Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Exploring Future Disease 'X' Outbreak Possibilities, Risk and Challenging Factors: A Comprehensive Analysis with Machine Learning Algorithms

Dr Prakash Kuppuswamy[1], Dr Saroj K Gupta[2], Dr Indhu Sharma[3], Dr. Saeed Q. Y Al Khalidi Al-Maliki[4], Ahmed Ali ShaikMeeran[5], Ahmed Hamed[6]

[1]Computer Science and Engineering Department, SRM University, Sonepat. E-Mail: prakashcnet@gmail.com
[2]Computer Science and Engineering Department, SRM University, Sonepat. E-Mail: saroj.k@srmuniversity.ac.in
[3]Computer Science Engineering Department, SRM University Delhi NCR
[4]Professor and Vice Dean, Department of Business Informatics, King Khalid University (KKU), Saudi Arabia. E-Mail: salkhalidi@kku.edu.sa
[5]College of Computer Science and Information Technology, Jazan University, Jazan, Saudi Arabia
[6]College of Applied Science, Jazan University, Jazan, Saudi Arabia

Background Pandemic refers to the widespread outbreak of a disease that transitions from an epidemic to a global scale, affecting multiple countries. Some of the most devastating diseases in human history, such as Cholera, Plague, smallpox, and influenza, have had a significant impact on populations worldwide. Smallpox alone has caused an estimated 300-500 million deaths over thousands of years.

Methods Using of artificial intelligence tools plays a crucial role in assessing risks and identifying areas that require immediate attention during pandemics. The objective of the research to explore potential future pandemic scenarios and their vulnerabilities. Healthcare professionals are urged to implement preventive measures to protect their communities and reduce human casualties. Artificial intelligence, machine learning, and deep learning algorithms can be valuable assets in forecasting future pandemic trends and assessing vulnerabilities. By adopting machine learning-based methodologies outlined in this research, healthcare operations can swiftly and effectively identify potential pandemic scenarios.

Results Linear regression is a statistical method used to predict the value of one variable based on another. It has an MSE of 567116.1297, indicating prediction error, and an R-squared value of 0.7779. The Random Forest algorithm, also with an MSE of 567116.1297 and an R-squared of 0.7779, shows similar prediction accuracy. Support Vector Machine (SVM) shares the same MSE and R-squared values as Random Forest, indicating comparable performance. Random Forest is noted to be better for predicting better result based on dataset trends.

Conclusions The emergence of Pandemic X, highlighted by advanced Machine Learning algorithms, raises significant concerns for humanity. Key factors affecting its timeline and impact include population density, climate, economic activity, healthcare resources, preventive measures, historical data, global travel, and the effectiveness of pandemic models. A comprehensive approach integrating these factors is vital for effective preparedness and response strategies in the future.

Keywords: Pandemic X, Machine learning, Artificial Intelligence, Pandemic, Covid-19, Influenza etc.,

1. INTRODUCTION

The large influenza pandemic of 1918 is kind of the anchor in historical perspective, but we have had some fairly milder influenza pandemics after that. In some reports, HIV is considered as a pandemic that has become endemic [1, 2]. Originally outbreaks of infectious diseases, that have the potential for a pandemic break out and spread quite frequently in history [3]. Humanity has also experienced both major pandemics and epidemics, the most famous ones being plague, cholera or influenza) as well as atypical coronaviruses (severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV)[4, 5]. We are in the midst of a new pandemic: coronavirus disease 2019 (COVID-19). Over the past decade, 13 major zoonotic diseases originated from birds and mammals that have been transmitted to humans due to increased interaction with animals through domestication; breeding pangolins emerged as victims of a long-standing illegal global trade in which wildlife is hunted or obtained unsustainably [6-8].

Isolation, quarantine and border control are all public health measures taken over centuries to manage the transmission of infectious pathogens so that society can continue to function. These containment strategies have even been applied today to combat the COVID-19 pandemic [9-11]. Health sector's should be considered that during a pandemic, rapid diagnostic and testing technologies, contact tracing strategies. repurposing of drugs, cues to the severity of disease along with new vaccine development platforms and production facility are much needed for an effective response [12, 13]. Epidemics of infectious diseases occur when a cascade of risk factors come together. A 1998 El Niño weather event resulted in four East African countries, Kenya, Somalia, Sudan, and Tanzania [14-16].

