(REVIEW ARTICLE)

Check for updates

# Data analytics in public health, A USA perspective: A review

Abdulraheem Olaide Babarinde [1, *], Oluwatoyin Ayo-Farai [2], Chinedu Paschal Maduka [3], Chiamaka Chinaemelum Okongwu [4] and Olamide Sodamade [5]

[1] The Heller School for Social Policy and Management, Brandeis University, Waltham Massachusetts.
[2] Jiann-Ping Hsu College of Public Health, Georgia Southern University, Georgia, U. S.
[3] Institute of Human Virology, Nigeria
[4] Faculty of Community Health and Primary Care, Department of Public Health, University of Lagos, Nigeria.
[5] Africa Voices HQ, Nigeria.

## Abstract

The integration of data analytics into public health practices represents a transformative paradigm shift in the United States. This review provides a comprehensive analysis of the impact and implications of data analytics on public health strategies, with a focus on disease surveillance and health policy within the USA. In the context of disease surveillance, data analytics has emerged as a crucial tool for real-time monitoring and early detection of health threats. Leveraging diverse datasets, including electronic health records and social media, allows for swift identification of trends and anomalies, enabling proactive responses to potential outbreaks. Advanced analytics techniques, such as machine learning and predictive modeling, contribute to the precision of surveillance efforts, facilitating targeted interventions and resource allocation. Beyond disease surveillance, data analytics significantly influences health policy. Evidence-based policy formulation is enhanced through data-driven insights, providing policymakers with a foundation for understanding the impact of interventions and designing strategies that align with the unique needs of diverse populations. Resource allocation strategies are optimized, ensuring efficient use of limited resources by analyzing health outcomes, service utilization patterns, and cost-effectiveness. Continuous monitoring and evaluation of implemented health policies through data analytics enable policymakers to adapt strategies in response to evolving health challenges, fostering a dynamic and adaptive public health ecosystem. As the landscape of public health evolves, data analytics in the USA continues to play a central role in shaping strategies and policies. The study delves into the historical context, key components, applications, and success stories, providing valuable insights for policymakers, public health professionals, and researchers aiming to navigate the complexities of data-driven public health management.

**Keyword:** Data Analytic; Public health; USA; Health management; Disease surveillance

## 1. Introduction

In the dynamic landscape of contemporary public health, the integration of data analytics has emerged as a transformative force, redefining how we perceive, manage, and address health challenges. This study navigates the realms of data analytics in public health, specifically exploring its nuances and applications within the United States.

At its core, data analytics in public health involves the systematic use of statistical methods, computational algorithms, and data-driven technologies to analyze vast datasets derived from various sources, elucidating patterns, trends, and insights relevant to population health (Manrai et al., 2017). It encompasses a spectrum of techniques, from traditional statistical analysis to advanced machine learning, all geared towards unraveling the complexities of health-related data (Shafqat, et al., 2020).

* Corresponding author: Abdulraheem Olaide Babarinde

The significance of data analytics in public health is multifaceted, ushering in a paradigm shift in how health practitioners, policymakers, and researchers approach the challenges of the modern era (Galetsi, Katsaliaki, and Kumar, 2019). By harnessing the power of data, public health professionals can gain deeper insights into disease patterns, risk factors, and healthcare outcomes. This, in turn, facilitates evidence-based decision-making, enhances the precision of interventions, and ultimately contributes to the overarching goal of improving population health.

In addition, data analytics plays a pivotal role in predicting and mitigating the impact of public health crises, offering real-time insights during disease outbreaks, and optimizing resource allocation (Razzak, Imran, and Xu, 2020). Its applications extend beyond epidemiology to include health disparities analysis, policy formulation, and personalized medicine, marking it as an indispensable tool in the public health arsenal.

Within the United States, the adoption of data analytics in public health is deeply embedded in the nation's commitment to harnessing technological advancements for the betterment of healthcare outcomes. Federal agencies, academic institutions, and healthcare organizations across the country have embraced data-driven approaches to address complex health challenges and inform policy decisions.

From the pioneering efforts of the Centers for Disease Control and Prevention (CDC) to the integration of electronic health records (EHRs) in healthcare systems, the USA has been at the forefront of leveraging data analytics for public health applications (Ranasinghe, 2023). The country's vast and diverse healthcare landscape, coupled with a robust digital infrastructure, provides a fertile ground for exploring innovative ways to utilize data analytics for public health surveillance, intervention planning, and health equity.

This study examines the intricacies of data analytics applications in the USA's public health domain, examining successes, challenges, and the evolving landscape that shapes the future of population health management. It also examines some selected cases in the United States of America.

## 2. Data Analytics on Public Health Strategies and Disease Surveillance and Health Policy in the USA

Data analytics has become a cornerstone in the realm of public health, revolutionizing strategies and decisions to address the dynamic healthcare landscape (Mooney, and Pejaver, 2018). In the United States, the integration of data analytics has played a pivotal role in shaping disease surveillance efforts and influencing health policy formulation. This study explores the profound impact of data analytics on public health strategies, with a specific focus on its contributions to disease surveillance and its far-reaching effects on health policy within the USA.

Data analytics enables real-time monitoring of health indicators, facilitating the early detection of disease outbreaks. By analyzing diverse datasets, including electronic health records, social media, and traditional surveillance systems, public health agencies can swiftly identify trends, clusters, and anomalies indicative of potential health threats. Real-time monitoring and early detection of diseases is an essential aspect of modern healthcare. By identifying diseases early, clinicians can intervene promptly and effectively, improving patient outcomes and reducing the overall burden of disease.

Several technologies are being developed and implemented to enable real-time monitoring and early detection of diseases. These technologies include wearable devices, such as smartwatches and fitness trackers, can collect a variety of data about an individual's health, including heart rate, sleep patterns, and physical activity levels. This data can be analyzed to identify potential health risks and early signs of disease. Figure 1 shows Industrial wearable technologies. (**A**) Evolution of wearable medical devices (**B**) Application of wearable devices in the healthcare and biomedical monitoring systems. Reproduced with permission from Hwang, I.; et al. Multifunctional smart skin adhesive patches for advanced health care; Wiley, 2018 and Yao, H.; et al. A contact lens with embedded sensor for monitoring tear glucose level; Elsevier, 2011 Guk et al (2019) did a detailed analysis of the evolution of wearable devices fitted with real-time disease monitoring with emphasis on personalized healthcare.

