
The Impact of Big Data Technology on Predicting and Controlling Infectious Disease Transmission

Sura R. Alatba

Al-Nisour University College, Iraq
surar@nuc.edu.iq

ABSTRACT

For the past few years, the Big Data technologies has emerged as a significant issue in medical research due to its capacity to rapidly gather, store, and analyze massive amounts of diverse data. Innovative possibilities for enhancing the forecasting and control of the dissemination of infectious illnesses have recently become available as an outcome of research. The research study helps to evaluate the impact of Big Data plays in forecasting and managing the global dissemination of infectious illnesses related to disease. The research idea is particularly interested in locating efficient strategies for utilizing big data impacts to forecast disease transmission trends and coordinate countermeasures to these variations. Hence, the research proposed a framework called Prediction Analysis for Infectious Disease using BigData (PAID-BD) for easy identification during pandemic illness. The aim of this investigation the application of big data in forecasting and controlling the spread of infectious diseases, with a particular emphasis on improving existing efforts to prevent and avoid epidemics. The construction of forecasting systems allows for the anticipation of epidemics of transmission diseases and the improved effectiveness of intervention health measures. These models are developed using various health information sources and cutting-edge significant data analytics procedures. A strong emphasis is placed on the prompt adoption of responsive measures and the adaptation of plans to changing epidemiological conditions.

Keywords: Big data platforms; epidemics outbreak; transmission disease prediction; Prediction Analysis; infectious disease.

1. Introduction

The technological revolution is driven by Big Data solutions that leverage AI to simplify and minimize the mental demands of obtaining and analyzing enormous amounts of data. Thus, AI and big data may have been used to battle the pandemic since they enabled analytics related to big data and produced incredible outcomes in fighting infectious diseases worldwide [1]. AI and big data could have been utilized under the circumstance of outbreak. This is because both technologies enabled Analytics (BDA and produced tangible outcomes in the battle against infectious illnesses worldwide [2]. AI methodologies, procedures, collections of data, and evaluation processes aid users in making informed decisions and handling situations in disease prediction. Considering the many available resources, AI still has to improve in some situations. Combining AI and big data is growing in popularity to enhance the capacity for adaptation and the potential of both systems in forecasting and controlling the transmission of infectious diseases [3].

From following the COVID outbreak around the country to acquiring alternative critical medications to improving immunization distribution networks, analysis of big data and intelligence are being employed [4]. By utilizing the information, which encompasses the history of illness, daily routines, and genetic data, medical professionals can build therapeutic treatment methods. Healthcare organizations employ big data tactics to increase effectiveness and analyze healthcare knowledge acquisition utilizing extensive data case studies and analytics [6]. The epidemic changed global lifestyles, and learning is among the most affected attributes by pandemic protocols, intriguing illnesses, and social estrangement [7]. Managing enormous volumes of data can be inefficient; establishing Data analytics integrated with an organization model reduces the number of inaccuracies that occur in healthcare due to con In addition to internet-based studies, big-data resources for testing and pre-diagnostics in both state and private medical facilities enable new pandemic detection and prediction methodologies [9].

This study aims to determine how big data analytics [10] can help during pandemics and epidemics when data are collected in enormous amounts from numerous sources and have characteristics that are challenging to relate without proper analysis. Monitoring platforms can identify and regulate transmissible disease outbreaks, help manage restricted assets, and generate skills, strategies for action, and resource allocation to avoid repeat infections [11]. AI-enhanced big data tracking systems can increase pattern identification and generate probabilistic illness outbreak estimates by analyzing vast amounts of information combining individuals, infectious agents, vectors generated and environmental elements [12].

The research contributes the following ideas:

- Enhance epidemic readiness by leveraging big data technology, such as PAID-BD, to develop prediction models to anticipate infectious illness.
- To identify the areas with the greatest need for intervention, allowing for strategic resource allocation to minimize disease transmission and provide timely medical care.
- To empower evidence-based decision-making in epidemic response efforts by providing actionable insights derived from big data analytics and more effective epidemic control.

An outline of the research paper is given below. In Section 2, the recent publications using various big data concepts related to infectious illness transmitted are forecasted thoroughly. The implemented big data ideas, research contributions and controlling factors with data preprocessing steps are detailed in Section 3. Forecasting infectious illness results with prediction and control performance metrics is covered in Section 5, which explores the summarized key findings and scope for future work.