Flooding most of these countries forced cattle and humans alike to share the remaining dry parts of the land. Because the latter was unable to secure sufficient vaccines for all its animals, the cattle were not vaccinated against the Rift Valley Fever, one of the most common animal ruminant infections in the region [17-19]. The flooding resulted in the creation of more breeding grounds for mosquitoes, which increased the mosquito population at an unprecedented rate. Vaccines are not an option for humans, because one of the means by which the Rift Valley Fever is transmitted is a mosquito: from animal to human and from human to human. This led to the emergence of the virus within human populations, which was subsequently transmitted from person to person [20-22]. The combination of these risk factors resulted in a significant outbreak of Rift Valley Fever among the human population in the region [23].

The medical sector is actively exploring innovative technologies and techniques to better control and manage the spread of the pandemic. Artificial intelligence, machine learning, and deep learning methods are being investigated as potential tools to enhance and sustain healthcare systems and organizations in effectively monitoring pandemic situations [24-27]. These advanced technologies have the potential to significantly impact the fight against the pandemic by reducing the death rate, identifying high-risk areas, tracking the spread of the virus, and controlling infections in real-time. Moreover, they can also aid in predicting fatality risks by analyzing historical data effectively [28, 29]. The emergence of new diseases, exemplified by the hypothetical Disease X, underscores the importance of being prepared for future outbreaks. The inevitability of future disease outbreaks, potentially originating from animals and spreading to humans, necessitates a proactive approach in addressing and mitigating the impact of such health crises on a global scale [30, 31].

Recent history has shown an increase in outbreaks caused by pathogens capable of causing severe diseases and fatalities in humans, further emphasizing the need for continuous advancements in technology and proactive measures to combat such health emergencies effectively [32]. In the 21st century, various outbreaks of infectious diseases such as SARS-CoV-1, MERS, Zika, and other new and re-emerging viral diseases have highlighted the importance of global preparedness for large-scale outbreaks or pandemics [33, 34]. The advancements in technologies for vaccines, diagnostic tests, and therapeutics have played a crucial role in enhancing the world's readiness for future health crises. These technologies have been developed over years of research investment, with the COVID-19 pandemic significantly accelerating their progress [35, 36].

In the event of another influenza or coronavirus pandemic, these advanced technologies hold the promise of rapidly developing effective vaccines, diagnostics, and treatments. However, the primary obstacles to preparedness for the next pandemic are likely to be political in nature [37, 38]. Balancing national sovereignty with the need for global solidarity is essential to ensure a coordinated and effective response to future health emergencies. One proposed solution to address these challenges is an Artificial Intelligence and Machine Learning-based healthcare framework designed to analyze, predict, and confront the potential challenges posed by future pandemic diseases [39, 40]. This framework leverages data analytics to conduct four types of analysis: descriptive, diagnostic, predictive, and prescriptive. By harnessing the power of AI and machine learning, this innovative approach aims to enhance preparedness and response strategies for future pandemics on a global scale [41, 42].

The table-1 lists various historical epidemics and pandemics, showing their names, the time period they occurred, the type of disease, and the estimated number of deaths caused by each outbreak. This information helps us understand the impact of these diseases on human populations throughout history. This table highlights the significant impact of infectious diseases throughout history. It shows how different pathogens, whether viruses or bacteria, have caused massive loss of life. Understanding these epidemics helps us learn about disease transmission and the importance of public health measures to prevent future outbreaks.