Biosensors are devices that can detect specific biomarkers in the body, such as glucose levels or proteins associated with certain diseases. Biosensors can be implanted or worn on the body and can provide real-time data about an individual's health status. AI is being used to analyze large amounts of data from wearable devices, biosensors, and other sources to identify patterns and predict potential health risks. AI can also be used to develop personalized risk assessments and early warning systems for disease. Figure 2 shows the various applications of biosensor
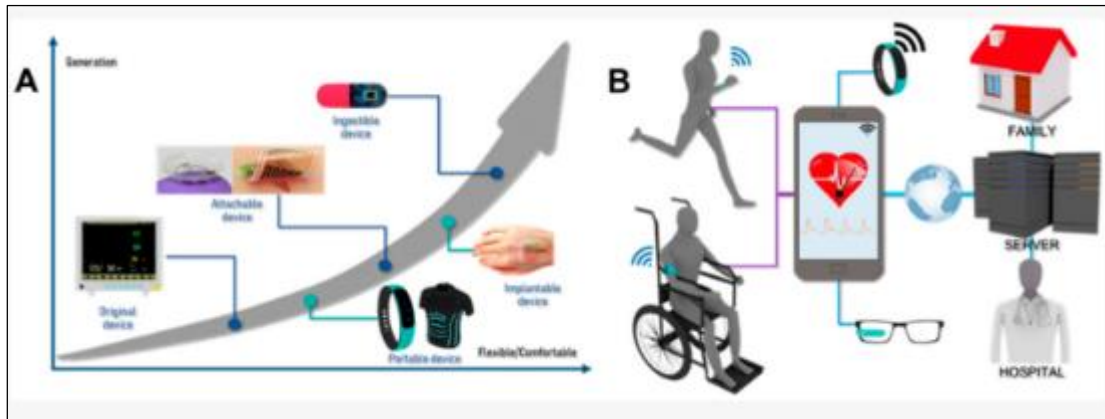
**Figure 1** Industrial wearable technologies (Guk et al., 2019)
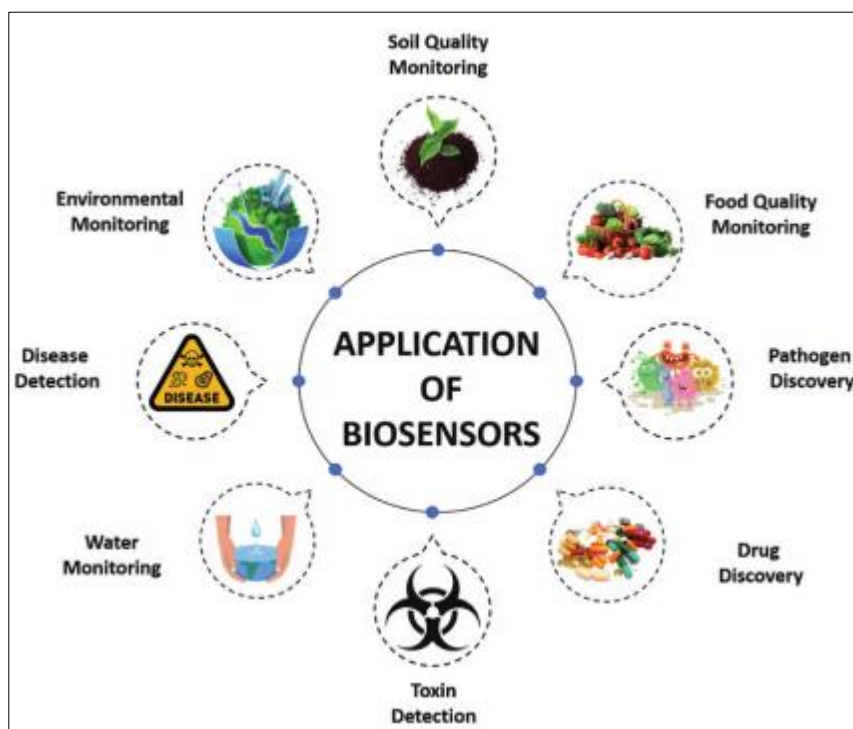


**Figure 2** Schematic representation of Biosensor applications (Tetyana et al., 2021)

Tetyana et al., 2021 did a comprehensive review of design, development and applications of biosensors. In their study, they opined that the quality of life has greatly improved as a result of the capacity to precisely monitor processes that affect both human nature and its surroundings and to identify even the smallest physiological change in the human body. Biosensors remain indispensable in numerous domains such as environmental monitoring, drug development, food regulation, biomedical diagnosis, and therapy and illness progression monitoring.

Biomarkers are substances that can be found in the blood, urine, or other bodily fluids that can indicate the presence of a disease. For example, elevated levels of certain proteins or enzymes can be indicative of cancer or other diseases. Figure 3 shows the biomarkers for detection and prognosis of diseases.

Medical imaging techniques, such as MRI and CT scans, can be used to detect structural abnormalities that may be indicative of disease. For example, an MRI can be used to detect tumors in the brain. AI is being used to develop algorithms that can analyze medical data, such as wearable sensor data, biomarkers, and medical imaging scans, to identify early signs of disease. AI can also be used to develop personalized risk assessment tools that can help to identify individuals who are at high risk of developing certain diseases. These technologies are still under development, but they have the potential to revolutionize the way we diagnose and treat diseases. By using these technologies, we can move

away from a reactive approach to healthcare, where we wait for patients to become symptomatic before we intervene, to a proactive approach, where we can identify and treat diseases early, when they are most treatable.
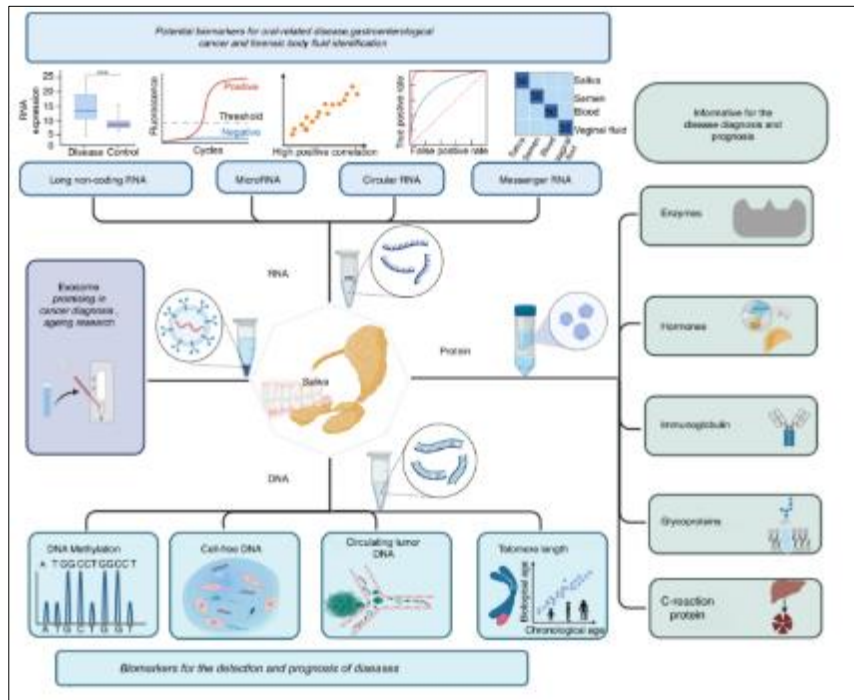


