PREDICTING DISEASE OUTBREAKS USING AI AND BIG DATA: A NEW FRONTIER IN HEALTHCARE ANALYTICS

Sanjay Ramdas Bauskar[1*], Chandrakanth Rao Madhavaram[2],Eswar Prasad Galla[3], Janardhana Rao Sunkara[4], Hemanth Kumar Gollangi[5]

Abstract

Disease forecasts how many people can be infected and die if no medicines and vaccines have been issued and offered to the population free of charge; if less than 70% of the population takes them, many will die in a short time in an outbreak. Many times, a few pockets of unvaccinated individuals will remain, thus necessitating 100% global vaccination. The risk can then lead to the development of appropriate overall clinical care and critical care for the first wave of people caught up in the outbreaks. The global distribution of resources, including such basic goods, is not in question in this short essay. A forecast for more than 300 diseases is possible using human biology, dry laboratory work, and artificial intelligence. However, this predictive health care with precise timing for the start of the disease is still experimental, due to the lack of financial support for such research. The "frontiers of science and medicine" are often overlooked by university and government agencies when they do not have financial means or thought leaders to enforce their use.This essay addresses the use of AI and Big Data in predicting disease outbreaks in a given area of a country or globally. This is a relatively new area for healthcare analytics, an expansion of what is today an extremely important component of the field. The predictive advantage of such a tool is that it will enable the world's governments to pre-order vast quantities of vaccines and antivirals once a forecast is made, thereby protecting the planet from the new contagious disease. Currently, a major disease outbreak and local epidemics can be contained if full doses of these products are delivered to 70% of the world's population within 20 to 30 days of the earliest clinical symptoms of infection. These pre-ordered vaccines and drugs can be kept in a chest somewhere in each country and region, with a "best before" date of three to four years. Thus, this is a project to protect the lives of all people in the world.

Keywords: Disease outbreak prediction, AI in healthcare, Big data analytics, Healthcare analytics, Epidemic forecasting, Predictive modeling, Machine learning in epidemiology, Data-driven disease prediction, AI-driven healthcare solutions, Health informatics, Pandemic prediction, Data science in public health, Risk assessment algorithms, Real-time outbreak monitoring, Healthcare data integration, Predictive analytics in medicine, Epidemiological data analysis, Big data in disease control, Artificial intelligence healthcare applications, Outbreak simulation models.

[1*]Pharmavite LLC, Sr. Database Administrator, sanjayramdasbauskar@outlook.com
[2]Microsoft Support Escalation Engineer, Craoma101@outlook.com
[3]Dept. of Comp. Sci. Univ. of Central Missouri, EswarPrasadGalla@outlook.com
[4]Siri Info Sol. Inc. Sr. Oracle DB Admin, JanardhanaRaoSunkara@outlook.com
[5] KPMG Consultant, HemanthKumarGollangi12@outlook.com

***Corresponding Author:** Sanjay Ramdas Bauskar
*Pharmavite LLC, Sr. Database Administrator, sanjayramdasbauskar@outlook.com

DOI: 10.53555/ecb.v11:i12.17745

1. Introduction

Epidemics and natural catastrophes occur frequently, with the more significant losses being attributed to neglected populations that are primarily situated in flood plains and fault lines. Till now, the healthcare sector lacked tools to anticipate when and where epidemics will erupt. Similarly, the power businesses do not possess the respective attribute to anticipate when and where the next sprawl of upsurge in healthcare spending will materialize. With the advancements in sensing, AI, machine learning, and data analysis, these health industry laggers are developing new big data platforms to catch up with the spree of technology spearheads in the banking and other consumer-facing verticals. How these technologies are being used to better understand and forecast disease outbreaks in these systems is what this essay will explore.In the early 2000s, SARS was among the first diseases to pioneer the then-dubious advancements of artificial intelligence (AI) and big data. With the COVID-19 pandemic, the progression and acceptance of these technologies picked up speed. Yet, while these subjects are currently center stage at gatherings from Silicon Valley to Davos, only a few discussions are devoted to a space where the use of analytics has had significant effects. In thematic sectors such as disease outbreak and preparedness, AI, big data, and predictive analytics are beginning to define a distinct set of vertices.

Epidemics and natural disasters disproportionately impact neglected populations situated in vulnerable areas, such as flood plains and fault lines, yet the healthcare sector has historically lacked the tools to predict when and where these crises will arise. Similarly, power companies struggle to anticipate surges in healthcare spending, highlighting a significant gap in predictive capabilities. However, recent advancements in sensing technologies, artificial intelligence (AI), machine learning, and data analytics are enabling the health industry to catch up with technology leaders in banking and consumer services. This essay delves into how these innovations are transforming the understanding and forecasting of disease outbreaks. The early 2000s saw SARS as a catalyst for the integration of AI and big data in health, but it was the COVID-19 pandemic that truly accelerated the adoption of these technologies. Despite their increasing prominence in discussions from Silicon Valley to Davos, the substantial impacts of analytics in sectors like disease outbreak and preparedness remain underexplored. As AI and predictive analytics carve out new pathways, they are beginning to redefine how we approach public health challenges and resource allocation in times of crisis.

