of Significant Analysis

Matrix	Original sample (O)	Sample Mean (M)	Standard Deviation (SD)	T-Statistic	P-Values
ID1<-ID a	0.112	0.411	0.521	0.215	0.083
ID10<-ID e	-1.005	0.442	0.546	1.840	0.066
ID2<-ID a	-0.987	0.324	0.650	1.518	0.129
ID3<-ID b	-0.636	0.388	0.565	1.125	0.0261
ID4<-ID b	0.770	0.230	0.671	1.148	0.215
ID5<-ID c	0.742	0.465	0.623	1.192	0.233
ID6<-ID c	-0.697	0.215	0.599	1.164	0.244
ID7<-ID d	0.588	0.471	0.505	1.166	0.244
ID8<-ID d	0.742	0.517	0.462	1.605	0.109
ID9<-ID e	0.124	0.454	0.528	0.235	0.814

The results of above table 3 show that the original sample value, sample mean value, standard deviation rates, T statistic level, and P value of each matrix are all represented in the significant analysis result. ID1~IDa is the first matrix. It demonstrates that the first sample value was 0.112. With a standard deviation rate of 0.521, the mean is 52% off. T statistic level, according to the outcome, is 0.215. The current positive P value indicates an 8% significant level between them. Similarly, the ID 10<-ID e reveals a 1.840 relationship between them that is now positive and has a 6% significant relationship, with a significant level of 0.066. There is a 1.166 and 24% significant relationship between them, according to ID7-ID d. In addition, the IDs 9 and e show a 23% positive and an 81% significant relationship with

big data analytics in determining the transmission of infectious diseases.

Conclusion

After examining the implications of big data analytics for understanding the transmission of infectious diseases, we may draw some conclusions about how helpful it is to the field of pathology. Even though infectious diseases can currently be identified and tracked early in Africa's healthcare systems, a thorough analysis of the literature revealed that significant advancements in readiness and response action—prevention, prediction, and control, in particular—remain necessary. A multi-stakeholder approach is required for capacity

building initiatives. Governments must create big data regulations, academic institutions must carry out extensive research, and IT professionals must create user-centered applications and training materials that are specifically adapted for the African environment. The research study based on primary data to determine the research used smart PLS software and generate result overall research concluded that direct significant link between them. A possible limitation of the research is that relevant literature published in non-

open access journals may offer further context for the results made about the application of big data analytics in African healthcare systems. Some publications were inaccessible by the writers because of outside factors, like limitations on online resources and subscription journals. The study's findings present an opportunity for academics to expand their inquiry by giving them unique information on the surveillance of infectious diseases from many countries, particularly those in Africa