2. Research Works

The application of big data techniques was suggested by Judijanto et al. [13] to the management and prediction of infectious illness transmission. Big Data platforms provide a thorough understanding of the elements that contribute to the spread of disease by combining data from multiple sources, including medical facilities, socioeconomic, and behavioural information. To locate patterns, important discoveries, and gaps in the body of literature, the study uses qualitative methodologies such as reviewing related works, studies of cases, and discussions with health professionals. The research seeks knowledge of applying technological advances in big data related to healthcare incident forecasting and control i. The possible influence in the selection of instance reports and healthcare experts questioned could be one research drawback, affecting how broadly applicable the conclusions are.

The use of artificial intelligence (AI) and big data in the fight against the pandemic was examined by Sood et al. [14], highlighting the importance of these resources for monitoring, controlling, and treating the disease. Technology can provide substantial assistance with the initial identification, evaluation, and action to prevent the propagation of the virus and lower death rates. Although they exhibit potential in combating the pandemic, more studies are needed to develop generally recognized protocols and recommendations(rec) for their efficient application after the present epidemic. Consistent reports and persistent research and development efforts are essential to optimize the potential advantages of Big Data and AI in avoiding and controlling epidemics. Challenges in pandemic analysis and assessment include limited data sources, sluggish gathering, lack of governance, and low data accuracy.

Social and infectious illness systems have challenged traditional modelling methods after the pandemic was discussed by Shi, Honghao, et al. [15]. The objective of this study helps to upgrad the compartment layout of the model employed with agents and models based on AI to handle epidemiology challenges. The researchers apply multi-model, compartment separation, adaptive variables, agent-based simulations, and AI-based modelling tools. The researchers examined the mobility of humans, age groups, healthcare resources, responses from humans, and vaccination. Modelling approaches were used to objectively examine socioeconomic networks' effects and make recommendations for spreading infectious diseases and management.

Wang et al. [16] carry out a preliminary investigation on the construction of a big data control and prevention model for health-related emergencies by employing the grounded theory, which is a method of qualitative study, alongside information available, regulations, and guidelines serving as research samples, and then conducts a theoretical analysis utilizing three-level coding and saturated state test. The proposed model is a framework that determines the variations regarding data demands of various populations throughout a pandemic and highlights many creative methods to enhance the sharing of resources and joint management.

Amusa et al. [17] developed a study related to infection epidemiological research; the purpose of the present research was to find hotspots of massive data and identify studies that have been done previously. According to the findings of the bibliometric analysis, the most frequently exploited big data sources for monitoring infectious illnesses or modelling are internet searches and transactions on social platforms. As a result of the investigation, the study was identified as the primary player in this research field. The major study themes highlighted are epidemic tracking and surveillance, the practical value of medical records, the methodology foundation for medical big data tools, and machine learning and deep learning. Infectious illness monitoring and control have been improved by introducing novel significant data streams, generating exciting opportunities.

Torres Munguía et al. [18] demonstrated an updated database of infectious disease outbreaks compiled from the News and the Covid Dashboard that the World Health Organization developed. The study includes details about 50+ transmissible illnesses and various incidents related to health in 233 nations and provinces around the world between 1996 and March 2022. Timespan. Using this analysis method, planners can determine the locations most likely to be affected by epidemics of transmissible diseases. Taking into mind the level of willingness and risk of these zones, they can establish big data measures that have the potential to help restrict the spread of future occurrences of incidents related to transmissible health diseases.

3. Proposed Work

The main task is to extract valuable information from multiple input sources of big data related to the epidemic, integrate it across borders, circulate it among numerous victims, thereby do preparation to forecast and control. The wider public is the smallest measurement that can be used to respond to health emergencies affecting individuals related to outbreak situations. Additionally, the public is the fundamental demographic community for most of the large amounts of knowledge about epidemics, including transportation facts, patient history-related patterns data, conveyance data, etc. The general population has a tremendous influence on the forecasting and controlling the epidemics through the activities that they engage in. The necessity for most requested data and information from the general population mainly comprises two factors. Perception of risk is of the utmost importance. The general public must be aware of information on the distribution of this pandemic, the medical care of individuals, and the impact of the circumstances from the collected big data. After that, they need to understand the risk of general population is to increase the need for understanding healthcare facilities and resources, such as clinics designated explicitly for treating infectious diseases, virtual consultations, instantaneous prescription decisions, and other similar services, is an additional source of demand.