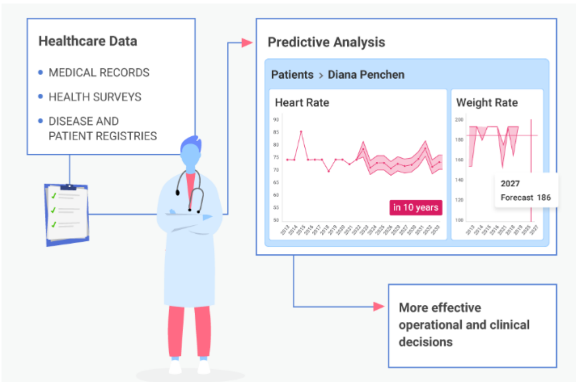

Fig 1 : Predictive Analytics In Healthcare

1.1. Background and Rationale

While standard approaches to diagnose infectious illnesses and predict their outbreaks exist, they have limitations. Moreover, non-infectious illnesses such as hypertension, obesity, coronary diseases, type 2 diabetes mellitus, and others have become a major addition to the disease bouquet and are also included in the factors contributing to high disease mortality. With the growing prevalence of significant diseases even in the richer and educated general community, one approach to decrease their condition prevalence in a country and decrease cure and therapy cost may be to try to anticipate them and begin management on a preventive path, if suitable in various ways. Public health surveillance systems may be required for infectious disease monitoring, intervention, and mitigation.Thanks to breakthroughs in modern data analytics techniques, the precision of predictive analytics, such as forecasting epidemics and other health occurrences, especially in informal data-rich conditions, has surpassed the potential of early warning devices for forecasting epidemics and other health occurrences. With the growing availability and prevalence of both big data and data analytics instruments and tactics in both academia and industry, the quality gap between the three study sites should be diminished to guarantee the generalizability of the outcomes. In order to identify the constraints of our study, an examination of the relative evidence published in the scope of informatics in the health sector and public health surveillance is necessary.In public health, non-infectious diseases are increasingly contributing to the general disease burden, which has resulted in numerous fatalities and an overwhelming demand for treatments. In order to assist public health authorities in rapidly detecting and controlling an outbreak, advanced techniques such as AI and Big Data must be used to forecast future disease outbreak trends. Public health officers aim to anticipate and control a future outbreak of any disease before it can cause any harm to citizens. The various possibilities that AI and Data Science bring to the table can be discussed in this article.

Equ 1: Basic Reproduction Number

$$\frac{\partial E_S}{\partial t} = \underbrace{\Gamma}_{\substack{\text{influx of fresh}\\ \text{environments}}} - \underbrace{\lambda E_S H_I}_{\substack{\text{environment}\\ \text{infection}}} - \underbrace{\delta E_S}_{\substack{\text{environment}\\ \text{decay}}} \qquad [I]$$

$$\frac{\partial E_I}{\partial t} = \underbrace{\lambda E_S H_I}_{\substack{\text{environment}\\ \text{infection}}} - \underbrace{\delta E_I}_{\substack{\text{environment}\\ \text{decay}}} \qquad [II]$$

$$\frac{\partial H_S}{\partial t} = \underbrace{g(H_I + H_S)\left(1 - \frac{H_I + H_S}{K}\right)}_{\text{logistic growth of the host population}} - \underbrace{\omega H_S E_I}_{\substack{\text{host}\\ \text{infection}}} - \underbrace{m H_S}_{\substack{\text{host}\\ \text{death}}} \qquad [III]$$

$$\frac{\partial H_I}{\partial t} = \underbrace{\omega H_S E_I}_{\substack{\text{host}\\ \text{infection}}} - \underbrace{m H_I}_{\substack{\text{host}\\ \text{death}}} \qquad [IV]$$

$$\frac{\partial E_S}{\partial t} = \underbrace{\Gamma}_{\substack{\text{influx of fresh}\\ \text{environments}}} - \underbrace{\lambda E_S H_I}_{\substack{\text{environment}\\ \text{infection}}} - \underbrace{\delta E_S}_{\substack{\text{environment}\\ \text{decay}}} \qquad [I]$$

$$\frac{\partial E_I}{\partial t} = \underbrace{\lambda E_S H_I}_{\substack{\text{environment}\\ \text{infection}}} - \underbrace{\delta E_I}_{\substack{\text{environment}\\ \text{decay}}} \qquad [II]$$

$$\frac{\partial H_S}{\partial t} = \underbrace{g(H_I + H_S)\left(1 - \frac{H_I + H_S}{K}\right)}_{\text{logistic growth of the host population}} - \underbrace{\omega H_S E_I}_{\substack{\text{host}\\ \text{infection}}} - \underbrace{m H_S}_{\substack{\text{host}\\ \text{death}}} \qquad [III]$$

$$\frac{\partial H_I}{\partial t} = \underbrace{\omega H_S E_I}_{\substack{\text{host}\\ \text{infection}}} - \underbrace{m H_I}_{\substack{\text{host}\\ \text{death}}} \qquad [IV]$$

1.2. Research Aim and Objectives

Prior studies have presented a diverse value of dengue hazards, contemporary and past hums of value to look at. There is only one study with a retrospective and present value forecasting model of dengue in the study of Ng et al. aimed for 4-6 weeks of dengue outbreak predictions with the use of AI and big data techniques. The prediction values set in the study are roughly about 70-80%, with low MASE value, and higher scores from sensitivity, precision, specificity, and accuracy. Additionally, this study validates a new dataset of variables of dengue outbreak prediction which have normal BMI, malnourished, and obese dataset values to precede study efficiency. This is the first evidence-based documentation for the outflow of dengue outbreak forecast using alternative variants of models.

The world measures the quality of research by the aims and objectives the researcher set in the beginning. The research aimed at proposing a conceptual framework for the early prediction of the chance outbreaks of dengue by examining the demographics of a specific area for low resource configuration. The proposed framework included the details of the data pre-processing, the data sampling, and the algorithm of haze detection used in the forecasting module. Section 1 raised the intuition of the study. Objective 1 indicated that the research aimed to propose an initial framework for early prediction of dengue outbreaks using big data analytics methods for low-resource areas. Given the current state of research studies, it also proposed some initial hypotheses that may be tested in future studies. Therefore, the proposed study aimed to fill this gap and to answer the following research questions.Prior research on dengue outbreak forecasting has explored a range of methodologies, with a notable study by Ng et al. utilizing AI and big data techniques to predict dengue outbreaks 4-6 weeks in advance. This study achieved prediction accuracy between 70-80% and demonstrated low MASE values alongside high scores in sensitivity, precision, specificity, and overall accuracy. It also introduced a novel dataset incorporating variables like BMI categories—normal, malnourished, and obese—to enhance prediction efficiency, marking a significant advancement in dengue forecasting. The research aimed to establish a conceptual framework for early dengue outbreak prediction, especially tailored for low-resource settings. This framework detailed data pre-processing, sampling, and haze detection algorithms employed in the forecasting model. Section 1 of the study articulated the foundational intuition and objectives, focusing on leveraging big data analytics for early dengue prediction and proposing hypotheses for future validation. By addressing this research gap, the study sought to offer a comprehensive approach to predicting dengue outbreaks, potentially guiding future investigations and interventions.The study by Ng et al. represents a significant advancement in dengue outbreak forecasting by leveraging AI and big data techniques to predict outbreaks 4-6 weeks in advance with notable accuracy between 70-80%. This research achieved low MASE values and high sensitivity, precision, specificity, and overall accuracy, highlighting its robustness. A key contribution of the study is its novel dataset that incorporates variables such as BMI categories—normal, malnourished, and obese—thus refining prediction efficiency. The research aimed to establish a conceptual framework for early dengue prediction, particularly for low-resource settings, detailing the processes of data pre-processing, sampling, and haze detection algorithms. By addressing a critical research gap, the study not only proposes a new predictive framework but also sets the stage for future hypotheses and investigations, enhancing the ability to anticipate and mitigate dengue outbreaks effectively.