Table 1. Historical Epidemics

Years	Type of Pandemics	Routes
165-180	Antonine Plague	Cattles
541-543	Justinian Plague	Fleas
735-737	Japanese smallpox	Camel
1347-1351	Black Death	Fleas
1520-till	New world smallpox	cattles
1817-1824	1st Cholera	Water contamination
1827-1835	2nd Cholera	Water contamination
1839-1856	3rd Cholera	Water contamination
1863-1875	4th Cholera	Water contamination
1881-1886	5th Cholera	Water contamination
1890	Yellow Fever	Mosquitoes
1885 –till	3rd plague	Fleas
1889-1893	Russian Flu	Avian
1899-1923	6th Cholera	Water contamination
1918-1919	Spanish flu	Avian
1957-1959	Asian flu	Avian
1961- till	7th Cholera	Water contamination
1968-1970	Hong Kong flu	Avian
1981-till	AIDS/HIV	Chimpanzees
2002-2003	SARS	Bats
2009-2010	Swine flu	Pigs
2014-2016	Ebola	Wild animals
2015- till	MERS	Bats, Camels
2019-till	COVID-19	Bats, pangolins

The table-2 presents information about various pandemics throughout history. It includes the years when these pandemics occurred, the type of pandemic, the number of people affected (in millions), and the total population during that time. This column shows the time period during which each pandemic took place. Some pandemics lasted for just a few years, while others spanned many years. Type of Pandemics column lists the names of the diseases that caused the pandemics. Each name represents a different illness that affected many people. By studying these pandemics, we can learn lessons that may help us deal with current and future health crises.

Table 2. The Pandemic Data Table

Years	**Type of Pandemics**	**in Million**	**Total Population**
165-180	Antonine Plague	5	200
541-543	Justinian Plague	50	210
735-737	Japanese smallpox	1	226
1347-1351	Black Death	200	392
1520-till	New world smallpox	56	461
1817-1824	1st Cholera	0.2	989
1827-1835	2nd Cholera	0.2	990
1839-1856	3rd Cholera	0.2	1030
1863-1875	4th Cholera	0.1	1263
1881-1886	5th Cholera	0.1	1276
1890-1920	Yellow Fever	0.15	1654
1885 –till	3rd plague	0.13	1563
1889-1893	Russian Flu	1	1654
1899-1923	6th Cholera	0.8	1912
1918-1919	Spanish flu	50	1912
1957-1959	Asian flu	1.1	2557
1961- till	7th Cholera	1	3043
1968-1970	Hong Kong flu	1	3712
1981-till	AIDS/HIV	30	4451
2002-2003	SARS	0.001	6242
2009-2010	Swine flu	0.2	6788
2014-2016	Ebola	0.01	7169
2015- till	MERS	0.0001	7247
2019-till	COVID-19	7	8169

In the following figure-1 and figure 2, refers "overall death", the total number of people who have died within a specific period, usually measured over a year. This can include deaths from all causes, such as diseases, accidents, natural disasters, and other factors. We are talking about the total number of deaths expressed in millions. Overall death in millions is a vital statistic that provides insight into the health of populations. By understanding this concept, we can better appreciate the factors that contribute to mortality and the importance of addressing health issues to improve life expectancy and quality of life.