The module automates data collection, organizing, analyzing, and graphing, creating a comprehensive data repository for pandemic details like infectious disease transmission. This streamlines gathering data, relationships between healthcare, and dissemination, improving the health of the public knowledge rapidity and credibility of disease prediction. The subjective element of the proposed model analyzes the information demands of many stakeholders, including government, healthcare organizations, businesses, the news media, and the general population, during an outbreak of an illness. The authorities require worldwide data and information. Healthcare providers require resources from the network, knowledge of the prevention of epidemics, individual medical information, and illness diagnosis reference databases. Businesses need an organizational collection of data that accurately reflects the general public's views, actions, and motives. Media demands encompass information sources, validity, and timeliness. Public needs include understanding risk and healthcare facility. The application model examines big data technology applications during the earlier prevention and recovery phase of modern outbreaks. During the initial preserving phase, the essential functions related to pandemic impact management and precise information diffusion. During pandemics, big data's critical applications include tracing and anticipating transmission paths, following infected individuals, screening close contacts, and conducting intelligent medical and medication research. During recovery, the essential applications include assisting businesses in resuming operations, fostering government-public relations, and optimizing control plans for epidemics.

a. Data Acquisition and Cleaning phase

This research classifies the various data sources involved in healthcare events under some groups. The first big data category is the pandemic, which includes information about infectious illnesses' nature, dimension, level, and risk level. A second category is the big data in the pharmaceutical and medical industries, which primarily encompasses medical care, drug development and research, individual electronic medical records, and other similar areas. Third, the government has freely available big data, which primarily includes information about shifting populations, belonging, congestion, the environment, and other issues. Fourth, the big data generated by the internet, which typically originates through online involvement of general population, records of searching, transportation information transaction data, and so on. The information that individuals produce through the use of handheld devices has the potential to become "inactive" accumulating resources.

Analyzing the data is the central component of the entire process of big data. This includes the analysis of the importance of pandemic details, the subjective attitude ability, and the forecast of epidemic progression, among other things. It accomplishes this by transforming data from several sources into mathematical models following the logic of the industry for needs in the business. This allows technology that is capable of generating autonomous outcomes in accordance with the demands of the algorithm in order to satisfy the process of making decisions requirements. On the other hand, the essence of data analytics is data modelling to support business decision-making. Furthermore, this dictates that the technique of analysis needs to be altered from fundamental investigation to skill learning, from reasoning to associative regulation, as well as from qualitative assessment to the enhanced interpretation of the matching of patterns. These changes are necessary since they are necessary in order to achieve the desired results. Table 1 defines the collected data preliminaries for analyzing the impact of big data.

Table 1: Data Preliminaries

Source Tag Information	Data Extraction & Category
Public health info	Clinical record prediction
Classification of infection	Pandemic risk category
Type of infection	Behavioural pattern
Reverse analysis	Risk identification alert
Close screening	Decision plan
Environment	Healthcare resource
Bigdata era	Transmission trace

Gathering, organizing, and interpreting knowledge through big data is an essential step in the evolution of innovation. Data will undoubtedly have as much impact as blood in the real-world applications used to prevent and manage epidemics. This indicates that individuals have to provide the data as a condition of participation to take advantage of both intelligent and instructive services. Consequently, the application of considerable data information to assist in preventing and managing the spread of illnesses or outbreaks while also taking into consideration the protection of individual data and privacy assurances should become a primary focus of research to fully exploit the actual value and beneficial effects that result from big data technology. The primary objective is to add value to massive amounts of multi-source data related to the pandemic spread and its infection by gathering important details, integrating international in-nature knowledge, circulating knowledge among numerous participants, and preparing for effective control and prevention. Data sources, including social networks, medicine transactions, web searches, and electricity use, lack relevant information for this pandemic and cannot substitute medical records as a significant source. Combining unstructured information with structured details, along with the analysis of correlation and tool refinement, increases the availability of information and provides a more valuable perspective than medical information. Figure 1 explains the overall working procedure of the PAID-BD.