2. The Role of AI and Big Data in Healthcare Analytics

AI and big data have disrupted various industries with their potential to understand detailed real-time data and derive useful insights from them. They are found to be very useful in the field of healthcare analytics, especially in predicting the next public health emergency. Today's machine learning techniques can analyze vast amounts of structured and unstructured data to predict how diseases evolve and spread, understand their ramifications, and customize treatments for patients. This results in an estimated economic impact of AI ranging from $7,740 billion to $15,770 billion in 2030.So, how have AI and big data proven to be so beneficial in the field of healthcare analytics, and what are the best examples of these technologies predicting impending health crises or being especially helpful in battling them? The many functions of AI and big data in healthcare analytics have made the field ripe for extraordinary growth. These technologies hold a lot of potential in predicting outbreaks, analyzing genomes and diseases, and managing medical records, among other things. AI and big data are used to make policy decisions, prioritize emergency funds, and minimize loss to economies from disruptions like travel restrictions, business closures, or supply chain breakdowns.

Apart from real-time monitoring and disease modeling, AI is also used in patient monitoring at the Mayo Clinic to monitor patients hospitalized with COVID-19. This method is safe and doesn't necessarily put more strain on other healthcare professionals. It's a process that can't be very easily automated, but it's certainly not without the help of AI and healthcare analytics. Data from CT scans or X-rays are analyzed to see who is more prone to the severity of the disease. It can, to an extent, help frontline healthcare professionals know who to prioritize, especially if healthcare resources are particularly strained.AI and big data have revolutionized healthcare analytics by enhancing predictive capabilities and optimizing patient care. These technologies excel in analyzing vast amounts of structured and unstructured data, offering invaluable insights into disease progression and outbreak forecasting. For instance, AI algorithms can sift through diverse datasets, including real-time health records, genomic data, and social media trends, to predict potential public health emergencies with remarkable accuracy. A notable example is the use of AI at the Mayo Clinic, where it assists in monitoring COVID-19 patients by analyzing CT scans and X-rays to predict disease severity. This approach helps prioritize patient care, especially in overwhelmed healthcare settings, by identifying individuals at higher risk and enabling more efficient resource allocation. The economic impact of AI in healthcare, projected to reach between $7,740 billion and $15,770 billion by 2030, underscores its potential to transform the field by improving decision-making, managing resources, and mitigating the economic consequences of health crises.AI and big data have profoundly transformed healthcare analytics by enhancing predictive capabilities and optimizing patient care. These technologies excel in processing vast amounts of both structured and unstructured data to provide deep insights into disease dynamics and outbreak forecasting. For example, AI algorithms can analyze diverse data sources, including real-time health records, genomic information, and social media trends, to predict potential public health emergencies with impressive accuracy. A prominent case is the Mayo Clinic's use of AI to monitor COVID-19 patients; the technology analyzes CT scans and X-rays to assess disease severity, enabling healthcare professionals to prioritize care and allocate resources more effectively in strained settings. This innovative approach not only improves patient outcomes but also highlights the transformative economic potential of AI in healthcare, with estimates suggesting an impact ranging from $7,740 billion to $15,770 billion by 2030.

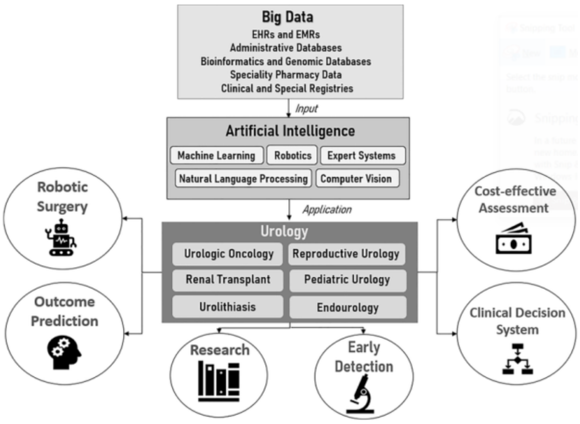

Fig 2: Role of AI and Big Data in Healthcare

2.1. Overview of AI and Big Data Technologies

Technically, AI seeks to create systems that can learn and adapt, making decisions based on data, while typically having very complex relationships and features. One of the most powerful tools of AI is deep learning, which identifies and consumes vast amounts of structured and unstructured data and makes informed predictions. It is mostly used to recognize objects in images, the semantic meaning of text (natural language processing), and the sound of spoken words (speech recognition). Machine learning, the study, and construction of algorithms that can learn from and make predictions on data, is central to predicting outbreaks. This type of AI is often used – and recommended by global health experts – for this kind of study on data from previous outbreaks to distinguish between different patterns and disease dynamics. Big data analysis is premised on a systematic approach to collecting, processing, and interpreting structured and unstructured data affordably and efficiently. In healthcare, the application of big data extends beyond analyzing data from individual sources to data mining to uncover hidden patterns, while re-contextualizing data to provide tools for recognizing symptoms and finding solutions to improve healthcare services. It is clear that AI and big data offer a new perspective on dealing with large amounts of data and identifying hidden patterns that the healthcare industry has not previously encountered. This includes tailoring diagnoses, treatments, and care programs to suit each individual, based on data.AI and big data offer exciting new opportunities for providing better healthcare to people all over the world. AI applications in healthcare are wide-ranging and include robot-assisted surgery, virtual nursing assistants, clinical trials, drug research and development, aided by AI agents, electronic health records, monitoring systems, hospital management systems, drug interaction, and precision medicine. A key driver for these applications is big data, that is, the enormous amount of healthcare data generated and recorded every day.

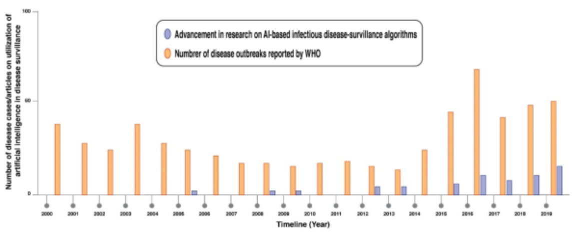

Fig: Advancement of AI-based infectious disease-surveillance algorithms

2.2. Applications in Disease Surveillance and Outbreak Prediction

In the healthcare and public health sector, the number of use cases that leverage AI (artificial intelligence) and big data has been growing rapidly, including practice at the point of care (e.g., personalized medicine and genomics studies), health-related research (e.g., precision public health), and public health programs (e.g., disease surveillance and event detection). Research shows there is a growing need for global surveillance systems to share information across national, regional, and global boundaries to detect outbreaks before they become global pandemics and to develop appropriate countermeasures against the pathogens that cause them. A number of projects have demonstrated the possibility of integrating, analyzing, and utilizing this diverse set of data to yield early detection of historic infectious disease outbreaks. Would this data combination work in detecting novel threats to the U.S.? What are the biological, social, and communications chokepoints that could be targeted for distribution in a bioterrorism event?

