

GLOBAL HEALTH INTELLIGENCE PLATFORMS DATA SCIENCE INFRASTRUCTURE FOR REAL-TIME EPIDEMIOLOGICAL SURVEILLANCE AND PANDEMIC FORECASTING

Md Imtiaz Faruk¹, Tushar Roy², Md Jubayar Hossain³, Muhammad Adnan⁴, Emran Hossain⁵,
Mahmudul Haque Rijvi⁶, Mohammed Majid Bakhsh⁷, Syed Mohammed Muhive Uddin^{*8}

^{1,2,3}St. Francis College, Brooklyn, New York, USA,

⁴Trine University, Detroit, Michigan, USA,

⁵Humphreys University, Stockton, California, USA,

^{6,7,*8} Washington University of Science and Technology, Alexandria, Virginia, USA

¹mfaruk@sfc.edu, ²troy3@sfc.edu, ³mhossain8@sfc.edu, ⁴muhammadadnan441@outlook.com,
⁵hu0112358@student.humphreys.edu, ⁶mrijvi.student@wust.edu, ⁷mohammedmajidb@gmail.com
^{*8}muhive.jiku@gmail.com

DOI: <https://doi.org/10.5281/zenodo.21155490>

Keywords

Infectious Disease Surveillance,
Real-Time Epidemiological
Monitoring, Pandemic Forecasting,
Digital Epidemiology, Public
Health Surveillance, Epidemic
Prediction

Article History

Received: 25 October 2024

Accepted: 04 December 2024

Published: 21 December 2024

Copyright @Author

Corresponding Author: *
Syed Mohammed Muhive
Uddin

Abstract

The increasing rate of infectious disease outbreak and globalization has created the necessity to suggest new sophisticated surveillance and monitoring systems with the ability to monitor the situation in real-time and provide accurate forecasting of a pandemic. This paper describes how the surveillance of infectious diseases is transformed by combining artificial intelligence (AI), digital epidemiology, and big data infrastructure. It initially examines the causes of inefficiencies in the conventional disease surveillance systems, such as poor stakeholder participation, poor use of infrastructure, regulatory barriers, and poor awareness of policymakers. The paper next explains the application of AI in the detection of outbreaks with a focus on predictive analytics, geospatial maps, and natural language processing in the earliest identification of disease trends. Moreover, the shift to digital epidemiology based on classical epidemiology is discussed, and the opportunities and challenges related to the large-scale digital sources of data are outlined. A data science infrastructure of real-time surveillance, such as healthcare analytics, big data processing frameworks, and a layered architecture aimed at dealing with heterogeneous health data is also provided in the paper. Lastly, the paper will discuss the significance of diverse data sources in forecasting an epidemic and outline future prospects, including AI-driven analytics, secure data sharing with blockchain, and global collaboration efforts. Development of new policy-guided collaboration with advanced data science technologies can help improve global preparedness and response to up-emerging infectious diseases threats by a significant margin.

I. INTRODUCTION

A transition in the global order that has profoundly impacted the socioeconomic landscape and the provision of healthcare is the result of the melting of national barriers and the rising interconnectivity of nations, which has been hastened by technology [1]. A robust health system is one that can withstand the shock of an emergency like Ebola while still provide standard medical care, allowing other parts of the nation to continue operating normally [2]. Throughout ages, infectious illnesses have placed alongside starvation and conflict as the worst threats to human survival and advancement. They continue to rank among the world's most common causes of death and disability. The worldwide burden of infections is significantly increased by the periodic emergence of epidemics of both novel and old infectious diseases against a backdrop of persistent infections [3]. Studies of these new illnesses demonstrate the evolutionary characteristics of harmful microorganisms and the dynamic interactions among bacteria, their hosts, and the environment.

Infectious illnesses that are either recently discovered in a community or that have been there but are spreading more quickly are referred to as emerging infectious diseases (EID). To put it simply, these could include previously unknown infections that emerge in regions undergoing ecological transformation [4], identified infections that disseminate to novel populations or geographic regions, emergent infections arising from alterations or development of pre-existing organisms, or re-emergent infections due to antibiotic resistance in established pathogens or deficiencies in public health interventions.

The ongoing coronavirus disease 2019 (COVID-19) outbreak has caused substantial disruptions due to its rapid transmission and elevated mortality rate, rendering it a worldwide crisis. The COVID-19 pandemic has presented a significant threat to worldwide public health [5]. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which was first identified and isolated in individuals exposed at a seafood market in Wuhan City, Hubei Province, is the cause of COVID-19, China, in December 2019 [6]. immunological traits are increasingly recognised as possible indicators for

disease progression and prospective therapeutic targets for COVID-19, resulting from the elucidation of the relationship between immune responses and the disease [7].