**Figure 3** Biomarkers for detection and prognosis of diseases (Song et al., 2023)

Figure 4 shows the various Functions of Artificial Intelligence (AI) for Infectious-Disease Surveillance. The references



**Figure 4** Various Functions of Artificial Intelligence (AI) for Infectious-Disease Surveillance (Brownstein et al., 2023).

Real-time monitoring and early detection of diseases offer several benefits. Early detection of diseases can lead to earlier intervention and treatment, which can improve patient outcomes and reduce the risk of complications (Gardy and Loman, 2018). Early detection can also reduce healthcare costs by preventing the need for more expensive treatments and hospitalizations (Tahamtan, and Ardebili, 2020). Early detection can also improve an individual's quality of life by allowing them to manage their disease more effectively and make informed decisions about their health. A characteristic of hereditary metabolic diseases like methylmalonic acidemia (MMA) is chronic kidney disease (CKD), which can induce low metabolic activity and stunted growth, delaying the diagnosis and treatment of renal disease. KowsalyaDevi Pavuluri and colleagues at the interdisciplinary departments of Functional Brain Imaging, Metabolic Genetics, Radiology and Radiological Science, and Pediatrics in the United States developed an alternate method to monitor renal function, which is now available in a new paper published on Science Advances. The technique was developed by the research team using a pH-sensitive MRI agent to monitor the course of the disease in an MMA mouse model (Pavuluri et al., 2019). The study effectively made it easier to employ MRI for CKD monitoring and early identification.

Several real-time monitoring and early detection applications are already in use or under development, including Continuous glucose monitoring (CGM) systems are used by people with diabetes to monitor their blood glucose levels in real time. This data can help them to manage their blood sugar levels and prevent complications. Wearable devices and biosensors can be used to monitor heart rate, rhythm, and other cardiac parameters. This data can help to identify potential heart problems early on. Wearable devices and biosensors can be used to monitor respiratory rate, oxygen levels, and other respiratory parameters. This data can help to identify potential respiratory problems early on.

Despite the significant progress in real-time monitoring and early detection of diseases, there are still some challenges that need to be addressed, including Data privacy and security, Accuracy and reliability, and Integration with healthcare systems. Ensuring the privacy and security of health data is essential. The accuracy and reliability of real-time monitoring technologies need to be continuously improved. Real-time monitoring technologies need to be seamlessly integrated with existing healthcare systems.

As research and development continue, we can expect to see even more advanced real-time monitoring and early detection technologies emerge in the future. These technologies have the potential to revolutionize healthcare by enabling earlier detection and intervention for a wide range of diseases.

Advanced analytics techniques, such as machine learning and predictive modeling, empower public health officials to forecast disease outbreaks (Aldahiri, Alrashed, and Hussain, 2021, Singhal, Baxi, and Mago, 2022). By analyzing historical data, environmental factors, and population dynamics, predictive models contribute to proactive decision-making, allowing for timely interventions and resource allocation.

Data analytics enables precision surveillance, allowing public health practitioners to tailor interventions to specific populations or geographic areas (Jamshidi, et al., 2022). This targeted approach not only enhances the effectiveness of public health measures but also minimizes unnecessary disruptions to communities.

## 2.1. Health Policy Impacts of Data Analytics

This includes Evidence-Based Policy Formulation, Resource Allocation and Efficiency, and Monitoring Policy Effectiveness. Data-driven insights serve as a foundation for evidence-based health policy formulation. Policymakers leverage analytics to understand the impact of interventions, assess health disparities, and design policies that align with the unique needs of diverse populations. Data analytics guides resource allocation strategies, ensuring that limited resources are directed where they are most needed. By analyzing health outcomes, service utilization patterns, and cost-effectiveness, policymakers can optimize the allocation of funds and improve the efficiency of public health initiatives. Continuous data analytics allows for the ongoing monitoring and evaluation of implemented health policies (Manohar, and Keerthana, 2023). This iterative process enables policymakers to assess the effectiveness of interventions, identify areas for improvement, and adapt strategies in response to evolving health challenges.

In the USA, data analytics has become an indispensable tool for public health, transforming the way disease surveillance is conducted and policies are crafted. The integration of advanced analytics not only enhances the precision of disease monitoring but also empowers policymakers to make informed decisions that positively impact population health. As data analytics continues to evolve, its role in shaping public health strategies and policies is poised to become even more influential, ushering in an era of proactive and data-driven healthcare management.

## 2.2. Historical Context of Data Analytics in US Public Health

The historical evolution of data analytics in public health in the United States traces a trajectory from rudimentary record-keeping to the sophisticated and technology-driven approaches employed in contemporary healthcare systems. Understanding this historical context provides insight into the development, challenges, and significance of data analytics in shaping public health strategies. The earliest forms of data collection in public health involved basic record-keeping of vital statistics, such as births, deaths, and disease occurrences. The field of epidemiology emerged during this period, with pioneers like John Snow utilizing basic data analysis to map and understand the spread of diseases such as cholera. The 20th century witnessed the formalization of public health surveillance systems. National and state health agencies started collecting and analyzing data systematically to monitor and control the spread of infectious diseases. The focus was primarily on manual data collection and analysis. The late 20th century saw the gradual transition from paper-based health records to electronic health records (EHRs). The digitization of health data allowed for more efficient data storage and retrieval, laying the foundation for more sophisticated data analytics in the future. The proliferation of computing power and advancements in statistical methods contributed to the integration of data analytics into public health practices. Statistical analyses became more sophisticated, allowing for better interpretation of large datasets. Geographic Information Systems (GIS) gained prominence, enabling spatial analysis for disease mapping and hotspot identification.