Applications in disease surveillance and outbreak prediction. Early detection is critically important to curtail economic and human losses due to disease outbreaks. Currently, the healthcare industry has begun a data-inspired change where vast amounts of data regarding disease and outbreak events are being accumulated by researchers, epidemiologists, and healthcare professionals across the globe. However, the sheer magnitude of widespread global data collection poses a unique and complex problem characterized not only by data abundance but also by data heterogeneity, time lags, biases, and other sources of noise and misinformation. With the additional challenge of emerging and reemerging diseases in animals and wildlife, controlling and preventing an outbreak continues to require methods that can integrate diverse sources and multiple streams in real-time for accurate event detection. Many existing surveillance and public health tools have been based on the growing abundance of data from electronic health records (EHRs), telecommunication data, internet search queries, and social media posts, as well as informal communication or communication in petit networks where confidentiality is an issue.

3. Challenges and Limitations

Much current research has not only concentrated on health monitoring systems that employ AI and massive data of human blood and different measures but has also highlighted the processing methods and evidence utilized to understand and forecast risks or opportunities within populations. In conclusion, experts' apparent discovery of efficient AI and big data processing capabilities for forecasting illness outbreaks depends on the systematic processing of available population data. However, there are many challenges to effective implementation. An ethics barrier is first and foremost. Given the perilous political circumstances, leaders and civil society are universally aggravated. Finally, the feasibility of the data is essential for the AI advancement project. The effort of preparing the data requires time and cost for implementation, execution, and explanation, given that a lot of measures need to be transferred to generate the outcomes, thereby extending the control loop of illness prevention. Disasters are unpredictable events that transcend random occurrences. As the world continues to face a wider range of tragedies owing to the COVID-19 epidemic and other emergent diseases, an increasing number of studies have relied on modern advances in artificial intelligence (AI) and large data sets to forecast and alert the population to disorder outbreaks. However, global AI development is still in its early stages, and it is not yet uniformly common. This raises issues about moral risks and the accuracy of such improvements. Even if AI and Big Data can be used to anticipate illness outbreaks, there is still an ethical barrier to identifying possible danger areas due to privacy. Current research has increasingly focused on health monitoring systems that leverage AI and large datasets derived from human blood measures and other health indicators to understand and predict risks within populations. While experts have made significant strides in utilizing AI and big data for forecasting disease outbreaks, the success of these initiatives hinges on the systematic processing of available population data. However, several challenges hinder effective implementation, with ethical concerns at the forefront. The precarious political climate often exacerbates tensions between leaders and civil society, complicating data usage. Additionally, the feasibility of data for AI projects poses significant hurdles; the time and costs associated with preparing and processing this data can be substantial, often extending the timeline for effective disease prevention. As the world grapples with an increasing frequency of disasters, spurred by the COVID-19 pandemic and emerging diseases, reliance on AI and big data for early warning systems has grown. Yet, the uneven development of AI capabilities raises critical questions about moral implications and the accuracy of predictive models, particularly concerning privacy issues when identifying at-risk areas. Thus, while these technologies hold great promise, ethical barriers must be addressed to ensure responsible and effective utilization.

Fig 3 : Challenges

3.1. Ethical and Privacy Concerns

Although the positive effects of utilizing AI and big data to predict highly contagious outbreaks significantly limit the risks and improve the effectiveness of outbreak management, it is critical to take many ethical considerations into account before deploying these systems. Moreover, forecasting broad trends about possible outbreak locations and timelines may lead to greater social benefits by allowing early, state-to-state warning and coordination. As such, the tradeoff between potential benefits and risks must be carefully considered and enable flexible, adjustable systems. Consideration should be given to providing feedback loops that allow false alarms to be resolved and for information about true alarms to be provided in a rapid and fair manner. Further, warnings should be matched with resources and realistic plans to manage the information provided and illness reports sure to follow a large-scale warning; disasters are not a joke. Decision makers, the public, and human subjects should be provided with enough information to understand the natural capabilities and limitations of any warning system to make an informed decision about the risks and benefits of participation.

Risk assessments should seriously consider and address direct and indirect concerns about the potential impact of the warning system on the local and global economy, individual search behavioral modification (e.g., increased healthcare-seeking), and public behavioral modification (e.g., fear and avoidance of reported locations, firms, products, and/or markets).Opportunities must also be taken to offer authentic informed consent for participation in any big data outbreak surveillance or warning research across all forms of data when these data have been previously collected and stored; when they start to be collected and stored for use in warning; and particularly when they are collected specifically for warning system development on human subjects who might be noted to be in-vivo during the R&D phase. Mistakes and regulations governing privacy practices must be guided by solid ethical considerations that are up to the times in our fast-forward world. Technology limits and robust plug-in—always knowing laws (and showing them in privacy impact assessments) must underscore and make possible applications suggested by rapid methodological advancements in topic prediction, classification, and warning systems. Social assessment impacts associated with building effective big data forecasts and AI-based warning systems for disease transmission ahead of time must be studied and made publicly available. Given the potential inaccuracies of current high-speed long-distance bio-surveillance using AI technologies, ethical considerations about the use of big data, e.g., from international file-sharing of individual travel records, phone records, or commerce transactions, need to be made. When reported illness cases are used to forecast geographic spread, timing, and relative intensity of emerging infectious diseases prior to the availability of evidence, warning timing and level must be based on the best available judgment from surveillance data at the time combined with evidence from warning errors previously documented and real-time analytical judgment. For these reasons, big data and AI health surveillance researchers should provide advanced, in-depth ethical reasoning when seeking approval from institutional review boards or equivalents for any project domain in vivo. While the application of AI and big data in predicting contagious outbreaks offers substantial benefits for managing and mitigating risks, it is essential to address a range of ethical considerations before deployment. The potential for AI-driven forecasts to significantly improve outbreak management and coordination between states must be balanced against the risks of false alarms and the subsequent social and economic impacts. It is crucial to implement feedback mechanisms to resolve false positives and provide timely, accurate information for true alarms, ensuring that warnings are matched with appropriate resources and realistic response plans. Furthermore, transparency about the capabilities and limitations of warning systems is vital for informed decision-making by the public and stakeholders. Risk assessments should address both direct and indirect impacts, such as changes in healthcare-seeking behavior and public reactions to warnings, which could affect local economies and individual behaviors. Authentic informed consent must be obtained for the use of data in surveillance and warning systems, with adherence to current privacy regulations and ethical standards. Social impact assessments should accompany the development of AI-based forecasting tools, and any use of international data must be carefully regulated. Given the potential inaccuracies of current technologies, the best available judgment and evidence should guide warning decisions, with robust ethical considerations incorporated into the approval process for research projects.