The way people live and work has changed as a result of globalisation. Consequently, it has been determined that travel and trade play a significant role in the spread of illness. The increasing urbanization and enhanced integration of the global economy have also promoted global connectedness. The spread of disease is now largely dependent on globalisation. It is improbable that other nations would be able to achieve China's accomplishment at such a rapid pace due to differences in political regimes' ability to impose limits as well as individual ideas, customs, and lifestyles. In any event, South Korea's success in testing, treating, and isolating all cases [8] as well as evaluating and separating COVID-19 patients' close connections [9], provides optimism. The impact of globalization was defined in relation to mobility, economy, and healthcare systems. The extent of people's movement was assessed using mobility data and statistics on airline and seaport trade [10]. The economic impact was determined using the labor, academic institutions, food and agriculture, event cancellations, and supply chain [11]. When assessing the healthcare capability, national preparation and healthcare system indicators were taken into considerations. The purpose of the following paper is to discuss the transformation of the process of infectious disease surveillance based on the use of artificial intelligence, digital epidemiology, and data science systems to allow real-time epidemiological surveillance and forecasting pandemics. In a progressively connected and digitalized society, high-tech data-driven technologies are vital to enhance the early detection of outbreaks, disease surveillance, and global preparedness to health.

A. Structure of the Paper

The paper is structured in the following way: Section II explains how the process of infectious disease surveillance has altered, along with surveillance issues, AI, and digital epidemiology. Section III gives the data science infrastructure of real-time surveillance. The Section IV explains the forecasting of pandemics and future trends. Section V discusses

related literature, and Section VI closes the paper and provides future research directions.

II. TRANSFORMING INFECTIOUS DISEASE SURVEILLANCE: DETERMINANTS, ARTIFICIAL INTELLIGENCE, AND DIGITAL EPIDEMIOLOGY

For than a century, the management of infectious illnesses has been a cornerstone of international relations. However, regardless of numerous progresses, health threats on the planet still continue to loom large- and even get bigger. Health interdependence is increasing as the process of economic integration at the global level continues. The two chapters in this part discuss both conventional and novel global health hazards. More people have died from infectious diseases than from war throughout history [12]. Approximately one-third of Europeans died from the Black Death in the 14th century, while a 3rd of the Roman Empire's population died from what was probably the plague during the Justinian era. Between 60% and 90% of the indigenous people of the Americas from Mexico downstream perished from smallpox, measles, or influenza throughout the 16th and 17th centuries.

A. Determinants of Inefficiencies in Disease Surveillance Frameworks

Reducing mortality and morbidity and maintaining a high-quality public health system depend on the implementation of effective infectious disease surveillance systems. Public health surveillance lessens the impact and spread of unfavourable healthcare occurrences and makes data and information more accessible [13]. In turn, this facilitates quick public health response plans, permits efficient monitoring and assessment of suggested measures, makes it simple to identify new and emerging illnesses, and offers population health security and stability.

Low disease surveillance is caused by certain factors, including:

- Insufficient stakeholder participation and resource contributions that might improve the surveillance system.
- Insufficient use of the current surveillance framework to advance public health.
- Insufficient effort in formulating plans to prioritise the elimination of certain endemic infectious diseases.

- Unfavorable laws and rules pertaining to the use of surveillance technologies.
- A further variable contributing to subpar disease monitoring systems is the dearth of knowledge among policymakers about the potential effects and influence of monitoring on public health.

B. Role of AI in Outbreak Detection

The high incidence and transmission of infectious diseases have emerged as an increasing concern in the world, with the health systems of the people failing to effectively and efficiently respond to the situation. Conventional methods of disease surveillance do not always work in real-time outbreak detection [14], and the control of such outbreaks needs prompt decisions that are based on data. AI has become a revolutionary technology in disease surveillance, offering superior capabilities in data analysis, forecasting, and decision-making. The applications of AI in outbreak detection are as follows:

1) Real-Time Outbreak Detection Using Predictive Analytics

AI can have one of the most promising uses in disease surveillance by detecting an outbreak and predicting it before it becomes too large. Predictive analytics involves the use of historical data, including disease spread patterns, environmental variables and population movement to predict future outbreaks. Patterns and trends that are hard to detect by humans can be identified by ML algorithms, and early intervention can take place.

2) Geospatial Mapping and AI: Pinpointing Disease Hotspots

Discovering Disease Hotspots: Geospatial mapping, which is driven by AI, is important in monitoring the spread of diseases across the geographical region. AI systems have the potential to combine health surveillance report data, satellite images, weather data, and social variables to produce dynamic maps that forecast the spread of diseases in the regions. AI, with the assistance of Geographic Information Systems (GIS) can be used to track cluster zones of infections, create hotspots, and track movement trends of patients.

3) *Social media and Natural Language Processing (NLP) in Epidemic Trends*

The role of social media platforms and news outlets has increased as significant sources of real-time information, which in many instances gives an early signal of an outbreak of the disease. The social media posts, news articles, and blogs can be monitored using AI which relies on NLP to identify the spread of the disease, as well as concerns related to the health of the population. When AI systems scan and analyze large amounts of text data, they are able to recognize mentions of symptoms, odd health trends or disease outbreaks even before official reports are submitted.