The 21st century witnessed a data revolution in healthcare. The advent of big data technologies allowed for the analysis of massive datasets, including those generated by EHRs, genomics, and social determinants of health. Advanced analytics, including machine learning and predictive modeling, became integral to disease surveillance, outbreak prediction, and personalized medicine. The contemporary period is characterized by a focus on public health informatics and interoperability. Efforts have been made to standardize data formats and terminologies to enhance interoperability between different health systems. This facilitates seamless data exchange and collaborative research efforts. The global COVID-19 pandemic highlighted the critical role of data analytics in public health. Real-time data tracking, predictive modeling, and data visualization became instrumental in monitoring the spread of the virus, allocating resources, and guiding public health interventions.

The historical context of data analytics in US public health reflects a journey from basic record-keeping to a sophisticated and technology-driven approach. The ongoing evolution underscores the importance of adapting data analytics to address emerging health challenges and improve overall population health outcomes.

## 2.3. Key Components of Data Analytics in Public Health

The key components of data analytics in public health encompass a diverse set of elements that collectively contribute to the systematic analysis and interpretation of health-related data. These components play a crucial role in shaping public health strategies, guiding interventions, and facilitating evidence-based decision-making. Some of the fundamental elements are here discussed.

### 2.3.1. Data Sources and Types

Public health relies on surveillance systems to collect and monitor health data (Velasco et al., 2014, Thacker, and Berkelman, 1988). These systems may include disease registries, syndromic surveillance, and notifiable disease reporting. The digitization of health records allows for the systematic collection of patient information, treatment histories, and outcomes (Choi, 2012). EHRs contribute to comprehensive and centralized data storage. Beyond clinical data, social determinants such as socioeconomic status, education, and living conditions are essential components. Integrating these factors provides a holistic understanding of health disparities.

### 2.3.2. Data Collection and Processing Techniques

The use of big data technologies enables the processing and analysis of massive datasets. This includes tools and techniques for handling large volumes of structured and unstructured data efficiently. Advanced analytics techniques, such as machine learning algorithms and predictive modeling, enhance the ability to identify patterns, trends, and potential outcomes. These methods are used for disease prediction, risk stratification, and treatment optimization. GIS allows for the spatial analysis of health data, enabling the mapping of disease prevalence, identification of hotspots, and visualization of health disparities based on geographic locations.

### 2.3.3. Data Integration and Interoperability

Data integration and interoperability are two concepts that are often used interchangeably, but they have distinct meanings. Integration refers to the process of combining multiple applications to function together as a unified whole.

On the other hand, interoperability is the ability of different systems to communicate with each other and exchange data in a meaningful way. In healthcare, interoperability is crucial for achieving a holistic view of a patient's care journey. It allows for the connection of data between systems and devices, which can be crucial for effective delivery of solutions to the market. Data integration, on the other hand, involves combining data from several disparate sources, which are stored using various technologies and provide a unified view of the data. It is essential for almost every digitalization project, such as during the migration from a legacy ERP software to a new system (Petrasch and Petrasch, 2022). Standardizing data formats and terminologies ensures consistency and interoperability across different health systems. Common standards facilitate the exchange of data between diverse sources. As data analytics involve sensitive health information, ensuring privacy and adherence to ethical guidelines is crucial. Robust data governance frameworks are essential to protect individual privacy while extracting meaningful insights.

### 2.3.4. Data Analysis and Interpretation

Data analysis and interpretation are two important aspects of research. Data analysis is the process of examining data using statistical and analytical tools to identify patterns, relationships, and trends. Interpretation, on the other hand, involves making sense of the data by drawing conclusions and making inferences based on the results of the analysis. There are several methods and techniques for data analysis and interpretation, including qualitative and quantitative analysis. Qualitative analysis involves analyzing non-numerical data such as text, images, and videos, while quantitative analysis involves analyzing numerical data using statistical tools. Data interpretation is important because it helps researchers to categorize, manipulate, and summarize information in order to answer critical questions. It also helps to identify gaps in knowledge and areas for further research. Traditional statistical methods are employed to analyze health data, assess correlations, and derive significance. Descriptive statistics, inferential statistics, and hypothesis testing contribute to evidence-based decision-making. Communicating complex health data effectively is achieved through data visualization techniques. Graphs, charts, and dashboards enhance the accessibility and understanding of key findings for diverse stakeholders.

### 2.3.5. Public Health Informatics

Public health informatics involves the design and implementation of information technology systems to support data analytics. This includes databases, data warehouses, and health information exchanges. These systems assist public health professionals in making informed decisions by providing real-time data, trends, and relevant information to guide interventions and policies. These key components collectively contribute to the robust foundation of data analytics in public health. The integration of diverse data sources, advanced analytics techniques, and a commitment to privacy and ethical considerations enables public health practitioners to leverage data effectively for improved population health outcomes.

## 2.4. Applications of Data Analytics in Public Health

The applications of data analytics in public health are diverse and impactful, spanning from disease surveillance to policy formulation. The integration of data analytics enhances the efficiency and effectiveness of public health strategies, contributing to evidence-based decision-making and improved health outcomes. The key applications of data analytics in public health is here discussed.

### 2.4.1. Disease Surveillance and Outbreak Detection

Data analytics enables real-time monitoring of health indicators, facilitating the early detection of potential disease outbreaks. Timely identification of unusual patterns or clusters of cases allows for swift public health responses. Advanced analytics techniques, including predictive modeling, contribute to forecasting disease outbreaks. By analyzing historical data, environmental factors, and population dynamics, public health officials can anticipate and prepare for emerging health threats.

### 2.4.2. Health Risk Assessment and Stratification

Data analytics is utilized to stratify populations based on health risks. This allows for targeted interventions and resources to be allocated to specific groups, optimizing the impact of public health initiatives. By analyzing demographic and socioeconomic data, public health professionals can identify and address health disparities, ensuring that interventions are tailored to the unique needs of diverse populations.

### 2.4.3. Intervention Planning and Resource Allocation

Data analytics guides the efficient allocation of resources by analyzing health outcomes, service utilization patterns, and cost-effectiveness. This ensures that limited resources are directed where they are most needed. Public health officials

use data analytics to assess the effectiveness of interventions. Analyzing outcomes helps in refining strategies, identifying successful interventions, and adapting approaches for continuous improvement.

### 2.4.4. Policy Formulation and Evaluation

Data analytics provides a foundation for evidence-based policy formulation. Policymakers leverage analytics to understand the impact of interventions, assess health trends, and design policies aligned with population health needs. Continuous data analytics allows for the ongoing monitoring and evaluation of implemented health policies. Policymakers can assess the effectiveness of interventions, identify areas for improvement, and adapt strategies based on real-time data.