Equ 2: An adaptive social distancing SIR model

Input: $N,\ \beta,\ \gamma,\ \sigma,\ p,\ \tau,\ n$

Output: $S,\ I,\ R,\ D$

$S(0) \leftarrow N-1,\ I(0) \leftarrow 1,\ R(0) \leftarrow 0,\ D(0) \leftarrow 0;$

$tol \leftarrow 10^{-6};$

for $k \leftarrow 0$ ***to*** $n-1$ **do**

  Calculate $\Theta 0 \leftarrow \frac{N}{S(k)}\mathcal{L}\big(k,\ \frac{S(k)}{N}\big);$

  $Err \leftarrow 1;$

  while $\big(Err < tol\big)$ **do**

$$z_1 \leftarrow \tau(\gamma+\sigma)+1-\tau\frac{\beta}{N}(S(k)+I(k))\Theta_0;$$
$$z_2 \leftarrow \tau\tfrac{\beta}{N}(\tau(\gamma+\sigma)+1)\Theta_0;$$
$$I(k+1) \leftarrow \frac{\sqrt{z_1^2+4\,z_2\,I(k)}-z_1}{2z_1};$$
$$S(k+1) \leftarrow \frac{S(k)}{1+\tau\frac{\beta}{N}\Theta I(k+1)};$$
$$R(k+1) \leftarrow R(k)+\tau\gamma I(k+1);$$
$$D(k+1) = N-S(k+1)-I(k+1)-R(k+1);$$
$$\Theta \leftarrow \tfrac{N}{S(k+1)}\mathcal{L}\big(k+1,\ \tfrac{S(k+1)}{N}\big);$$
$$Err \leftarrow abs(\Theta 0-\Theta);$$
$$\Theta 0 \leftarrow \Theta;$$

  end

end

3.2. Data Quality and Accessibility

Real-time access to credible information can bridge the gap created by misinformation and fake news during a time of emergency; for example, the COVID-19 crisis elevated the importance and widespread recognition of reliable information in service of its distribution to the public. Both of these factors are fundamental in creating transparency for the data owner, who can leverage this knowledge to help build a new suite of data sharing with partners. In a disaster like Ebola, governments were known to ritualistically degrade and withhold relevant information for years (BBC 2014). Besides the advent of suitable AI approaches, the global dissemination of trustworthy, accessible data is essential for epidemic responses. The danger of AI models that use location data is that they will flag separate geographic areas either as hotspots or cold alerts, stating risk in imperial reds and safe zones colored locked green. For them, it is harder to distinguish social distance because a crowded room will harbor both the hunter and the victim. The model should work in real-time in any situation that includes spatial density, for if not instructed to assess an outbreak, it will naturally sense the ambiance and assess the true risk in that area.

While the pervasiveness of big data is now a moot point, it's the data quality that is considered a cause of concern. For any disease forecasting model designed to predict outbreaks and patterns, the data used for the model training has to be precise, reliable, and follow ethical and privacy standards. The World Health Organization (WHO 2020) defines data quality as data that are fit for purpose. It therefore becomes essential that data are collected consistent with appropriate standards. This is because the insights obtained from a model can vary greatly depending on the uncertainty and noise levels within the data. Furthermore, models trained on imperfect or unreliable data can result in major errors in forecasting, and therefore should be carefully labeled by trusted sources that hold accountable the authenticity of the records. Furthermore, the data provided by government bodies, and medical institutions in low-middle-income (LMI) countries is scarce and can be biased toward urban living and males, while largely leaving minority and rural female patients unacknowledged. To ensure the quality of the predictions, the magnitudes and the variances associated with bias, noise, and outlying values in the data must be considered during the model-building process. In times of crisis, such as the COVID-19 pandemic, real-time access to credible information has proven essential for combating misinformation and ensuring transparency. Effective data sharing and reliable communication can significantly enhance epidemic responses and public trust. For instance, during the Ebola outbreak, delayed and restricted information from governments hindered effective responses, underscoring the need for immediate and accurate data dissemination. AI models, while powerful, must be designed to account for spatial density and social interactions, as simplistic location-based risk assessments may overlook critical nuances. The quality of data used to train these models is crucial; it must be accurate, reliable, and adhere to ethical standards. According to the World Health Organization, data quality is defined as being fit for its intended purpose, which highlights the importance of precise and consistent data collection. Inaccurate or biased data can lead to flawed predictions and misinformed responses, particularly in low-to-middle-income countries where data availability may be limited and skewed. Addressing biases and ensuring comprehensive, representative data is essential for building effective forecasting models and making informed decisions during health crises.

4. Case Studies and Success Stories

Several platforms designed for infectious disease surveillance development have been examples of international frameworks. This type of system aims to provide the 'big picture' of infectious disease trends to complement other internet-based disease surveillance systems, widely known as syndromic surveillance systems. The increasing amount of interest in this type of internet-based data is underlined by the article of Olson et al., who presented an overview of automated disease surveillance and identified 762 systems in operation around the world as of June 2016. They classified 156 systems as event-based and 606 as indicator-based. Of these, 244 were used every week for public health tracking or to inform actual event monitoring (35 event-based and 209 indicator-based) in areas such as weekly influenza-like illness, animal health, emergency room visits, lab-based reporting of infectious diseases, open-source media, outpatient visits, participatory systems detecting symptoms or animal die-offs, disease diagnoses by autopsy, fiscal data, and mental health. Before 2009, there were only 14 systems. In this survey, we have looked at a few case studies that outline how the tools and technologies that are available today can be leveraged to effectively predict and manage disease outbreaks. This section briefly reports these case studies and success stories. In summary, the methodological choices made by these platforms can be introduced effectively in technology-scarce and human resource-scarce environments. The choice of technology should ideally follow the pressures of data access and availability, as well as human expertise and the setup and long-term strategic approach of the organization for which it is developed and deployed.