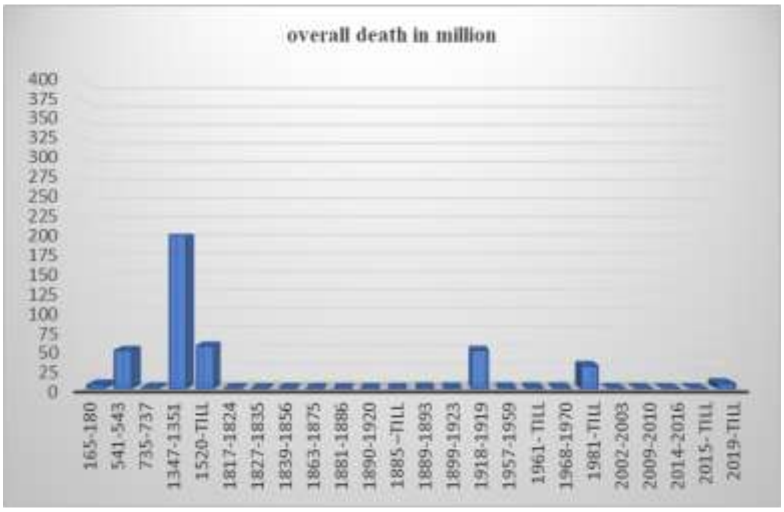

Figure 1. Pandemic Overall Death

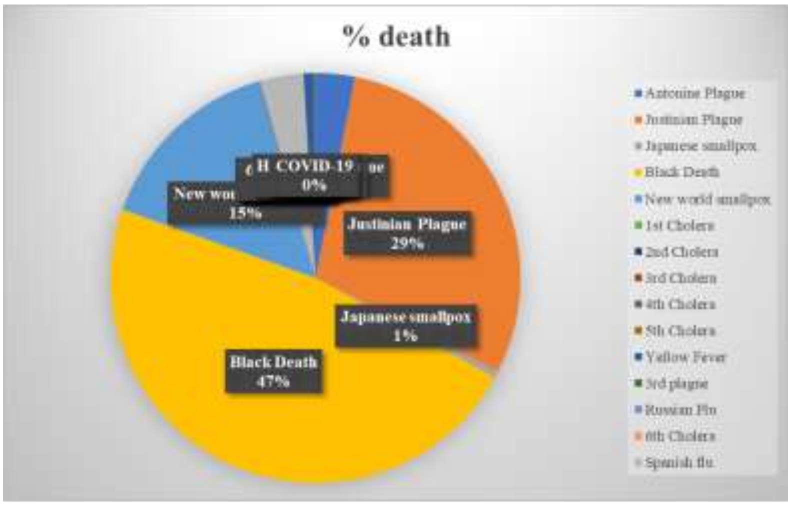

Figure 2. Percentage of death ratio

Zoonotic threats pose a significant risk to human populations, and virologists are constantly on the lookout for new potential threats [43]. While these threats may be unfamiliar at first, they often bear similarities to previously known viruses. Certain viral clades are more prone to making the jump from animals to humans, increasing the likelihood of continued spread within human populations [44, 45]. This pattern is evident in past outbreaks, such as the similarities between SARS-CoV-2 and its predecessor SARS-CoV, as well as the emergence of HIV-1 group M leading to multiple spillovers of similar primate viruses. Predicting and managing pandemics is crucial for public health, especially in times of uncertainty like the current COVID-19 pandemic [46, 47].

Data-driven predictions are essential but challenging due to the unpredictable nature of pandemics. The complexity of pandemics as "wicked problems" further complicates prediction efforts. Wicked problems are characterized by their novelty, complexity, and evolving nature, with incomplete and changing requirements. They involve intertwined social, economic, and political factors, as well as stakeholders with diverse views and dynamic reactions, creating a challenging and ever-changing landscape for pandemic response and prediction [48, 49]. The COVID-19 pandemic presents a unique challenge due to the presence of a novel virus without a known cure, accompanied by complex political, economic, and societal factors on a global scale. Various intricate and uncertain elements make it challenging to understand and model. Despite the availability of diverse predictive models, their validity cannot be definitively confirmed or refuted as there is no established criterion for assessing their precision. Different from "tame problems", which have clear solutions based on set criteria, "wicked problems" lack stable agreements and can only be resolved through the application of suitable methodologies [50, 51].

Researchers, healthcare professionals, epidemiologists, and policymakers are confronted with a significant challenge due to the current pandemic. As part of the global response. The healthcare sector stands to benefit significantly from machine learning algorithms, given the vast amount of data generated per patient. Consequently, the healthcare industry has witnessed several successful instances of applying machine learning techniques. Machine learning offers a broad spectrum of capabilities for addressing various healthcare tasks and diverse research endeavors. In healthcare, machine learning plays a pivotal role in enabling a range of services [52-55].

2. LITERATURE REVIEW

Alzahrani SI, Aljamaan IA, Al-Fakih EA (2020) The notable escalation in the quotidian tally of newly reported coronavirus cases on a global scale is a matter of significant concern, prompting numerous researchers to currently employ a diverse array of mathematical and machine learning-driven predictive models to ascertain the prospective trajectory of this pandemic. In the present investigation, we implemented the Autoregressive Integrated Moving Average (ARIMA) model to estimate the projected daily incidence of COVID-19 cases in Saudi Arabia for the forthcoming month. It assessed four distinct forecasting methodologies: the Autoregressive Model, the Moving Average, a hybrid model (ARMA), and an integrated ARMA (ARIMA), with the objective of determining the most appropriate model, ultimately concluding that the ARIMA model exhibited superior efficacy. Subsequently, it fine-tuned the parameters of the ARIMA model to enhance its predictive accuracy, ensuring that it accounted for seasonal variations and trends inherent in the data [56].