### 2.4.5. Personalized Medicine and Healthcare Delivery

Personalized medicine is a medical practice that uses an individual's genetic profile to guide decisions made in regard to the prevention, diagnosis, and treatment of disease. It is rapidly having an impact on how drugs are discovered and developed; how patients are diagnosed and treated; and how health care delivery is channeling its resources to maximize patient benefits. The International Consortium for Personalised Medicine (ICPerMed) envisions that personalized medicine will lead to the next generation of healthcare by 2030. This vision focuses on five perspectives: individual and public engagement, involvement of health professionals, implementation within healthcare systems, health-related data, and the development of sustainable economic models that allow improved therapy, diagnostic and preventive approaches as new healthcare concepts for the benefit of the public. The implementation of personalized medicine will result in more efficient and equitable healthcare, access to modern healthcare methods, and improved control by individuals of their own health data, as well as economic development in the health sector.

Data analytics contributes to the stratification of patients based on their health status and characteristics. This personalized approach guides healthcare delivery, ensuring that individuals receive tailored interventions and treatments. By analyzing individual health data, predictive analytics assists in forecasting health outcomes for specific patients. This informs healthcare providers about potential risks and allows for proactive management.

### 2.4.6. Behavioral Health Interventions

Data analytics incorporates social determinants of health to analyze and understand health behaviors. This information is crucial for designing targeted interventions and health promotion programs. Advanced analytics models can predict health-related behaviors based on historical data, enabling the development of strategies to modify behaviors and improve public health outcomes.

### 2.4.7. Response to Public Health Emergencies

Public health emergencies are situations that require immediate action to prevent or mitigate the spread of disease and protect public health. The World Health Organization (WHO) has a Health Emergencies Programme that works with countries and partners around the world to prepare for, prevent, detect, and respond to disease outbreaks and other health emergencies. The programme focuses on building core public health and health system capacities that reduce the health risks and consequences of all types of emergencies. WHO supports member states to evaluate, develop, and strengthen core capacities laid out in the International Health Regulations, to detect, assess, notify, and report events and to respond promptly and effectively to public health emergencies. The Centers for Disease Control and Prevention (CDC) also provides information for the public and emergency responders on how to stay safe during public health emergencies. During public health emergencies, such as pandemics, data analytics facilitates real-time data tracking. This includes monitoring the spread of infectious diseases, tracking healthcare resource utilization, and informing rapid response strategies. Analytics helps in optimizing the allocation of resources during emergencies, ensuring that healthcare facilities, personnel, and supplies are directed where they are most needed.

The applications of data analytics in public health continue to evolve as technology advances and new data sources become available. These applications collectively contribute to a data-driven and proactive approach to addressing public health challenges and improving overall population health.

## 2.5. Success Stories and Case Studies

### 2.5.1. Notable Public Health Initiatives Utilizing Data Analytics

Public health initiatives that utilize data analytics have the potential to improve the quality of healthcare delivery and patient outcomes. One such initiative is the use of electronic health records (EHRs) to track and analyze patient data.

EHRs can be used to identify patients who are at risk for certain diseases, monitor the effectiveness of treatments, and improve patient outcomes. Another example is syndromic surveillance, which aggregates data on chief complaints or diagnoses to identify possible surges of infectious disease or bioterrorism. Big data analytics can also be used to inform healthcare by improving decision-making processes with real-time analytics, enhancing patient-centric healthcare and quality of care, detecting health threats, and improving disease monitoring. Case studies have been developed for public health issues that enable healthcare leaders to use big data analytics tools in their work. Project DETECT (Digital Engagement and Tracking for Early Control and Treatment) a collaboration between the Mayo Clinic and nference, leveraged data analytics to detect and monitor the spread of COVID-19 in the United States. By analyzing diverse datasets, including clinical and non-clinical sources, the project aimed to identify COVID-19 hotspots, predict disease trends, and facilitate early interventions. Project DETECT demonstrated the potential of data analytics in providing real-time insights during a pandemic, aiding public health responses and informing policy decisions.

Another case study is the Flu Near You. The Flu Near You, a crowd-sourced health surveillance initiative, utilized data analytics to track and monitor flu-like symptoms reported by individuals across the United States. By analyzing user-reported data, the initiative aimed to provide early warnings of flu outbreaks, contributing to timely public health interventions. Flu Near You showcased the power of citizen-generated data in augmenting traditional surveillance systems, providing a complementary and real-time source of information for public health planning.

### 2.5.2. Positive Impact of Data Analytics on Health Outcomes

Data analytics has the potential to revolutionize healthcare by improving decision-making processes with real-time analytics, enhancing patient-centric healthcare and quality of care, detecting health threats, and improving disease monitoring. Healthcare data analytics can be used to identify patients who are at risk for certain diseases, monitor the effectiveness of treatments, and improve patient outcomes. Electronic health records (EHRs) can be used to track and analyze patient data, while syndromic surveillance aggregates data on chief complaints or diagnoses to identify possible surges of infectious disease or bioterrorism. Big data analytics can also be used to inform healthcare by improving decision-making processes with real-time analytics, enhancing patient-centric healthcare and quality of care, detecting health threats, and improving disease monitoring.

Case studies have been developed for public health issues that enable healthcare leaders to use big data analytics tools in their work. Reducing Hospital Readmissions - Mount Sinai Hospital. The Mount Sinai Hospital in New York implemented a data analytics-driven program to reduce hospital readmissions. By analyzing patient data, including medical history, social determinants, and post-discharge activities, the hospital developed predictive models to identify patients at risk of readmission. The program resulted in a significant reduction in hospital readmissions, improving patient outcomes and optimizing healthcare resource utilization. Another case is the Optimizing Diabetes Management - Geisinger Health System. The Geisinger Health System implemented a data analytics program to optimize diabetes management. By analyzing patient EHRs, lifestyle data, and socio-economic factors, the health system developed personalized care plans and interventions for diabetes patients. The data-driven approach led to improved glycemic control, reduced complications, and enhanced overall health outcomes for individuals with diabetes.

### 2.5.3. Lessons Learned from Implementations

Successful data analytics initiatives in public health require collaboration among interdisciplinary teams, including data scientists, healthcare professionals, and public health experts. The success of Project DETECT highlighted the importance of bringing together expertise from healthcare, data science, and epidemiology for effective pandemic response. The accuracy and reliability of data are critical for meaningful analysis. Standardization of data formats and terminologies is essential for interoperability. Initiatives like Flu Near You emphasized the need for robust data quality assurance processes to ensure the credibility of user-generated health data. Maintaining patient privacy and adhering to ethical guidelines are paramount in data analytics initiatives. Transparency in data use and clear communication with stakeholders is crucial. Programs like the diabetes management initiative at Geisinger underscored the importance of ethical considerations in handling sensitive health information. Public health data analytics initiatives should be dynamic and adaptable. Continuous iteration based on feedback and evolving health challenges is necessary for sustained success. Programs addressing hospital readmissions emphasized the need for ongoing evaluation and adjustment of predictive models to remain effective over time.