Fig 4: AI in Healthcare: Real-World Success Stories and Case Studies

4.1. Examples of Disease Outbreak Prediction Systems

1. Argus. The Argus project, developed with funding from the European 7th Framework Program, aimed to develop an early warning system for the detection, monitoring, or forecasting of health risks derived from novel pathogens or environmental changes that affect communicable diseases.

Argus "was built on the reasoning that the current requirements on new functions of modern epidemiology in the Global Health domain cannot be managed relying solely on traditional, descriptive methods. New approaches to Health Surveillance, dealing with the dynamics, complexity, and high rate of changing conditions of modern societies, often enforced by globalization, and enabling the proactive prediction and detection of events that can endanger human health, need to be developed". This project explored a number of different types of data, including notices, reports, trained on disease outbreak data from 12 countries in order to predict future outbreak signals for 49 infectious diseases. The system's architecture used Big Data tools (e.g. ETL, Hadoop), deriving Big Data storage and data processing.

2. HealthMap. This is one of the earliest AI-based initiatives used for the early detection of outbreaks, beginning in 2006. HealthMap was designed to address the "problem of sifting through a daily deluge of health information to identify and respond to events of significance". The aim of HealthMap is to help in the early detection of a disease outbreak, as well as provide insight into where and why the outbreak occurred. It searches, aggregates, filters, and visualizes various data and alerts, to build statistical and machine-learning models

sources in multiple languages, and has been used to identify outbreaks such as H1N1 and H5N1. More than 65,000 users access HealthMap each month. The health information comes from a range of resources such as lay-public and news sources, as well as more official sources such as discussion forums, mailing lists, and other content from the group of Promed partners, which includes ministries of health around the world. HealthMap uses various NLP tools for filtering information and discovering anomalies. geospatial referencing, from which data mining of search query patterns helps to identify hot spots in real-time as a possible outbreak.

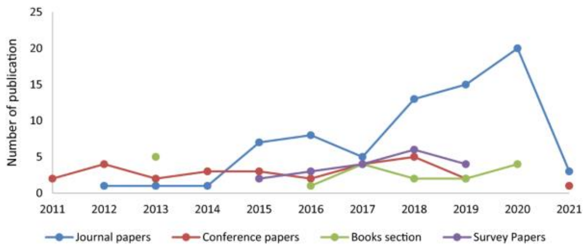

Fig : Evolution of final database by type of article and year

5. Future Directions and Implications

It is crucial to combine two or more variables for algorithms to learn because it adds predictive power to learn about the variables and outcomes. Disease prediction models will improve, with a growing trend to collect data to deal with disease maintenance, through these combinations and leverage an experience portfolio that may otherwise go unnoticed and may consequently affect the accuracy of predictions. The growing interest in healthcare analytics is that organizations are now moving from not only understanding what happened in the past, to using descriptive and diagnostic (what and why anything happened in the past) and predictive (to predict what will happen in the future) analytic tools into the realm of prescribing behaviors that will lead to certain outcomes. The use of predictive analytic tools is highly appropriate for the exploration of future events, based on historical data that inform certain outcomes. The future directions in this study include the need to start predicting various other diseases, which is crucial in implementing proactive public health interventions and policies. This may involve learning automated ways to detect and predict various aspects of communicable and noncommunicable diseases nationwide. The increasing use of AI for predicting the uptake of preventive measures will strongly reflect the growing interest of people in adopting these preventive strategies in time to prevent themselves from being infected. However, it becomes important for us to predict the impact of AI in causing panic in the healthcare system if the AI algorithm predicts a high healthcare consulting rate to avoid overstraining the system. The future implications of predicting disease outbreaks include different features that are associated with advanced public health initiatives and healthcare analytics. The first is the use of AI for disease maintenance and outbreaks. The integration of multiple variables in disease prediction models is crucial for enhancing their predictive power and accuracy, as it allows for a more comprehensive understanding of the complex relationships between variables and health outcomes. With the increasing emphasis on healthcare analytics, organizations are shifting from merely analyzing past events to employing descriptive, diagnostic, and predictive tools to forecast future occurrences and prescribe preventive measures. This advancement facilitates proactive public health interventions by enabling the prediction of various diseases and automating the detection of both communicable and noncommunicable diseases on a national scale. The growing utilization of AI in this context underscores its potential to influence preventive health behaviors and inform strategic responses. However, it is essential to balance these advancements with the potential risks, such as the possibility of AI-induced panic within the healthcare system due to predictions of high consultation rates, which could strain resources. Future research should focus on refining AI applications for disease management and ensuring that predictive models support effective and sustainable public health strategies.

Equ 3: Recurrent Neural Networks (RNNs)

$$\nabla_{\boldsymbol{c}} L = \sum_t \left(\frac{\partial \boldsymbol{o}^{(t)}}{\partial \boldsymbol{c}}\right)^{\top} \nabla_{\boldsymbol{o}^{(t)}} L = \sum_t \nabla_{\boldsymbol{o}^{(t)}} L$$

$$\nabla_{\boldsymbol{b}} L = \sum_t \left(\frac{\partial \boldsymbol{h}^{(t)}}{\partial \boldsymbol{b}^{(t)}}\right)^{\top} \nabla_{\boldsymbol{h}^{(t)}} L = \sum_t \operatorname{diag}\left(1 - \left(\boldsymbol{h}^{(t)}\right)^2\right) \nabla_{\boldsymbol{h}^{(t)}} L$$

$$\nabla_{\boldsymbol{V}} L = \sum_t \sum_i \left(\frac{\partial L}{\partial o_i^{(t)}}\right) \nabla_{\boldsymbol{V}^{(t)}} o_i^{(t)} = \sum_t \left(\nabla_{\boldsymbol{o}^{(t)}} L\right) \boldsymbol{h}^{(t)\top}$$

$$\nabla_{\boldsymbol{W}} L = \sum_t \sum_i \left(\frac{\partial L}{\partial h_i^{(t)}}\right) \nabla_{\boldsymbol{W}^{(t)}} h_i^{(t)} = \sum_t \operatorname{diag}\left(1 - \left(\boldsymbol{h}^{(t)}\right)^2\right) \left(\nabla_{\boldsymbol{h}^{(t)}} L\right) \boldsymbol{h}^{(t-1)\top}$$

$$\nabla_{\boldsymbol{U}} L = \sum_t \sum_i \left(\frac{\partial L}{\partial h_i^{(t)}}\right) \nabla_{\boldsymbol{U}^{(t)}} h_i^{(t)} = \sum_t \operatorname{diag}\left(1 - \left(\boldsymbol{h}^{(t)}\right)^2\right) \left(\nabla_{\boldsymbol{h}^{(t)}} L\right) \boldsymbol{x}^{(t)\top}$$