C. *Digital Epidemiology in Transition*

Epidemiology refers to the investigation of population health trends and factors. It is based on a

variety of health-related information from several sources, including sociodemographic data, lab tests, and questionnaires. Digital Epidemiology, a branch of Epidemiology, has become widespread in recent years due to the rise of the digital technology and the availability of computational power [15]. Digital epidemiology was first described as the monitoring and analysis of health and illness patterns among individuals using digital information sources, such as social media, mobile phone trace, and other internet sources. The Digital Epidemiology and its potential benefits of making surveillance faster, cheaper, and broader is not a novel concept. An especially compelling instance is the establishment of a computerized record network for sentinel physicians in France in 1984, aimed at monitoring syndromic influenza. Table I summarizes the significant sources of bias in digital and classical epidemiology

TABLE I. BIASES AND MITIGATION STRATEGIES IN CLASSICAL AND DIGITAL EPIDEMIOLOGY

Bias	Classical Epidemiology	Digital Epidemiology	Key Differences
Sampling and Representation	Often relies on clinical or survey-based samples that may not represent the whole population.	Uses large-scale digital data but may exclude populations with limited internet or technology access.	Classical bias arises from limited sampling; digital bias arises from unequal digital access.
Measurement and Information Bias	Data collected through standardized instruments but may suffer from reporting errors or delays.	Data from sensors, apps, or online activity may contain inconsistencies or behavioural noise.	Classical methods are more controlled; digital methods provide larger but less standardized data.
Technological and Platform Bias	Limited technology influence; data mainly collected through established clinical systems.	Platform design, algorithm rules, and device differences affect how data are captured.	Digital surveillance depends heavily on technological infrastructure.
Attrition and Behavioral Bias	Participant dropout occurs but is easier to track in controlled studies.	Users frequently stop using apps or change behaviour online, affecting data reliability.	Digital data is more sensitive to behavioural changes and

			platform engagement.
Confounding and Temporal Bias	Traditional epidemiology addresses confounding through study design and statistical control.	Digital data often includes many uncontrolled variables and rapid temporal changes.	Classical studies rely on structured methods; digital studies require advanced analytics to manage complexity.
Confirmation and Anchoring Bias	Researchers may favor hypotheses supported by existing theories or prior findings.	Algorithms and models may reinforce patterns found in training data.	Both approaches risk reinforcing prior assumptions but through different mechanisms.
Algorithmic Bias	Statistical models are usually transparent and interpretable.	Machine learning models may produce biased predictions if trained on unrepresentative data.	Digital epidemiology faces additional risks from opaque algorithms and biased training datasets.

III. DATA SCIENCE INFRASTRUCTURE FOR REAL-TIME SURVEILLANCE

In recent years, pandemics and infectious diseases have exposed the flaws in conventional public health surveillance systems. The world and cities are expanding faster, which makes it easier to spread the epidemic, and this is why real-time monitoring is required to raise awareness of health problems and dynamically respond to them in order to contain the situation. Regardless of the current improvement in the collection of data and epidemiological models, the methods of monitoring are largely based on retrospective analysis, which is not always in time to prevent widespread transmission. Public health monitoring has never been more significant because of the regular outbreak of infectious diseases, the

increase in urbanization and the problem of climate change. Timely surveillance of disease outbreaks, trend tracking, and intervention necessitate robust public health surveillance systems. These systems often collect and analyze data in the past, delaying the response to health issues.

A. Analytics of Healthcare Domain

The proportion of elderly people in society is rising globally, and the ageing of humanity is a phenomenon that affects healthcare services in a number of ways, particularly in terms of costs [16]. In such circumstances, millions of individuals may see a decrease in their quality of life if they continue to rely on traditional methods.