N. Kumar and S. Susan (2020) The worldwide COVID-19 pandemic has profoundly impacted millions, with numerous individuals succumbing to the virus. It is imperative to anticipate future rates of infection and the velocity of virus transmission to enhance the preparedness of healthcare systems and mitigate mortality. Forecasting the advancement of COVID-19 poses a multifaceted and analytical challenge for researchers. This endeavor necessitates the model of the COVID-19 epidemic's trajectory and the application of ARIMA and Prophet time series forecasting methodologies for predictions. The efficacy of these models is evaluated using metrics such as mean absolute error, root mean square error, root relative squared error, and mean absolute percentage error. This analytical framework contributes to a deeper comprehension of disease trends and offers valuable insights into the epidemiological phases of various nations. Investigation reveals that ARIMA models exhibit greater reliability in forecasting the prevalence of COVID-19. The resultant forecasting findings could assist governments in formulating strategies to mitigate the proliferation of the virus [57].

Dr. Saeed Q Al-Khalidi Al-Maliki, Dr. Prakash Kuppuswamy, Dr. Rajan John, Dr. Nithya Rekha Sivakumar (2022) This scholarly article express a systematic methodology for comprehending the mutations associated with COVID-19 through the utilization of sophisticated data analytics, accentuating the significance of machine learning in enhancing healthcare results amidst the pandemic. By employing extensive datasets, the objectives were to discern patterns and forecast prospective variants, ultimately informing public health strategies. The researchers employed artificial intelligence and machine learning techniques, notably the J48 and Linear Regression algorithms, to tackle the challenges posed by COVID-19 mutations. Their findings suggested that these algorithms are capable of categorizing mutations and anticipating their implications on transmissibility and vaccine effectiveness, thereby augmenting readiness for future outbreaks. This investigation emphasizes the importance of artificial intelligence and machine learning within the healthcare sector, particularly during pandemics, positing that these technologies could facilitate more rapid and precise diagnoses in forthcoming public health emergencies [58].

Albzeirat M, Zulkepli N, Qaralleh (2022) The emergence of the pandemic has engendered widespread apprehension on a global scale across various domains, encompassing public health, economic stability, international relations, political coherence, and social cohesion. Its onset was abrupt, impacting the world swiftly and unexpectedly. Presently, there remains an insufficiency of comprehensive data regarding the virus, including its origins, symptomatic manifestations, modes of transmission, diagnostic approaches, and therapeutic interventions. As a result, over one million individuals have succumbed to the virus, precipitating considerable economic ramifications. This manuscript will investigate the potential of artificial intelligence in forecasting and monitoring viral outbreaks, as well as in mitigating prospective future threats through the application of artificial intelligence, algorithmic methodologies, and cognitive fission theory [59].

Maher MC, Bartha I, Weaver S, Di Iulio J, Ferri E, Soriaga L, Lempp FA, Hie BL, Bryson B, Berger B, Robertson DL (2022) The progression of SARS-CoV-2 represents a considerable threat to both vaccine-mediated and naturally conferred immunity, as well as the efficacy of therapeutic monoclonal antibodies. In order to bolster public health preparedness, our objective was to ascertain which pre-existing amino acid mutations in SARS-CoV-2 could conceivably give rise to forthcoming variants of concern. The results revealed that the transmissibility facilitated by ACE2 and the virus's capacity to circumvent population-level immunity have varied temporally, acting as pivotal determinants in its evolutionary trajectory. The outcomes of the study demonstrated a pronounced level of precision in retrospectively identifying these trends. The behavior of the model conformed to a credible causal framework, wherein epidemiological variables encapsulate the influence of a diverse array of evolving factors that affect viral fitness. The findings exemplify the practicality of predicting the mutations that may arise in new SARS-CoV-2 variants of concern. This approach can be adapted to any rapidly evolving pathogens that possess extensive genomic surveillance data, such as influenza and potential future pandemic viruses [60].