5.1. Potential Impact on Public Health Policies and Interventions

One realm with many transformative applications of AI and big data is in public health emergency response in areas that could include notifiable diseases in humans, animals, or crops as well as natural and man-made disasters. In areas with low coverage given the public expenditure involved, the most compelling cases are where AI could provide a lead of at least a month over current early warning systems linked to big data available in governments and global sources about how an acute or subacute event is progressing or its likely peak such that a timely and appropriate quality and scale of response will save more lives and prevent more harm. Beyond these areas, more features and analyses may be required, for example, aspects of One Health linking possible animal reservoirs with humans, modeling of the damage that a known disease is already causing to human health or animal health and production, or the possibility of collateral or knock-on effects. The use of machine learning and big data in predicting early disease outbreaks is perceived to impact public health policies and interventions. The application of AI and big data in public health, more fundamentally, is a set of tools that could foster a new paradigm in evidence-based public health practice, which healthcare professionals have been struggling to introduce. Given the multiplicity of changes that public health would have to introduce to address the challenges of AI and big data, this form of information should also facilitate the re-engineering of public health systems and organizations to cope with new decision-making regimes and resource allocations that follow. The most compelling case is one in which precise geospatial data linked to an infectious disease can be used to guide policy actions.AI and big data have the potential to revolutionize public health emergency response by offering advanced predictive capabilities for notifiable diseases and disaster events. In regions with limited public health infrastructure and funding, AI could provide critical early warnings—up to a month in advance—by analyzing vast datasets from governmental and global sources. This lead time allows for timely and scaled responses, potentially saving lives and mitigating harm. Beyond immediate disease detection, AI can integrate complex factors such as One Health considerations, linking animal reservoirs to human outbreaks, and modeling the broader impacts on health and production systems. The integration of AI and big data into public health could usher in a new era of evidence-based practice, offering tools to improve policy and intervention strategies. This shift demands significant changes in public health systems, including the re-engineering of decision-making processes and resource allocation to adapt to these advanced technologies. The use of precise geospatial data in conjunction with AI can particularly enhance policy actions, providing a more nuanced and proactive approach to managing infectious disease threats.

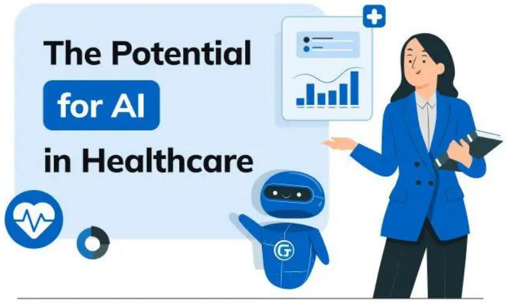

Fig 5: The Potential for AI in Healthcare

6. Conclusion

The challenges and opportunities in predicting disease outbreaks mostly revolve around two interconnected technologies: AI and big data. Nevertheless, disease outbreaks are only predicted based on historical and contemporary data within these frameworks, while all the evidence of historical and contemporary data is conjointly referred to as big data. Machine learning models begin to function when big data is provided and go through multi-staged and multi-method procedures to reply to the questions needed. Other studies indicate the prediction of unknown diseases by the functioning of a computational model that incorporates AI into predictive systems. The use of AI in big data-based computational models enhances the reliability and, to some extent, the timeliness of a prediction model to manage the impact of an outbreak situation, such as the discovery of a new pathogen. The increasing incidence of unusual and local diseases raises new issues for medical informatics, including the handling of imminent pandemic threats, the prediction of unknown local outbreaks, and the geographical spread of emerging diseases. The increasing incidence of unusual outbreaks has attracted global attention to the development of systems and frameworks that can predict the start of epidemics more accurately and in a shorter period of time. In this regard, it appears that big data and artificial intelligence have the potential to bring healthcare analytics to a new level by introducing innovative prediction methods and algorithms. Although more research would be needed to thoroughly understand the subtleties of introducing AI-generated predictions driven by big data into hospitals and clinics, early evidence suggests that such a frontier in AI and health informatics brings an era that is rich in possibilities. The use of AI-driven predictions could offer a new direction to epidemic management in order to enhance healthy lifestyles and improve individual and herd immunity. Big data can provide the necessary information for AI, as long as considerations and privacy issues are addressed.

6.1. Future Trends

By combining key terms with major topics, I have been able to carry out a systematic search of the three major databases to identify emerging issues, research gaps, and other indicators of future trends. There are four emerging issues that have attracted the attention of researchers in the domain of big data and AI. We expect that the more predictive analytics expands to this area, the more types of data sources will be used. Future researchers are encouraged to investigate text-based search engines for predicting disease outbreaks. AI can refine the identification and discovery of an epidemic or outbreak of either known or an emerging disease considering the gauge of burden, for instance, cases, deaths, and sicknesses. Enhanced sensitivity in picking up outbreaks will be applicable to different maladies characterized by AI prospects such as digital health and big data in healthcare.

As far as healthcare analytics are concerned, there is a growing interest from healthcare companies to shift away from standard descriptive analytics to predictive and prescriptive analytics due to the capabilities that AI possesses. In the future, AI and big data have great potential to change how disease outbreaks are predicted in terms of reducing falling ill and dying from diseases globally. As far as AI is concerned, it has the ability to detect and identify zoonosis events early across the globe by searching for patterns that detect diseases and healthcare changes through requested searches and digital information. As big data has been combined with healthcare big data and digitized data, it can help predict disease outbreaks in the future in the healthcare sector.