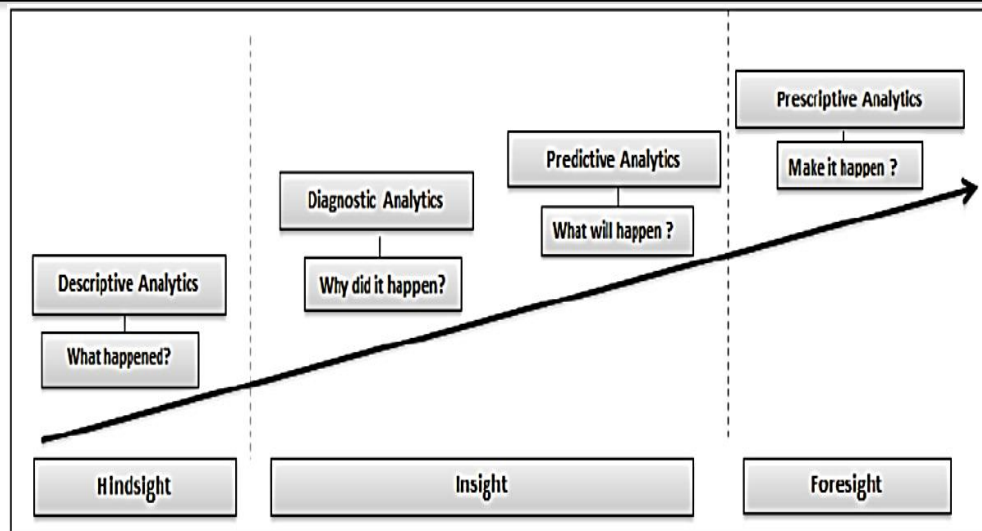


Fig. 1. Analytics for the Healthcare Domain

Due to the extensive use of wearable and mobile sensor technologies, the number of data sources in the healthcare industry has had a significant expansion in recent years, inundating the area with a vast quantity of data. [17]. As a result, it becomes difficult to analyze healthcare data using conventional techniques that are inadequate to deal with the large amount of diverse medical data. Figure 1, which depicts the analytics stages for the healthcare domain and is further examined, shows the four kinds of analytics that are often present in the healthcare domain. Medical systems may function better if big data technologies are included into healthcare analytics:

- **Descriptive Analytics:** It entails reporting on and characterising current events. At this level of analytics, a number of methods are employed. Descriptive analytics employ tools from descriptive statistics, such as histograms and charts.
- **Diagnostic Analytics:** The purpose of this is to provide light on the cause-and-effect relationships between certain occurrences. Diagnostic analysis, for instance, attempts to comprehend the causes of some patients' regular readmission by utilising a variety of techniques, including clustering and DT.
- **Predictive Analytics:** It demonstrates the ability to foretell future occurrences, aids in pattern detection, and estimates the likelihood

of uncertain outcomes. Predicting whether a patient may experience complications is an illustration of its function. ML is frequently utilized in creating predictive models.

- **Prescriptive Analytics:** It aims at suggesting appropriate measures that culminate in the best decisions. For example, where there is a significant likelihood of a negative side effect, prescriptive analysis could advise against a certain treatment [18]. Prescriptive analytics may be carried out using techniques such as DT and Monte Carlo simulations.

B. Big Data Processing Approaches

Big data technologies have generated a lot of interest because of their capacity to process high-volume data to better effects than the traditional methods [19]. Structured, semi-structured, and unstructured data are all supported by big data frameworks, which also offer a number of functionalities [20]. These elements, which enhance decision-making by locating relevant data, include massive data mining tools and predictive model development. Table II shows the various processing frameworks.

TABLE II. BIG DATA PROCESSING SOLUTIONS

Framework	Type	Latency	Developed by	Stream Primitive	Stream Source
Hadoop	Batch	Minutes or more	Yahoo	Key-value	HDFS
Strom	Streaming	Subseconds	Twitter	Tuples	Spouts
Spark Streaming	Batch/streaming	Few seconds	Berkley AMPlay	DStream	HDFS
S4	Streaming	Few seconds	Yahoo	Events	Networks
Flink	Batch/streaming	Few seconds	Apache Software Foundation	Key-value	KAFKA

C. An Extensible Big Data Architecture for Healthcare

Developing a versatile system that can handle a range of scenarios, such as early disease detection and emergency detection, is a benefit [21]. In order to

manage large medical data from various sources and formats, a novel architecture is proposed in this study. As seen in Figure 2, this architecture is described as the following data management scenario.

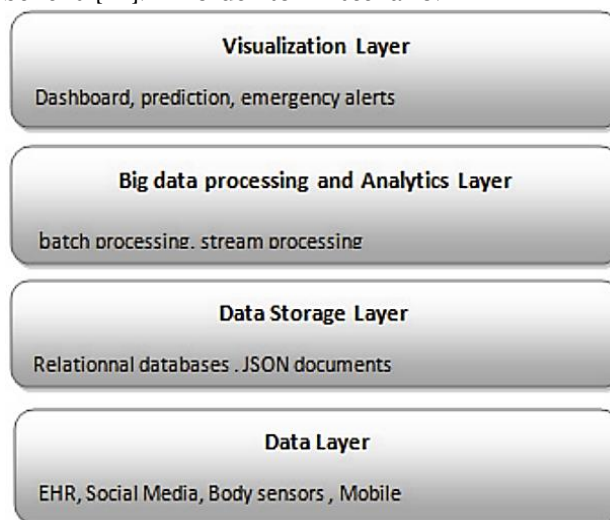


Fig. 2. The Layer Architecture

1) Data Layer

The system's data stores raw health-related data from several sources. Such sources include Electronic Health Records (EHRs), social media, body sensors and wearable devices, and mobile applications. The data obtained in this case may include patient symptoms, mobility, environmental conditions, and health behavior indicators. Since such datasets are sometimes large, heterogeneous, and constantly evolving, this layer is necessary for real-time epidemiological surveillance and the early prevention of potential outbreaks.

2) Data Storage Layer

The data storage layer has the responsibility of storage and organization of the data collected such that it can be accessed and processed efficiently. It normally employs relational databases to store structured data and JSON or document-based to store semi-structured or unstructured data. The layer is providing data integrity, scalability, and secure storage and offers the ability to retrieve it quickly to perform the analysis. It also helps in the assimilation

of heterogeneous data that is provided by various sources of surveillance.