Charlotte, Houldcroft (2023) The authors highlight the vital role of both ancient and modern DNA in studying the historical evolution of infectious diseases. This genetic evidence is essential for understanding how these diseases have influenced human evolution and societal development over time. The paper argues that analyzing historical pandemics can yield valuable insights and strategies for managing future outbreaks. By examining ancient diseases, scholars can derive lessons applicable to current public health efforts, enhancing our understanding of disease patterns and improving preparedness against emerging health threats. The manuscript calls for a thorough re-evaluation of historical perspectives on infectious diseases, emphasizing the need to integrate genetic data to fully comprehend the complex relationship between humans and pathogens throughout history. It highlights the importance of interdisciplinary collaboration among historians, geneticists, and epidemiologists to create a more comprehensive framework for studying these interactions [61].

3. RESEARCH PROBLEM

The rapid mutations observed in pathogens pose a challenge to the comprehension of their transmissibility and virulence. The process of diseases transferring from wildlife and livestock to humans, known as zoonotic spillover, remains a complex and unpredictable phenomenon. The influence of globalization, urbanization, and travel patterns on disease spread is substantial, rendering human behavior difficult to anticipate. The lack of sufficient public health surveillance, particularly in low- and middle-income nations, impedes the timely detection of outbreaks. The escalating human-animal interactions resulting from climate change and habitat destruction add a layer of complexity to public health forecasting. The presence of incomplete or inconsistent data undermines the precision of predictive models utilized for outbreak analysis. Disparities in political stability and healthcare systems have a significant impact on the responses to infectious diseases. The potential of AI and machine learning to assist in predictions is contingent upon the quality of the data employed. International collaboration is imperative; however, discrepancies in resources and policies can impede the efficacy of responses. Variations in the adherence of the public to health guidelines can profoundly alter the trajectories of outbreaks. In summary, overcoming these multifaceted challenges necessitates interdisciplinary methodologies, enhanced surveillance mechanisms, and strengthened global partnerships to enhance the capacity for predicting and mitigating forthcoming pandemics. Predicting future pandemics is challenging due to various complex factors, including:

- ❖ Evolution of Pathogens
- ❖ Transmission of Zoonotic Diseases
- ❖ Impact of Human Activities
- ❖ Deficiencies in Surveillance Systems
- ❖ Effects of Environmental Transformation
- ❖ Existence of Data Deficiencies
- ❖ Influence of Political and Social Elements
- ❖ Advancements in Technology
- ❖ Necessity of Global Cooperation
- ❖ Role of Public Awareness

4. PROPOSED METHOD

The research methods are divided into three comprehensive sections. The first section describes the objectives of the study, detailing the specific goals it aims to achieve. Next, the methodology is meticulously designed, incorporating various approaches and techniques for collecting and analyzing data. In addition, the research settings are described, providing context for the research environment and conditions. In order to conduct an effective and rigorous investigation, these elements must be considered together.

4.1Objectives

The World Health Organization (WHO) has expressed significant concern regarding the potential emergence of a future pandemic, often referred to as Pandemic X. Considering this, advanced strategies are essential to predicting and mitigating outbreaks of this type. Machine Learning (ML) and Artificial Intelligence (AI) algorithms will be pivotal tools to identify efficient drugs against infectious diseases in the future. In addition to epidemiological data, social media trends, traveling patterns, and environmental factors, these cutting-edge technologies can process and analyze huge datasets. As a result of this predictive capacity, health authorities can implement timely interventions, reducing transmission rates and safeguarding the public's health.

4.2Design

Preprocessing is the first step in preparing data for analysis or machine learning. It involves cleaning and organizing the data to make it suitable for the next steps. The portion of the data used to teach the model. It's like studying for a test; you need to practice with the material before the actual exam. After training the model on the training data, the testing data to see if it can make accurate predictions. Preprocessing, data splitting, and testing data are essential steps in data analysis and machine learning. They ensure that the data is clean, organized, and ready for building models that can make accurate predictions.