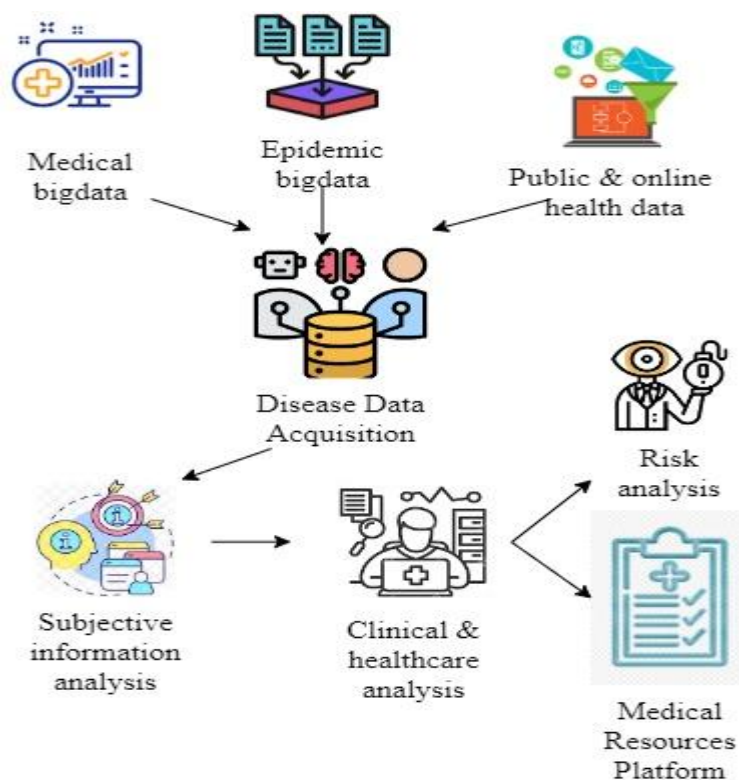


Figure 1: Schematic Representation of Proposed Model

b. Infectious Disease Transmission Forecasting Process

The analysis of the additional value method for applying big data for forecasting and controlling emergencies was presented earlier reveals that the heart of this process for extending the resources of an individual mental ability. The research unexplored with the support of information and computing and to gradually transform individual working practices while concurrently addressing the obstacles of pandemic control and prevention.

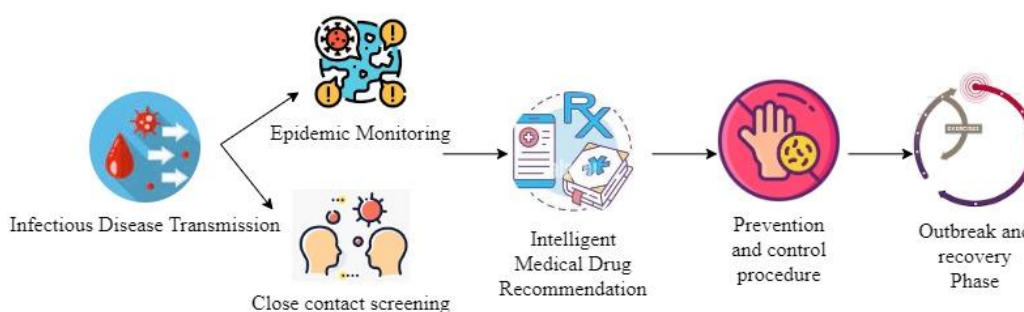


Figure 2: Prediction of Infectious Disease Transmission using BigData

Figure 2 discusses epidemic monitoring in medical and health-related big data, which relies on citizen behaviour and illness tracking, with data from drug retail and healthcare electronic records as primary support. Drug retail data, including respiratory, digestive, antibacterial, and influenza drugs, can partially predict disease prevalence and provide early warnings based on location. Studies indicate a strong correlation between antiviral medicine sales and influenza cases, as well as temperature and hand sanitizer consumption.

c. Application of Big Data in infectious disease Transmission prediction

Participants in the control of epidemics primarily consist of the authorities, healthcare facilities, organizations, the news media, and the general population. The demands of participants in the epidemic response procedure can symbolize the population's preferences in selecting and motivating continual enhancement of the effort to prevent and control infectious problems. The authorities are responsible for assuming The task of general strategy and collaboration in reaction to situations involving the health of people. The government's primary objective is to successfully combine the assets of individuals involved with collaborative and identical connections. This will result in forming an integrated control design that consists of sharing risks and integration according to the healthcare assistance from the outbreak of government.