3) *Big Data Processing and Analytics Layer*

This layer uses data processing, integration and advanced analytics. It uses technologies capable of handling large volumes of health data through batch processing (for analyzing large historical datasets) and stream processing (for analyzing real-time incoming data). These may be utilised in ML algorithms, statistical models, and predictive analytics to detect anomalous disease patterns, identify outbreaks, and predict possible epidemic patterns. This layer produces sense out of raw data.

4) *Visualization Layer*

The visualization layer provides the analyzed data in a format that is easy to understand and to act upon by the officials in the field of public health, as well as policymakers and researchers. It also has interactive dashboards, predictive models and emergency alert systems reporting real-time disease trends, geographical spread and risk prediction. This layer helps make quick decisions that result in an effective response to the public health needs by converting complicated analytical tasks into visual presentations.

IV. PANDEMIC FORECASTING AND FUTURE DIRECTIONS

Human history has been highly impacted by epidemics since the Black Death in the 14th century and the COVID-19 epidemic in more modern times. Such outbreaks bring a high morbidity and mortality rate, disruption of the economy and erosion of healthcare systems [22]. The international effect of epidemics highlights why it is important to establish appropriate management control mechanisms that can reduce the effects of epidemics. Early detection and forecasting are important in this regard. In my opinion, as it allows the public health authorities to initiate timely actions against the spread of an infectious disease, assign resources rationally, and avoid the disruption of the whole society, it allows a timely detection of the potential outbreak and control its spread.

A. *Role of Data in Epidemic Forecasting*

Information plays a central role in the prediction of epidemics, which forms the basis of model construction and forecasting. The use of different types of data in these environments provides different information on disease dynamics.

1) *Health Records*

Clinical data [23] and electronic health records (EHRs) are detailed records of individual cases, including symptoms, treatments, and outcomes. This kind of information is crucial for understanding the dynamics of clinical illnesses and trends in their incidence and spread.

2) *Social Media*

Mobile phone data, GPS and transportation network data provide information about the patterns of movement of people. The potential transmission of infectious illnesses, particularly those transferred by intimate contact, may be modeled using information about people's movement both within and across areas.

3) *Environmental Data*

Temperature, humidity, air quality are among some of the factors that can determine the proliferation of some diseases. By incorporating environmental data in the models, the predictive power of the models can be optimized in predicting diseases, which are sensitive to climatic conditions.

4) *Genomic Data*

Pathogen genomics is a source of information about the genetic composition of infectious organisms. Genomic information is able to trace the development and transmission of pathogens, detect mutations, and so inform vaccines.

B. *Future Directions and Recommendation*

In the future, a number of emerging technologies are set to transform big data analytics to revolutionize epidemic forecasting even more.

1) *Technological Advancements*

AI is in the limelight, and ML algorithms are becoming more complex to predict the occurrence of diseases on massive datasets. AI possesses the

capability to assess trends and abnormalities in real-time, delivering unparalleled precision and rapidity in epidemic prediction. Also, the implementation of blockchain technology would help in solving the data security and privacy issues. Through the use of safe and transparent information transmission methods, it can maintain the integrity of data and foster confidence among all parties involved, blockchain technology offers a decentralized and immutable store of health information.

2) *Policy Development*

Adaptive and progressive policy frameworks are necessary in order to utilize these technological developments. Existing policies tend to be slow to keep up with technological advancement, which leaves loopholes in regulatory supervision and kills potential gains. This is why the necessity of flexible policies, which could be adjusted to the changes in technologies, is pressing. Such policies need to promote innovation without neglecting ethical values and the safeguarding of data. Policy improvements that should be recommended include establishing regulatory sandboxes to test new technologies in a restricted setting prior to large-scale implementation. The given strategy enables policymakers to comprehend what new technologies imply and adjust regulations to this fact.

3) *Collaboration and Global Initiatives*

International partnerships play a vital role in combating health threats at the international level. Epidemics are not considerate of boundaries; hence, the international community needs to collaborate. International organizations like the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) are crucial to data exchange and teamwork when it comes to epidemic prediction. Programs and collaborations to integrate and share health data at the global level are crucial. The Global Health Security Agenda (GHS), a multinational endeavor to increase the ability to avoid, identify, and address potential risks from infectious diseases, is one example of this. These efforts can be extended into data sharing and collaborative analysis on a more thorough level with the help of big data technologies.

V. LITERATURE REVIEW

This section introduces prior research on the epidemiological surveillance and real-time prediction of pandemics. The methodical review of previous studies, focusing on AI-based prediction systems, digital surveillance systems, and data-driven forecasting methods, along with their major findings, is condensed in Table III.