REFERENCES

- [1] J. B. Awotunde, R. G. Jimoh, I. D. Oladipo, M. Abdulraheem, T. B. Jimoh, and G. J. Ajamu, "Big data and data analytics for an enhanced COVID-19 epidemic management," in *Artificial Intelligence for COVID-19*. Springer, 2021, pp. 11-29.
- [2] Z. Li *et al.*, "Monitoring the spatial spread of COVID-19 and effectiveness of control measures through human movement data: proposal for a predictive model using big data analytics," *JMIR Research Protocols*, vol. 9, no. 12, p. e24432, 2020.
- [3] S. Pyne, A. K. S. Vullikanti, and M. V. Marathe, "Big data applications in health sciences and epidemiology," in *Handbook of statistics*, vol. 33: Elsevier, 2015, pp. 171-202.
- [4] Q. Jia, Y. Guo, G. Wang, and S. J. Barnes, "Big data analytics in the fight against major public health incidents (Including COVID-19): a conceptual framework," *International journal of environmental research and public health*, vol. 17, no. 17, p. 6161, 2020.
- [5] S. J. Alsunaidi *et al.*, "Applications of big data analytics to control COVID-19 pandemic," *Sensors*, vol. 21, no. 7, p. 2282, 2021.
- [6] P. M. Kasson, "Infectious disease research in the era of big data," *Annual Review of Biomedical Data Science*, vol. 3, no. 1, pp. 43-59, 2020.
- [7] S. Verma and R. K. Gazara, "Big data analytics for understanding and fighting COVID-19," *Computational intelligence methods in COVID-19: Surveillance, prevention, prediction and diagnosis*, pp. 333-348, 2021.
- [8] S. Bansal, G. Chowell, L. Simonsen, A. Vespignani, and C. Viboud, "Big data for infectious disease surveillance and modeling," *The Journal of infectious diseases*, vol. 214, no. suppl_4, pp. S375-S379, 2016.
- [9] L. Judijanto, H. Hermansyah, K. P. Ningsih, D. Anurogo, and M. Firdaus, "The Role of Big Data Technology in Predicting and Managing the Spread of Infectious Diseases," *Journal of World Future Medicine, Health and Nursing*, vol. 2, no. 2, pp. 219-230, 2024.
- [10] C. Garattini, J. Raffle, D. N. Aisyah, F. Sartain, and Z. Kozlakidis, "Big data analytics, infectious diseases and associated ethical impacts," *Philosophy & technology*, vol. 32, pp. 69-85, 2019.
- [11] S. Agrebi and A. Larbi, "Use of artificial intelligence in infectious diseases," in *Artificial intelligence in precision health*. Elsevier, 2020, pp. 415-438.
- [12] A. Alahmadi *et al.*, "Influencing public health policy with data-informed mathematical models of infectious diseases: Recent developments and new challenges," *Epidemics*, vol. 32, p. 100393, 2020.
- [13] R. E. Baker *et al.*, "Infectious disease in an era of global change," *Nature Reviews Microbiology*, vol. 20, no. 4, pp. 193-205, 2022.
- [14] A. Banerjee, C. Chakraborty, A. Kumar, and D. Biswas, "Emerging trends in IoT and big data analytics for biomedical and health care technologies," *Handbook of data science approaches for biomedical engineering*, pp. 121-152, 2020.
- [15] K. Batko and A. Ślęzak, "The use of Big Data Analytics in healthcare," *Journal of big Data*, vol. 9, no. 1, p. 3, 2022.
- [16] N.-C. Chiu *et al.*, "Impact of wearing masks, hand hygiene, and social distancing on influenza, enterovirus, and all-cause pneumonia during the coronavirus pandemic: retrospective national epidemiological surveillance study," *Journal of medical Internet research*, vol. 22, no. 8, p. e21257, 2020.
- [17] C. Connolly, R. Keil, and S. H. Ali, "Extended urbanisation and the spatialities of infectious disease: Demographic change, infrastructure and governance," *Urban studies*, vol. 58, no. 2, pp. 245-263, 2021.
- [18] Y. He *et al.*, "CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis," *Scientific data*, vol. 7, no. 1, p. 181, 2020.
- [19] S. Jain *et al.*, "Internet of medical things (IoMT)-integrated biosensors for point-of-care testing of infectious diseases," *Biosensors and Bioelectronics*, vol. 179, p. 113074, 2021.
- [20] M. Karatas, L. Eriskin, M. Deveci, D. Pamucar, and H. Garg, "Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives," *Expert Systems with Applications*, vol. 200, p. 116912, 2022.
- [21] E. Khodadadi and S. Towfek, "Internet of Things Enabled Disease Outbreak Detection: A Predictive Modeling System," *Journal of Intelligent Systems & Internet of Things*, vol. 10, no. 1, 2023.
- [22] R. K. R. Kummitha, "Smart technologies for fighting pandemics: The techno-and human-driven approaches in controlling the virus transmission," *Government Information Quarterly*, vol. 37, no. 3, p. 101481, 2020.
- [23] I. Maljkovic Berry *et al.*, "Next generation sequencing and bioinformatics methodologies for infectious disease research and public health: approaches, applications, and considerations for development of laboratory capacity," *The Journal of infectious diseases*, vol. 221, no. Supplement_3, pp. S292-S307, 2020.
- [24] N. Peiffer-Smadja *et al.*, "Machine learning for clinical decision support in infectious diseases: a narrative review of current applications," *Clinical Microbiology and Infection*, vol. 26, no. 5, pp. 584-595, 2020.
- [25] X. Qian and S. V. Ukkusuri, "Connecting urban transportation systems with the spread of infectious diseases: A Trans-SEIR modeling approach," *Transportation Research Part B: Methodological*, vol. 145, pp. 185-211, 2021.
- [26] M. I. Razzak, M. Imran, and G. Xu, "Big data analytics for preventive medicine," *Neural Computing and Applications*, vol. 32, no. 9, pp. 4417-4451, 2020.
- [27] S. Ribeiro-Navarrete, J. R. Saura, and D. Palacios-Marqués, "Towards a new era of mass data collection: Assessing pandemic surveillance technologies to preserve user privacy," *Technological Forecasting and Social Change*, vol. 167, p. 120681, 2021.
- [28] R. Sameni, "Mathematical modeling of epidemic diseases; a case study of the COVID-19 coronavirus," *arXiv preprint arXiv:2003.11371*, 2020.
- [29] N. Sims and B. Kasprzyk-Hordern, "Future perspectives of wastewater-based epidemiology: monitoring infectious disease spread and resistance to the community level," *Environment international*, vol. 139, p. 105689, 2020.
- [30] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial Intelligence (AI) applications for COVID-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 337-339, 2020.



This is an open-access article distributed under the terms of the Creative Commons Attribution License.