On the one hand, it is necessary to make it possible to avert emergencies accurately and transmit monitoring and alert notifications. When it comes to efficiently controlling the spread of the epidemic, it must respond without delay. In addition to this, it is obligated to take full responsibility for the efficient distribution of resources and the reconstruction and recovery efforts that will be carried out after the epidemic. When it comes to preventing and controlling outbreaks, the requirement for knowledge and data from the government is global. This demand continues throughout the infection's development, epidemic, try and restoration phases. This phase emphasizes epidemic understanding, surveillance, and early prior notification. It involves analyzing epidemic characteristics, collecting data from multiple sources, and monitoring the global situation. To prioritize prevention, alerts will be delivered to appropriate organizations and society when monitoring results surpass the predefined level.

d. Recovery Phase Evaluation

During the recovering cycle, the pandemic is well managed, regions of individuals better life after the illness has been resolved. During this phase, the primary objective is to organize, assess, and interpret the knowledge on the entire process of a pandemic as well as to offer plans for recovery and the rebuilding process, to summarize the pandemic circumstance and early prediction. The average person is readily misled by false comments during the recovery period of the infection. These statements can potentially contribute to the primary and secondary risks associated with the outbreak; therefore, it is vital to improve the development of the general impact of collecting necessary data about individual health. Big data helps to analyze the behaviour of internet users and the guidelines that govern the dissemination of information. It can also handle knowledgeable communication according to a complete knowledge of the genuine requirements of the general population. This allows for The information is both present and precise to be delivered to the public regarding removing the outbreak and inquiry strategies. Additionally, it can prevent the adverse feedback triggered by reports from occurring in the first place.

Additionally, computer-generated A computerized reproduction of the whole representation of the development of the global epidemic is possible with the use of simulation technology that is based on big data. This technology can also assist the relevant departments in making organized breakdowns of the underlying factors, interruptions, and knowledge gained during the pandemic. Furthermore, it can facilitate optimising the plan for as well as the prevention and management of the pandemic and updating the preceding instance information repository. The module automates data collection, organizing, analyzing, and graphing, creating a comprehensive data repository for pandemic details like infectious disease transmission. The PAID-BD streamlines gathering data, relationships between healthcare, and dissemination, improving the health of the public knowledge rapidly and credibility of disease prediction.

4. Results and Discussion

For implementing the research, utilizing relational or NoSQL databases and info lakes are every instance of possible approaches employed in this research. The existing algorithms, such as AI-rec [14] and multi-model [15], are taken for comparison study with the following metrics: prediction accuracy, response time, adaptability ratio and transmission rate reduction analysis for analyzing the impact of big data application in infectious disease transmission.