These success stories and case studies highlight the positive impact of data analytics in diverse public health initiatives. They also underscore the importance of interdisciplinary collaboration, data quality, ethical considerations, and continuous adaptation for successful implementation and sustained positive health outcomes.

## 2.6. Challenges and Limitations in Data Analytics for Public Health

The use of data analytics in public health has the potential to revolutionize healthcare by improving decision-making processes with real-time analytics, enhancing patient-centric healthcare and quality of care, detecting health threats, and improving disease monitoring. However, there are several challenges and limitations that need to be addressed. One of the main challenges is the lack of interoperability of health data, which makes it difficult to share data between different systems. Another challenge is the problematic data conventions, which can lead to inaccurate or incomplete data. Institutionalized practices in care delivery and the misaligned incentives of various actors in the industry also pose challenges to the implementation of data analytics in public health. Public health data often suffer from incompleteness or inaccuracy, impacting the reliability of analytics outcomes. Inconsistent reporting and variations in data collection methods contribute to this challenge. Implementing robust data quality assurance processes, standardizing data collection methods, and promoting adherence to data quality standards can help address this challenge. Variability in data formats, coding systems, and terminologies across different healthcare systems hinders interoperability and standardized analysis. Advocating for and implementing standardized data formats and terminologies can improve consistency and facilitate data integration across diverse sources.

Also, Health data often exist in silos, limiting seamless data sharing between healthcare institutions, public health agencies, and research entities. Developing interoperable health information exchange systems, fostering collaboration, and addressing legal and privacy concerns can help break down data silos. Privacy regulations and concerns about data ownership create barriers to sharing health data. Balancing the need for data access with privacy protection is a delicate challenge. Establishing clear guidelines for data sharing, implementing de-identification techniques, and ensuring compliance with privacy regulations can help address these concerns.

The increasing digitization of health data makes it susceptible to cybersecurity threats, including data breaches and unauthorized access. Implementing robust cybersecurity measures, encryption protocols, and continuous monitoring can help safeguard health data from potential breaches.

The healthcare workforce may lack adequate training in cybersecurity, making them susceptible to social engineering attacks and unintentional security lapses. Providing ongoing cybersecurity training for healthcare professionals, administrators, and IT staff is essential to enhance awareness and strengthen the human element of cybersecurity.

Public health agencies and organizations may face budgetary constraints, limiting their ability to invest in the necessary infrastructure, technologies, and skilled personnel for effective data analytics implementation. Advocating for increased funding, exploring public-private partnerships, and prioritizing resource allocation for critical data analytics initiatives can help overcome resource limitations.

Initiatives may struggle with sustainability beyond initial funding periods, leading to challenges in maintaining and updating data analytics infrastructure. Incorporating sustainability plans into project designs, demonstrating the long-term value of data analytics, and exploring alternative funding sources contribute to ongoing success.

Addressing these challenges requires a concerted effort from stakeholders, including policymakers, healthcare providers, researchers, and technology experts. Overcoming these limitations is crucial to fully harness the potential of data analytics for improving public health outcomes. To address these barriers, federal policy should emphasize interoperability of health data and prioritize payment reforms that will encourage providers to develop data analytics capabilities.

## 2.7. Future Trends and Innovations in Data Analytics for Public Health

AI and machine learning will play a pivotal role in enhancing predictive modeling capabilities for disease surveillance (Alowais et al., 2023, Sawhney et al., 2023, Orikpete et al., 2023). These technologies can analyze complex patterns and large datasets to forecast disease outbreaks with greater accuracy. AI-driven algorithms will enable precision medicine by analyzing individual health data to tailor treatments based on genetic, lifestyle, and environmental factors. This personalized approach can improve treatment outcomes and reduce adverse effects. AI will be increasingly employed for automated pattern recognition in health data, aiding in the identification of subtle trends and correlations that may go unnoticed through traditional methods.

Early Warning Systems for Public Health Emergencies. Predictive analytics will evolve to create more robust early warning systems for public health emergencies (Khan et al., 2023, Ukoba et al., 2023). By analyzing a combination of clinical, environmental, and social data, these systems can provide timely alerts and recommendations for rapid response. Predictive analytics will move towards providing longitudinal health predictions, forecasting not only short-

term outcomes but also anticipating health trajectories over extended periods. This can support proactive and preventive healthcare strategies. Integration with Wearable and IoT Devices. Predictive analytics will increasingly integrate with wearable devices and the Internet of Things (IoT), enabling real-time monitoring of individual health parameters. This integration can enhance the accuracy of predictions and provide timely interventions. Real-Time Data Analytics for Rapid Response. Real-time data analytics will be employed for automated syndromic surveillance, allowing public health agencies to detect and respond to emerging health threats with unprecedented speed. Dynamic Resource Allocation in Public Health Emergencies. Real-time analytics will be utilized for dynamic resource allocation during public health emergencies. By continuously analyzing data on healthcare utilization, the allocation of resources can be adjusted in real-time to meet evolving needs.

Global Collaborative Platforms for Outbreak Response. Real-time data analytics will be facilitated through global collaborative platforms that allow for the rapid sharing of data and insights during outbreaks. These platforms will enable a coordinated and timely international response to emerging health crises. Blockchain for Secure and Transparent Data Sharing. Blockchain technology will be increasingly used to create secure and transparent data-sharing platforms. This can address privacy concerns and enhance trust among stakeholders involved in health data exchange. Interoperability Standards for Seamless Data Exchange. Efforts towards establishing interoperability standards will grow, enabling seamless data exchange between diverse healthcare systems. This will contribute to more comprehensive and integrated public health analytics. Collaborative data-sharing platforms will facilitate cross-sector collaboration, bringing together data from healthcare, social services, and other sectors. This integrated approach can provide a holistic understanding of health determinants.

These future trends and innovations in data analytics for public health signify a shift towards more advanced, responsive, and collaborative approaches. Embracing these developments holds the potential to significantly enhance the capabilities of public health systems in addressing emerging challenges and improving overall population health.