7. References

1. Kommisetty, P. D. N. K. (2022). Leading the Future: Big Data Solutions, Cloud Migration, and AI-Driven Decision-Making in Modern Enterprises. Educational Administration: Theory and Practice, 28(03), 352-364.
2. Yadav, P. S. Optimizing Data Stream Processing Pipelines: Using In-Memory DB

and Change Data Capture for Low-Latency Enrichment.
3. Mahida, A. (2022). Comprehensive Review on Optimizing Resource Allocation in Cloud Computing for Cost Efficiency. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-249. DOI: doi. org/10.47363/JAICC/2022 (1), 232, 2-4.
4. Tilala, M., Pamulaparti Venkata, S., Chawda, A. D., & Benke, A. P. Explore the Technologies and Architectures Enabling Real-Time Data Processing within Healthcare Data Lakes, and How They Facilitate Immediate Clinical Decision-Making and Patient Care Interventions. European Chemical Bulletin, 11, 4537-4542.
5. Chintale, P., Deshmukh, H., & Desaboyina, G. Ensuring regulatory compliance for remote financial operations in the COVID-19 ERA.
6. Avacharmal, R. (2022). ADVANCES IN UNSUPERVISED LEARNING TECHNIQUES FOR ANOMALY DETECTION AND FRAUD IDENTIFICATION IN FINANCIAL TRANSACTIONS. NeuroQuantology, 20(5), 5570.
7. Dilip Kumar Vaka. (2019). Cloud-Driven Excellence: A Comprehensive Evaluation of SAP S/4HANA ERP. Journal of Scientific and Engineering Research. https://doi.org/10.5281/ZENODO.11219959
8. Mandala, V., & Kommisetty, P. D. N. K. (2022). Advancing Predictive Failure Analytics in Automotive Safety: AI-Driven Approaches for School Buses and Commercial Trucks.
9. Yadav, P. S. (2022). Enhancing Real-Time Data Communication and Security in Connected Vehicles Using MQTT Protocol. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-E122. DOI: doi. org/10.47363/JAICC/2022 (1) E122 J Arti Inte & Cloud Comp, 1(3), 2-6.
10. Mahida, A. Predictive Incident Management Using Machine Learning.
11. Vaka, D. K. (2020). Navigating Uncertainty: The Power of 'Just in Time SAP for Supply Chain Dynamics. Journal of Technological Innovations, 1(2).
12. Pamulaparti Venkata, S. (2022). Unlocking the Adherence Imperative: A Unified Data Engineering Framework Leveraging Patient-Centric Ontologies for Personalized Healthcare Delivery and Enhanced Provider-Patient Loyalty. Distributed Learning and Broad Applications in Scientific Research, 8, 46-73.
13. Chintale, P., Korada, L., WA, L., Mahida, A., Ranjan, P., & Desaboyina, G. RISK MANAGEMENT STRATEGIES FOR CLOUD-NATIVE FINTECH APPLICATIONS DURING THE PANDEMIC.
14. Avacharmal, R., & Pamulaparthyvenkata, S. (2022). Enhancing Algorithmic Efficacy: A Comprehensive Exploration of Machine Learning Model Lifecycle Management from Inception to Operationalization. Distributed Learning and Broad Applications in Scientific Research, 8, 29-45.
15. Mandala, V., & Mandala, M. S. (2022). ANATOMY OF BIG DATA LAKE HOUSES. NeuroQuantology, 20(9), 6413.
16. Yadav, P. S. (2020). Minimize Downtime: Container Failover with Distributed Locks in Multi-Region Cloud Deployments for Low-Latency Applications. International Journal of Science and Research (IJSR), 9(10), 1800-1803.
17. Vaka, D. K. " Integrated Excellence: PM-EWM Integration Solution for S/4HANA 2020/2021.
18. Mahida, A. (2022). A Comprehensive Review on Ethical Considerations in Cloud Computing-Privacy Data Sovereignty, and Compliance. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-248. DOI: doi. org/10.47363/JAICC/2022 (1), 231, 2-4.
19. Mulukuntla, S., & Pamulaparthyvenkata, S. (2022). Realizing the Potential of AI in Improving Health Outcomes: Strategies for Effective Implementation. ESP Journal of Engineering and Technology Advancements, 2(3), 32-40.
20. Chintale, P., & Desaboyina, G. (2018). FLUX: AUTOMATING CLUSTER STATE MANAGEMENT AND UPDATES THROUGH GITOPS IN KUBERNETES. International Journal of Innovation Studies, 2(2).
21. Vaka, D. K. "Artificial intelligence enabled Demand Sensing: Enhancing Supply Chain Responsiveness.
22. Avacharmal, R. (2021). Leveraging Supervised Machine Learning Algorithms for Enhanced Anomaly Detection in Anti-Money Laundering (AML) Transaction Monitoring Systems: A Comparative Analysis of Performance and Explainability. African Journal of Artificial Intelligence and Sustainable Development, 1(2), 68-85.
23. Mandala, V., Premkumar, C. D., Nivitha, K., & Kumar, R. S. (2022). Machine Learning Techniques and Big Data Tools in Design and

Manufacturing. In Big Data Analytics in Smart Manufacturing (pp. 149-169). Chapman and Hall/CRC.

24. Yadav, P. S. (2021). Big Data Analytics and Machine Learning: Transforming Fixed Income Investment Strategies. North American Journal of Engineering Research, 2(2).
25. Mahida, A. A Review on Continuous Integration and Continuous Deployment (CI/CD) for Machine Learning.
26. Pamulaparti Venkata, S., & Avacharmal, R. (2021). Leveraging Machine Learning for Proactive Financial Risk Mitigation and Revenue Stream Optimization in the Transition Towards Value-Based Care Delivery Models. African Journal of Artificial Intelligence and Sustainable Development, 1(2), 86-126.
27. Chintale, P., & Desaboyina, G. (2018). FLUX: AUTOMATING CLUSTER STATE MANAGEMENT AND UPDATES THROUGH GITOPS IN KUBERNETES. International Journal of Innovation Studies, 2(2).
28. Mandala, V. (2022). Revolutionizing Asynchronous Shipments: Integrating AI Predictive Analytics in Automotive Supply Chains. Journal ID, 9339, 1263.
29. Yadav, P. S. (2021). Improving DevOps Efficiency with Jenkins Shared Libraries and Templates. European Journal of Advances in Engineering and Technology, 8(11), 116-120.
30. Mahida, A. A Comprehensive Review on Generative Models for Anomaly Detection in Financial Data.
31. MULUKUNTLA, S., & VENKATA, S. P. (2020). AI-Driven Personalized Medicine: Assessing the Impact of Federal Policies on Advancing Patient-Centric Care. EPH-International Journal of Medical and Health Science, 6(2), 20-26.
32. Perumal, A. P., Deshmukh, H., Chintale, P., Desaboyina, G., & Najana, M. Implementing zero trust architecture in financial services cloud environments in Microsoft azure security framework.
33. Mandala, V., & Surabhi, S. N. R. D. (2021). Leveraging AI and ML for Enhanced Efficiency and Innovation in Manufacturing: A Comparative Analysis.