a. Prediction Accuracy

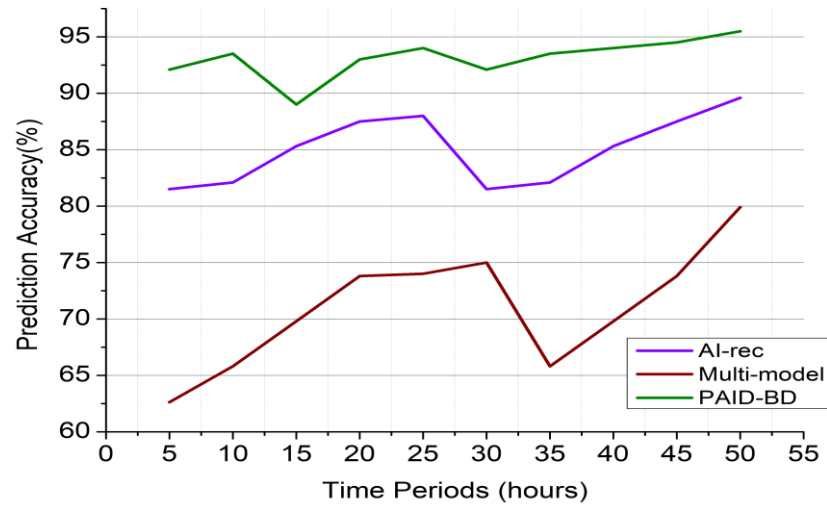


Figure 3: Prediction Accuracy Analysis

In Figure 3, An enormous quantity of varied information about infectious illnesses is gathered in this research study. These data include epidemiological data, which provides for cases reported and dissemination paths; environmental information, including features such as climate variables and atmosphere excellence; data on demographics, including features like the number of people and birth year for age circulation; and how individuals move and interpersonal relations. It is then subjected to preprocessing to organize, recover, and normalize the data in preparation for analysis. To carry out this process, it is necessary to identify critical indicators of disease transmission and to turn raw data into relevant characteristics of the input for modelling purposes. The findings of model evaluations are used to provide feedback, which is then used to modify and develop prediction models in an iterative manner using PAID-BD.

b. Response Time

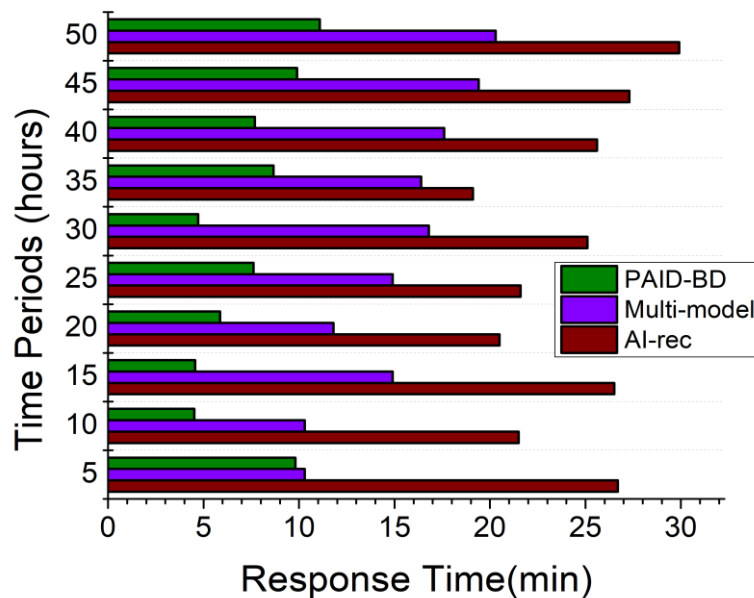


Figure 4: Response Time

In Figure 4, the computation of response time starts with identifying an infectious incident, which may involve the diagnosis of unique indicators, testing verification of a new infectious agent, or a rise in reported cases that exceed the expected amounts. When a contagious incident has been identified, the correct data is often communicated to the agencies in charge of public health or those involved in viral surveillance and response. When calculating response time, the time between the initial detection and receiving notice of agencies is considered. Following the completion of the illness event using the PAID-BD evaluation framework, the authorities in charge of public health will devise response plans and tactics to prevent the disease from spreading further and to safeguard public health. In the computation of response time, the amount of time it took to make these decisions and develop response plans is comprised.

c. Adaptability Ratio (%)

It is possible to gain knowledge of the capacity to adapt forecasting techniques to changing demographic circumstances by contrasting the actual precision with the anticipated accuracy. The ability to adjust the system is evaluated when actual reaction actions are assessed in terms of timeliness and efficacy compared to predefined agility criteria. This evaluation determines how quickly tactics may be adjusted in response to new threats. In these situations, researchers can evaluate the efficacy and robustness of big data-driven approaches to forecasting and preventing the transmission of infectious diseases by measuring the adaptability ratio. This research can provide valuable insights for future optimization efforts and serve as a roadmap for creating public health systems that are more adaptable and responsive, as shown in Figure 5 using Equation (1).

$$AR(\%) = \frac{\text{actual perf}}{\text{target perf on specific disease over a time}} \quad (1)$$