L. Liu et al. (2023) In the early phases of the COVID-19 pandemic, a DL-based method was proposed for the detection of stigmatized material on OSN platforms. In order to identify important geographical and temporal features of COVID-19 stigmatization for the purpose of early warnings and risk reduction, their approach employs a semantic-based quantitative analysis. The prediction utility of their approach are investigated through empirical assessments. The results of the co-occurrence network visualisation using Gephi show two different categories of stigmatised terms related to Wuhan residents and their eating habits, respectively [24].

H. Wang et al. (2022) proposed T-SIRGAN, this combines the benefits of DL models and epidemiological theories to simulate non-linear correlations and represent complex epidemic processes for a more accurate COVID-19 growth estimate. To begin, T-SIRGAN uses the epidemiology-based Susceptible-Infectious-Recovered (SIR) model to provide simulation data. The next step is to feed these datasets into a generative adversarial network (GAN) as adversarial examples to enhance the data. Next, using the obtained synthetic data, Transformers are utilised to forecast COVID-19's future patterns [25].

D. Machi et al. (2021) provided methods for COVID-19 pandemic planning and response that are scalable and enabled by high speed computing. With it, you may create forecasts and other types of hypothetical analysis for each of the 3140 counties in the United States, which includes all 50 states and DC, and perform fine-grained simulations every day because to the scalability of their technique. To achieve this goal, their operations make use of a hybrid cloud/cluster system that combines local and distant cluster computing resources with over 20,000 CPU cores operating for six to nine hours each day. Their models are used by their state (Virginia), state hospital network, university, the Centers for Disease

Control and Prevention, together with the Department of Défense [26].

Y. Dong and Y.-D. Yao (2021) discussed about how the system for preventing and controlling epidemics may be integrated with the IoT. In particular, they presented a potential IoT platform that combines fog and clouds for use in intelligent and systematic prevention and control of COVID-19. There are five initiatives on this platform: Tracking SARS-CoV-2 Mutations, Diagnosing COVID-19 Symptoms, Monitoring Quarantine, Tracking Contacts and Social Distancing, and COVID-19 Forecasting Outbreaks. In order to demonstrate the possibilities of IoT in combating future epidemics of infectious diseases or the present COVID-19 pandemic, they looked into and reviewed the most recent literature on these five approaches [27].

A. Ramchandani, C. Fan, and A. Mostafavi (2020) The proposed model is capable of learning intricate relationships between a wide range of heterogeneous characteristics, incorporating historical infection growth, social distancing data, intra-county and inter-county travel, and census data. Using information obtained from several sources, they calculated the range of an increase in infected people in each US county in seven days. Furthermore, they employed

the model to determine the most important characteristics for forecasting the spread of infection. Additionally, they examine feature pairings and calculate the amount of second-order interaction that has been seen. The proposed framework model be used in addition to traditional epidemiological models for tracking pandemics at the national level, as COVID-19, since experiments demonstrate that it achieves adequate prediction performance and reasonably interpretable feature analysis findings [28].

M. Rimsan et al. (2020) proposed a cutting-edge blockchain-based architecture that uses the design science research method to integrate intercountry COVID-19 monitoring for patients who have been screened or infected. Governments, aviation companies, health authorities, and citizens may find the proposed framework useful in making critical choices about infection detection, prediction, and prevention. The unanticipated COVID-19 epidemic and unrestrained worldwide expansion show that current global healthcare networks have limits in handling public safety concerns. When patients go overseas, it might be difficult to find infected or tested people worldwide [29]

TABLE III. LITERATURE REVIEW OF DATA-DRIVEN APPROACHES FOR REAL-TIME EPIDEMIOLOGICAL SURVEILLANCE

References	Study Focus	Input Data	Implemented Methods	Case Study	Key Contributions
Liu et al. (2023)	Detection of stigmatized content related to COVID-19 on online social networks for early risk alerts	Online social network posts and textual data related to COVID-19	Deep learning with semantic-based quantitative analysis and co-occurrence network visualization using Gephi	Early-stage COVID-19 discussions on OSN platforms	Identified spatial-temporal characteristics of stigmatization and revealed two major stigma clusters related to Wuhan residents and dietary practices
Wang et al. (2022)	Predicting COVID-19 infection growth accurately using deep learning and hybrid	Data from epidemiological simulations produced by the SIR model	T-SIRGAN framework combining SIR model, Generative Adversarial Network (GAN) for data	COVID-19 outbreak prediction	Improved modeling of complex epidemic dynamics and enhanced prediction accuracy through hybrid deep learning and