## 2.8. Policy Implications and Recommendations for Data Analytics in Public Health

Establishing a stringent Privacy Regulations. Governments should establish and enforce robust privacy regulations specifically tailored to health data. These regulations should outline clear guidelines for data collection, sharing, and storage while ensuring the protection of individuals' sensitive health information. Policymakers should encourage and incentivize healthcare organizations to adopt ethical data use practices. This includes obtaining informed consent, implementing de-identification techniques, and promoting transparency in data-sharing practices. There should be promotion of Cross-Border Data Governance. Collaboration between nations to create standardized cross-border data governance frameworks can ensure consistent privacy protections and facilitate the secure exchange of health data across international boundaries.

There should be standardization of data formats and terminologies. There should be development of National Data Standards. Governments should lead efforts to develop and enforce national data standards for health information, promoting uniformity in data formats and terminologies. This will enhance interoperability and facilitate seamless data exchange between healthcare systems. Policymakers should align national data standards with international frameworks to ensure compatibility and facilitate global collaboration in public health research and initiatives. Governments can provide incentives, such as grants or tax benefits, to healthcare organizations and technology vendors that adhere to standardized data formats and terminologies. This encourages widespread adoption and accelerates the transition to interoperable systems. More capacity building for public health professionals. By integrating data analytics training in public health education. Educational institutions and public health agencies should collaborate to integrate data analytics training into public health education programs. This ensures that future professionals are equipped with the skills needed for effective data-driven decision-making. There is need for continuous professional development programs. Governments and public health organizations should invest in continuous professional development programs focused on data analytics for existing public health professionals. This can include workshops, certifications, and online courses to keep professionals updated on evolving analytics techniques.

Governments can establish and support centers of excellence in data analytics for public health. These centers can serve as hubs for research, training, and the development of best practices in the application of data analytics to public health challenges. Governments should prioritize and allocate funding for the development and maintenance of health information technology infrastructure. This includes investing in electronic health records, data warehouses, and interoperability solutions. Policymakers should encourage and financially support research and development initiatives focused on advancing data analytics technologies for public health. This can drive innovation and the creation of cutting-edge tools and platforms. Governments can foster public-private partnerships to enhance technological infrastructure.

Collaborative efforts between public health agencies, private healthcare providers, and technology companies can accelerate the adoption of advanced data analytics capabilities.

Implementing these policy implications and recommendations can create a conducive environment for the responsible and effective use of data analytics in public health. By addressing privacy concerns, promoting standardization, building capacity, and investing in technological infrastructure, policymakers can foster a data-driven approach that improves public health outcomes while safeguarding individual privacy and data integrity.

## 3. Conclusion

In reviewing the landscape of data analytics in public health within the USA, several key findings emerge. The integration of data analytics has significantly transformed disease surveillance, intervention strategies, and policy formulation. Notable success stories, such as Project DETECT and Flu Near You, highlight the power of data analytics in providing real-time insights for effective public health responses. However, challenges, including data quality concerns, privacy issues, and resource constraints, underscore the need for a comprehensive and strategic approach.

The future of public health in the USA is intricately tied to the continued evolution of data analytics. Advancements in artificial intelligence, predictive analytics, and real-time data applications hold immense potential for improving health outcomes, enhancing early detection of outbreaks, and personalizing healthcare interventions. Standardizing data formats, ensuring data privacy, and building a skilled workforce will be crucial elements in shaping a future where data analytics becomes an integral part of public health practice.

Establish and enforce clear regulatory frameworks for data privacy to build trust and ensure responsible data use. Prioritize funding for health information technology infrastructure, capacity building, and research initiatives to advance data analytics capabilities. Embrace standardized data formats and terminologies to enhance interoperability and facilitate seamless data exchange. Invest in continuous training programs for healthcare professionals to ensure proficiency in data analytics applications.

Drive innovation in data analytics technologies while maintaining a focus on responsible and ethical data use. Develop solutions that support interoperability and collaboration across diverse healthcare systems. Integrate data analytics training into public health education programs to equip future professionals with essential skills. Conduct research to advance the understanding of the applications and implications of data analytics in public health. Engage in continuous professional development to stay abreast of evolving data analytics techniques and applications. Advocate for the adoption of data-driven practices within public health organizations to enhance decision-making and improve outcomes.

In conclusion, the future of public health in the USA hinges on the responsible and strategic integration of data analytics. Stakeholders must collaboratively work towards addressing challenges, promoting innovation, and fostering a culture where data is leveraged to its full potential. By doing so, we can usher in an era where data analytics becomes a cornerstone in shaping a healthier and more resilient nation.

## Compliance with ethical standards

*Disclosure of conflict of interest*

No conflict of interest to be disclosed.

## Reference

[1]    Aldahiri, A., Alrashed, B. and Hussain, W., 2021. Trends in using IoT with machine learning in health prediction system. *Forecasting*, *3*(1), pp.181-206.

[2]    Al Hossain F, Lover AA, Corey GA, Reich NG, Rahman T. FluSense: a contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020;4:1-28

[3]    Alowais, S.A., Alghamdi, S.S., Alsuhebany, N., Alqahtani, T., Alshaya, A.I., Almohareb, S.N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H.A. and Al Yami, M.S., 2023. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education*, *23*(1), p.689.

[4]   Bastani H, Drakopoulos K, Gupta V, et al. Efficient and targeted COVID-19 border testing via reinforcement learning. Nature 2021;599:108-113

[5]   Brownstein, J.S., Rader, B., Astley, C.M. and Tian, H., 2023. Advances in Artificial Intelligence for infectious-disease surveillance. *New England Journal of Medicine*, *388*(17), pp.1597-1607.

[6]   Choi, B.C., 2012. The past, present, and future of public health surveillance. *Scientifica*, *2012*.

[7]   Dantas LF, Peres IT, Bastos LSL, et al. App-based symptom tracking to optimize SARS-CoV-2 testing strategy using machine learning. PLoS One 2021;16(3):e0248920-e0248920.

[8]   Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. J Am Med Inform Assoc 2008;15:150-157.

[9]   Galetsi, P., Katsaliaki, K. and Kumar, S., 2019. Values, challenges and future directions of big data analytics in healthcare: A systematic review. *Social science & medicine*, *241*, p.112533.

[10]  Gardy, J.L. and Loman, N.J., 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics*, *19*(1), pp.9-20.

[11]  Guk, K., Han, G., Lim, J., Jeong, K., Kang, T., Lim, E.K. and Jung, J., 2019. Evolution of wearable devices with real-time disease monitoring for personalized healthcare. *Nanomaterials*, *9*(6), p.813.