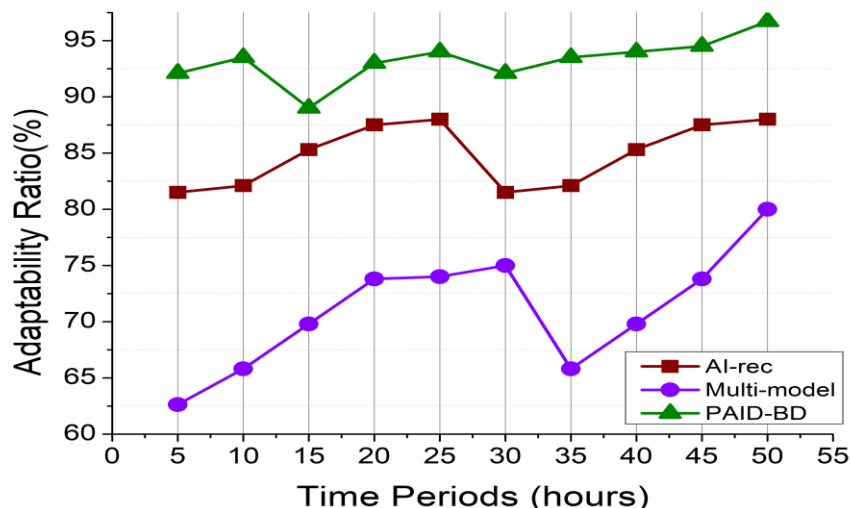


Figure 5: Analyzing Adaptability Ratio

d. Transmission Rate Reduction Analysis

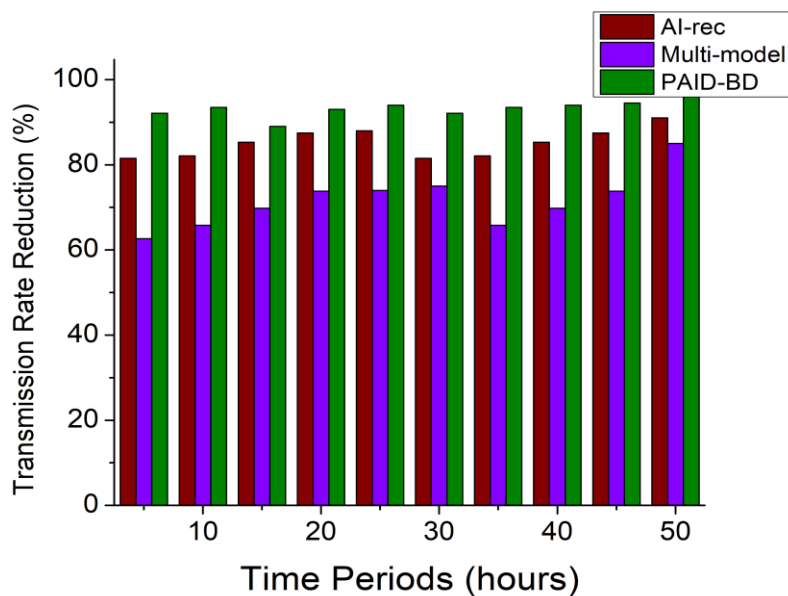


Figure 6: Analyzing Transmission Rate Reduction

Figure 6 shows the illustration of bigdata from various sources, such as demographic information, hospital tracks, communication feed, and environmental factors, is made possible through the use of analytics for big data. Researchers can get insights into mechanisms regarding illness propagation and pinpoint high-risk people or places by analyzing various data sources. Public health authorities are able to quickly recognize and respond to growing epidemics by continuously gathering and analyzing data streams. These data streams include hospitalizations, symptom news accounts, and mobility patterns. This allows the authorities to intervene to minimize transmission rates as quickly as possible during the outbreak. Researchers can construct models that can predict expected transmission rates under various scenarios and evaluate the potential impact of intervention strategies on reducing transmission by merging historical epidemiology data alongside real-time data streams. A significant reduction in the transmission rate indicates that the steps implemented, whether the use of

vaccination programs, protocols for quarantine, measures to isolate individuals, or other measures, have successfully reduced the spread of the illness within the local population.

5. Conclusion

The proposed research presents a novel approach to predicting infectious illness using big data ideas such as PAID-BD. The research findings on forecasting and preventing the spread of transmissible diseases through big data technologies have significant consequences for preparing public health interventions, responding to epidemics, and attempting to control diseases. The research outcome improves the prediction of disease dynamics, enhance the rapidity and efficiency of epidemic response, and reduce the impact of infectious diseases on global health. This will be accomplished through the integration of a variety of data resources and novel approaches for intervention. For forecasting disease outbreaks and informing prompt intervention measures, researchers can't predict by utilizing vast volumes of data from various sources. For predicting infectious illness during a pandemic and control efforts to be successful, cooperation must occur across domains and data exchange among academia, health officials, medical professionals, legislators, and other stakeholders. Stakeholders can cooperatively handle the complex problems of infectious illnesses and improve their readiness and response to epidemics if they share data, competence, and data resources using big data. The future enhancement investigated the implementation of sophisticated AI methodologies, such as deep neural networks and reinforcement neural networks, to analyze complicated datasets and recognize microscopic variations and patterns in the changing dynamics of infectious disease transmission.