	epidemiology models		augmentation, and Transformer for forecasting		epidemiological modeling
Machi et al. (2021)	Large-scale pandemic forecasting and planning using high-performance computing	Large epidemiological datasets for U.S. counties	High-performance computing workflows with hybrid cloud/cluster systems for large-scale simulations	COVID-19 forecasting across 50 U.S. states and 3140 counties	Enabled daily fine-grained simulations and county-level forecasts to support decision-making by government and health agencies
Dong & Yao (2021)	Internet of Things (IoT) integration with systems for controlling and preventing epidemics	Data from IoT devices, health monitoring systems, and surveillance platforms	Fog-cloud integrated IoT platform that facilitates mutation tracking, contact tracing, epidemic forecasting, diagnostics, and quarantine monitoring	COVID-19 pandemic control systems	Demonstrated how IoT-based infrastructures can enhance intelligent epidemic monitoring and management
Ramchandani et al. (2020)	Predicting future COVID-19 infection growth using heterogeneous data sources	Census data, mobility data, social distancing data, historical infection growth data	Machine learning model analyzing heterogeneous features and second-order feature interactions	COVID-19 prediction across U.S. counties	Predicted infection growth up to 7 days ahead and identified influential factors affecting pandemic spread
Rimsan et al. (2020)	Global tracking of infected or tested patients using a secure data-sharing framework	Global health records and travel-related infection data	Blockchain-based framework designed using design science research methodology	Global COVID-19 patient tracking	Proposed a secure intercountry system for infection tracking, prediction, and prevention to support government and health authorities

VI. CONCLUSION AND FUTURE WORK

Digital epidemiology, artificial intelligence, and data science infrastructure are required to transform infectious disease surveillance in order to enhance the preparedness of the world to new health threats. This paper has discussed the factors that contribute to inefficiencies in the conventional surveillance

systems and how inefficiency due to resource constraints, gaps in policies, and inadequate use of infrastructure may reduce effective disease surveillance measures. With AI-based analytics, predictive modelling, and geospatial capabilities, there is a great opportunity to detect an outbreak at its initial stages and conduct real-time

epidemiological surveillance. Additionally, to increase forecasting accuracy, the shift to digital epidemiology will allow data from many sources being integrated, including social media, EHRs, mobility, and environmental data. The suggested big data design and processing models facilitate the effective management of heterogeneous healthcare data and allow timely insights to be made to the decision-making process in relation to the population health. In the future, artificial intelligence, data-sharing technologies with a high level of security, and international cooperation will be of critical importance in enhancing the effectiveness of epidemic prediction systems and the more proactive and resilient global health surveillance systems.

The future development will apply the advanced artificial intelligence, the federation of learning, and real-time data sharing platforms to improve epidemic prediction accuracy. Genomic surveillance, IoT-based health surveillance, and secure data models like blockchain can also enhance the global disease surveillance process, allowing to detect and respond to diseases faster, coordinate their efforts, and offer more effective responses to them.

REFERENCES

- A. Warriar, "Real-Time Healthcare Event Processing: Stream Analytics for Clinical Decision Support," *Int. J. Emerg. Res. Eng. Technol.*, vol. 1, no. 4, December, pp. 47-54, 2020, doi: 10.63282/3050-922X.IJERET-V1I4P106.
- S. A. Tabish and N. Syed, "Future of Healthcare Delivery: Strategies that will Reshape the Healthcare Industry Landscape," *Int. J. Sci. Res.*, vol. 4, no. 2, pp. 2319-7064, 2013.
- D. M. Morens, G. K. Folkers, and A. S. Fauci, "The challenge of emerging and re-emerging infectious diseases," *Nature*, vol. 430, no. 6996, pp. 242-249, Jul. 2004, doi: 10.1038/nature02759.
- D. B. McArthur, "Emerging Infectious Diseases," *Nurs. Clin. North Am.*, vol. 54, no. 2, pp. 297-311, Jun. 2019, doi: 10.1016/j.cnur.2019.02.006.
- A. Warriar, "COVID-19 Contact Tracing: Privacy-Preserving Integration Architectures for Public Health Surveillance," *Int. J. AI, BigData, Comput. Manag. Stud.*, vol. 2, no. 1, pp. 88-97, 2021, doi: 10.63282/3050-9416.IJAIBDCMS-V2I1P109.
- Y. Shi et al., "An overview of COVID-19," *J. Zhejiang Univ. B*, vol. 21, no. 5, pp. 343-360, May 2020, doi: 10.1631/jzus.B2000083.
- L. Yang et al., "COVID-19: immunopathogenesis and Immunotherapeutics," *Signal Transduct. Target. Ther.*, vol. 5, no. 1, p. 128, Jul. 2020, doi: 10.1038/s41392-020-00243-2.
- N. Shrestha et al., "The impact of COVID-19 on globalization," *One Heal.*, vol. 11, p. 100180, Dec. 2020, doi: 10.1016/j.onehlt.2020.100180.
- P. Sinha et al., "Evaluation of Saline as Potential Alternative to Viral Transport Media for COVID-19 Samples Stored at Different Temperatures," *J. Clin. Diagnostic Res.*, vol. 15, no. 4, pp. 19-22, 2021, doi: 10.7860/jcdr/2021/48389.14831.
- H. Gajjar, U. Korat, and V. Gondaliya, "Telemedicine with Automated ECG Analysis," *Int. J. Eng. Res. Technol.*, vol. 2, no. 10, pp. 2253-2255, 2013.
- A. C. Senok, G. A. Botta, and O. O. Soge, "Emergence and Spread of Antimicrobial-Resistant Pathogens in an Era of Globalization," *Interdiscip. Perspect. Infect. Dis.*, vol. 2012, pp. 1-3, 2012, doi: 10.1155/2012/286703.
- T. A. Faunce, *Global Public Goods*, 2nd ed., vol. 4. London: Elsevier, 2012. doi: 10.1016/B978-0-12-373932-2.00340-9.
- A. T. Aborode et al., "Impact of poor disease surveillance system on COVID-19 response in africa: Time to rethink and rebuilt," *Clin. Epidemiol. Glob. Heal.*, vol. 12, p. 100841, Oct. 2021, doi: 10.1016/j.cegh.2021.100841.
- U. A. Salam, "Real-Time Disease Surveillance: AI in Tracking and Controlling Outbreaks," *Emerg. Trends Med.*, vol. 1, no. 1, pp. 9-13, 2021.