[12]  Hwang, I.; Kim, H.N.; Seon, G.M.; Lee, S.H.; Kang, M.; Yi, H.; Bae, W.G.; Kwak, M.K.; Jeong, H.E. Multifunctional smart skin adhesive patches for advanced health care. *Adv. Healthc. Mater.* **2018**, *7*, 1800275

[13]  Jamshidi, M., Roshani, S., Daneshfar, F., Lalbakhsh, A., Roshani, S., Parandin, F., Malek, Z., Talla, J., Peroutka, Z., Jamshidi, A. and Hadjilooei, F., 2022. Hybrid deep learning techniques for predicting complex phenomena: A review on COVID-19. *AI*, *3*(2), pp.416-433.

[14]  Khan, M., Husnain, G., Ahmad, W., Shaukat, Z., Jan, L., Haq, I.U., Islam, S.U. and Ishtiaq, A., 2023. Performance evaluation of Machine Learning models to predict heart attack. *Machine Graphics and Vision*, *32*(1), pp.99-114.

[15]  Liang Z, Powell A, Ersoy I, et al. CNN-based image analysis for malaria diagnosis. In: Proceedings and Abstracts of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), December 15–18, 2016. Shenzhen, China: Institute of Electrical and Electronics Engineers, 2016.

[16]  Lim S, Tucker CS, Kumara S. An unsupervised machine learning model for discovering latent infectious diseases using social media data. J Biomed Inform 2017;66:82-94.

[17]  Liu D, Clemente L, Poirier C, et al. Real-time forecasting of the COVID-19 outbreak in chinese provinces: machine learning approach using novel digital data and estimates from mechanistic models. J Med Internet Res 2020;22(8):e20285-e20285.

[18]  Manrai, A.K., Cui, Y., Bushel, P.R., Hall, M., Karakitsios, S., Mattingly, C.J., Ritchie, M., Schmitt, C., Sarigiannis, D.A., Thomas, D.C. and Wishart, D., 2017. Informatics and data analytics to support exposome-based discovery for public health. *Annual review of public health*, *38*, pp.279-294.

[19]  Manohar, B. and Keerthana, C.H., 2023. Effective strategies for public health management using data and analytics. *Journal of healthcare and life-science research*, *2*(9), pp.83-93.

[20]  Mollalo A, Mao L, Rashidi P, Glass GE. A GIS-based artificial neural network model for spatial distribution of tuberculosis across the continental United States. Int J Environ Res Public Health 2019;16:157-157.

[21]  Mooney, S.J. and Pejaver, V., 2018. Big data in public health: terminology, machine learning, and privacy. *Annual review of public health*, *39*, pp.95-112.

[22]  Orikpete, O.F., Gungura, N.M., Ehimare, E. and Ewim, D.R.E., 2023. A critical review of energy consumption and optimization strategies in the Nigerian aviation sector: challenges and prospects. *Bulletin of the National Research Centre*, *47*(1), pp.1-14.

[23]  Pascucci M, Royer G, Adamek J, et al. AI-based mobile application to fight antibiotic resistance. Nat Commun 2021;12:1173-1173.

[24]  Pavuluri, K., Manoli, I., Pass, A., Li, Y., Vernon, H.J., Venditti, C.P. and McMahon, M.T., 2019. Noninvasive monitoring of chronic kidney disease using pH and perfusion imaging. *Science advances*, *5*(8), p.eaaw8357.

[25]  Petrasch, R.J. and Petrasch, R.R., 2022. Data integration and interoperability: Towards a model-driven and pattern-oriented approach. *Modelling*, *3*(1), pp.105-126.

[26] Ranasinghe, H.D., 2023. Strategic Utilization of Big Data and Analytics for Enhancing Public Health Responses to the COVID-19 Pandemic. *Emerging Trends in Machine Intelligence and Big Data*, *15*(10), pp.1-15.

[27] Randhawa GS, Soltysiak MPM, El Roz H, de Souza CPE, Hill KA, Kari L. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. PLoS One 2020;15(4):e0232391-e0232391.

[28] Razzak, M.I., Imran, M. and Xu, G., 2020. Big data analytics for preventive medicine. *Neural Computing and Applications*, *32*, pp.4417-4451.

[29] Reich NG, McGowan CJ, Yamana TK, et al. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. PLoS Comput Biol 2019;15(11):e1007486-e1007486

[30] Sawhney, R., Malik, A., Sharma, S. and Narayan, V., 2023. A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease. *Decision Analytics Journal*, *6*, p.100169.

[31] Shafqat, S., Kishwer, S., Rasool, R.U., Qadir, J., Amjad, T. and Ahmad, H.F., 2020. Big data analytics enhanced healthcare systems: a review. *The Journal of Supercomputing*, *76*, pp.1754-1799.

[32] Singhal, A., Baxi, M.K. and Mago, V., 2022. Synergy Between Public and Private Health Care Organizations During COVID-19 on Twitter: Sentiment and Engagement Analysis Using Forecasting Models. *JMIR Medical Informatics*, *10*(8), p.e37829.

[33] Song, M., Bai, H., Zhang, P., Zhou, X. and Ying, B., 2023. Promising applications of human-derived saliva biomarker testing in clinical diagnostics. *International Journal of Oral Science*, *15*(1), p.2.

[34] Sundermann AJ, Chen J, Kumar P, et al. Whole genome sequencing surveillance and machine learning of the electronic health record for enhanced healthcare outbreak detection. Clin Infect Dis 2021 November 17 (Epub ahead of print).

[35] Tahamtan, A. and Ardebili, A., 2020. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert review of molecular diagnostics*, *20*(5), pp.453-454.

[36] Thacker, S.B. and Berkelman, R.L., 1988. Public health surveillance in the United States. *Epidemiologic reviews*, *10*(1), pp.164-190.

[37] Tetyana, P., Shumbula, P.M. and Njengele-Tetyana, Z., 2021. Biosensors: design, development and applications. In *Nanopores*. IntechOpen.

[38] Ukoba, K., Kunene, T.J., Harmse, P., Lukong, V.T. and Chien Jen, T., 2023. The Role of Renewable Energy Sources and Industry 4.0 Focus for Africa: A Review. *Applied Sciences*, *13*(2), p.1074.

[39] Velasco, E., Agheneza, T., Denecke, K., Kirchner, G. and Eckmanns, T., 2014. Social media and internet-based data in global systems for public health surveillance: a systematic review. *The Milbank Quarterly*, *92*(1), pp.7-33.

[40] Yao, H.; Shum, A.J.; Cowan, M.; Lähdesmäki, I.; Parviz, B.A. A contact lens with embedded sensor for monitoring tear glucose level. *Biosens. Bioelectron.* **2011**, *26*, 3290–3296