REFERENCES

- [1]. Awotunde, Joseph Bamidele, et al. "Application of artificial intelligence and big data for fighting COVID-19 pandemic." *Decision Sciences for COVID-19: Learning Through Case Studies* (2022): 3-26.
- [2]. Sakly, Houneida, et al. "Artificial Intelligence and Big Data for COVID-19 Diagnosis." *Trends of Artificial Intelligence and Big Data for E-Health*. Cham: Springer International Publishing, 2023. 83-119.
- [3]. Alshahrani, Saeed M., et al. "Systematic Survey on Big Data Analytics and Artificial Intelligence for COVID-19 Containment." *Computer Systems Science & Engineering* 47.2 (2023).
- [4]. Hasan, Iqbal, et al. "Data analytics and knowledge management approach for COVID-19 prediction and control." *International Journal of Information Technology* 15.2 (2023): 937-954.
- [5]. Epizitone, Ayogeboh, Smangele Pretty Moyane, and Israel Edem Agbehadji. "Health information system and health care applications performance in the healthcare arena: a bibliometric analysis." *Healthcare*. Vol. 10. No. 11. MDPI, 2022.
- [6]. Ravikumar, Rejitha, et al. "The Impact of Big Data Quality Analytics on Knowledge Management in Healthcare Institutions: Lessons Learned from Big Data's Application within The Healthcare Sector." *South Eastern European Journal of Public Health* (2023).
- [7]. Shorman, Samer, Muath Jarrah, and Ahmed R. Alsayed. "The websites technology for Arabic language learning through COVID-19 pandemic." *Future of Organizations and Work After the 4th Industrial Revolution: The Role of Artificial Intelligence, Big Data, Automation, and Robotics*. Cham: Springer International Publishing, 2022. 327-340.
- [8]. Dhote, Sunita, et al. "Cloud computing assisted mobile healthcare systems using distributed data analytic model." *IEEE Transactions on Big Data* (2023).
- [9]. Agbehadji, Israel Edem, et al. "Review of big data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing." *International journal of environmental research and public health* 17.15 (2020): 5330.
- [10]. Corsi, Alana, et al. "Big data analytics as a tool for fighting pandemics: a systematic review of literature." *Journal of ambient intelligence and humanized computing* 12.10 (2021): 9163-9180.

- [11]. Pley, Caitlin, et al. "Digital and technological innovation in vector-borne disease surveillance to predict, detect, and control climate-driven outbreaks." *The Lancet Planetary Health* 5.10 (2021): e739-e745.
- [12]. Rees, E., et al. "Early detection and prediction of infectious disease outbreaks." *CCDR* 45.5 (2019).
- [13]. Judijanto, Loso, et al. "The Role of Big Data Technology in Predicting and Managing the Spread of Infectious Diseases." *Journal of World Future Medicine, Health and Nursing* 2.2 (2024): 219-230.
- [14]. Sood, Sandeep Kumar, et al. "An intelligent healthcare system for predicting and preventing dengue virus infection." *Computing* 105.3 (2023): 617-655.
- [15]. Shi, Honghao, et al. "Big data technology in infectious diseases modeling, simulation, and prediction after the COVID-19 outbreak." *Intelligent Medicine* 3.2 (2023): 85-96.
- [16]. Wang, Huiquan, and Hong Ye. "Constructing big data prevention and control model for public health emergencies in China: A grounded theory study." *Frontiers in Public Health* 11 (2023): 1112547.
- [17]. Amusa, Lateef Babatunde, et al. "Big Data and Infectious Disease Epidemiology: Bibliometric Analysis and Research Agenda." *Interactive journal of medical research* 12.1 (2023): e42292.
- [18]. Torres Munguía, Juan Armando, et al. "A global dataset of pandemic-and epidemic-prone disease outbreaks." *Scientific data* 9.1 (2022): 683.