- A. E. Aiello, A. Renson, and P. N. Zivich, "Social Media- and Internet-Based Disease Surveillance for Public Health," *Annu. Rev. Public Health*, vol. 41, no. 1, pp. 101–118, Apr. 2020, doi: 10.1146/annurev-publhealth-040119-094402.
- J. Antony, P. Palsuk, S. Gupta, D. Mishra, and P. Barach, "Six Sigma in healthcare: a systematic review of the literature," *Int. J. Qual. Reliab. Manag.*, vol. 35, no. 5, pp. 1075–1092, May 2018, doi: 10.1108/IJQRM-02-2017-0027.
- C. Tayal, "Data Quality Assessment and Cleaning Framework for Healthcare Databases Using Python," *Int. J. Artif. Intell. Data Sci. Mach. Learn.*, vol. 3, no. 4, pp. 107–112, 2022, doi: 10.63282/3050-9262.IJAIDSML-V3I4P112.
- A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, Apr. 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.
- S. Suthaharan, "Big Data Analytics," in *Machine Learning Models and Algorithms for Big Data Classification*, 2016, pp. 31–75. doi: 10.1007/978-1-4899-7641-3_3.
- [. Suthaharan, "Big Data Essentials," in *Machine Learning Models and Algorithms for Big Data Classification*, Springer, Boston, MA, 2016, pp. 17–29. doi: 10.1007/978-1-4899-7641-3_2.
- N. El aboudi and L. Benhlila, "Big Data Management for Healthcare Systems: Architecture, Requirements, and Implementation," Jun. 2018. doi: 10.1155/2018/4059018.
- Y. Macha, "A Review of Cloud-Based CRM Systems in Healthcare: Advances, Tools, Challenges, and Best Practices," *Int. J. Curr. Eng. Technol.*, vol. 12, no. 6, pp. 848–856, 2022, doi: 10.14741/ijcet/v.12.6.20.
- V. Pal, "Foundation Models for Multi-Modal Clinical Decision Support Systems," *ESP J. Eng. Technol. Adv.*, vol. 2, no. 2, pp. 183–191, 2022, doi: 10.56472/25832646/JETA-V2I2P121.
- L. Liu, Z. Cao, P. Zhao, P. J.-H. Hu, D. D. Zeng, and Y. Luo, "A Deep Learning Approach for Semantic Analysis of COVID-19-Related Stigma on Social Media," *IEEE Trans. Comput. Soc. Syst.*, vol. 10, no. 1, pp. 246–254, Feb. 2023, doi: 10.1109/TCSS.2022.3145404.
- H. Wang *et al.*, "Predicting the Epidemics Trend of COVID-19 Using Epidemiological-Based Generative Adversarial Networks," *IEEE J. Sel. Top. Signal Process.*, vol. 16, no. 2, pp. 276–288, Feb. 2022, doi: 10.1109/JSTSP.2022.3152375.
- D. Machi *et al.*, "Scalable Epidemiological Workflows to Support COVID-19 Planning and Response," in *2021 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, IEEE, May 2021, pp. 639–650. doi: 10.1109/IPDPS49936.2021.00072.
- Y. Dong and Y.-D. Yao, "IoT Platform for COVID-19 Prevention and Control: A Survey," *IEEE Access*, vol. 9, pp. 49929–49941, 2021, doi: 10.1109/ACCESS.2021.3068276.
- A. Ramchandani, C. Fan, and A. Mostafavi, "DeepCOVIDNet: An Interpretable Deep Learning Model for Predictive Surveillance of COVID-19 Using Heterogeneous Features and Their Interactions," *IEEE Access*, vol. 8, pp. 159915–159930, 2020, doi: 10.1109/ACCESS.2020.3019989.
- M. Rimsan, A. K. Mahmood, M. Umair, and F. Hassan, "COVID-19: A Novel Framework to Globally Track Coronavirus Infected Patients using Blockchain," in *2020 International Conference on Computational Intelligence (ICCI)*, IEEE, Oct. 2020, pp. 70–74. doi: 10.1109/ICCI51257.2020.9247659.