Simple Linear Regression:

$$y=\beta 0+\beta 1x+\epsilon \text{--} \quad (1)$$

Where:

- y is the dependent variable (the variable you are trying to predict).
- x is the independent variable (the variable used to make predictions).
- β0 is the y-intercept (the value of y when x=0).

- β1 is the slope of the line (how much y changes for a one-unit change in x).
- ϵ is the error term (the difference between the observed and predicted values of y).

Multiple Linear Regression:

$y=\beta 0+\beta 1x1+\beta 2x2+\cdots+\beta nxn+\epsilon$--- (2)

Where $x_1,x_2,\ldots,x_n$ are the independent variables and

$\beta 1,\beta 2,\ldots,\beta n$ are the corresponding coefficients.

Random Forest

Y =majority vote$(T_1(x),T_2(x),\ldots,T_M(x))$--- (3)

Where:

- $T_1(x),T_2(x),\ldots,T_M(x)$ are the predictions of the M decision trees in the forest.

Support Vector Machine

SVM optimization $\min_{w,b} 1/2\ ||w||^2$--- (4)

Where

Weight vector www, which maximizes the margin between the two classes.

SVM Kernel

SVM kernel maps the input data into a higher-dimensional space where it is linearly separable.

$f(x) = \mathrm{sign}(\sum_{i=1}^{n} \propto_i y_i\, k(x_i, x)+b$-- (5)

Where:

- $K(x_i,x)$ is the kernel function (e.g., linear, polynomial, radial basis function),
- α_i are the Lagrange multipliers obtained during the optimization process.

4.3Research Setting

Research setting are shown in In Figure 3, we present a comprehensive overview of the proposed architecture for pandemic prediction, which leverages a diverse array of machine learning algorithms. This architecture is meticulously designed to analyze and interpret complex datasets, enabling the identification of patterns and trends that may signal the onset of a pandemic. By integrating multiple machine learning techniques, the framework enhances its predictive accuracy and robustness, allowing for a more nuanced understanding of the factors contributing to pandemic outbreaks. Each algorithm is selected for its unique strengths, ensuring a holistic approach to data analysis that can adapt to the dynamic nature of infectious disease spread.

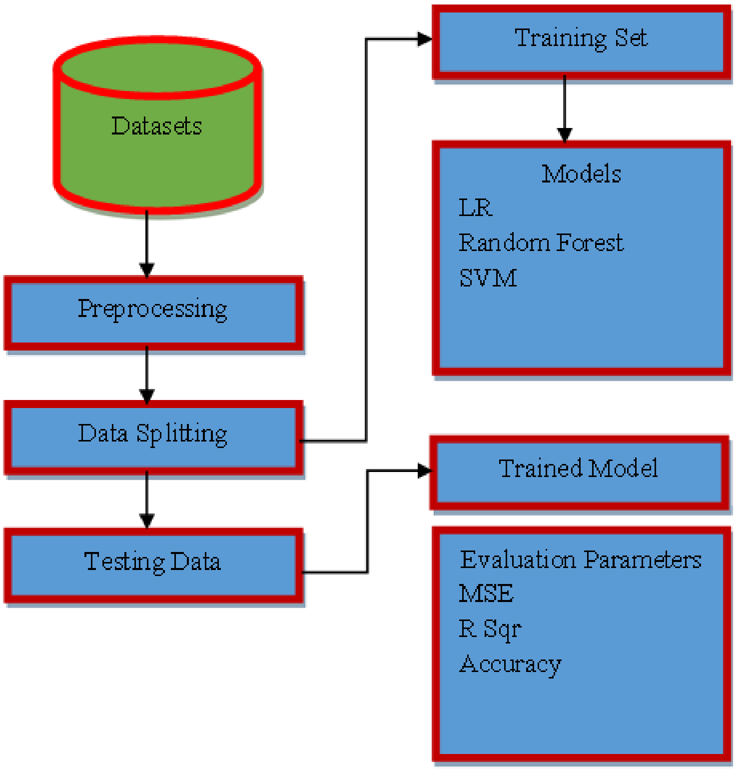

Figure 3. Proposed structure

5. RESULTS

Understanding pandemic deaths by country is crucial for learning how to manage and respond to pandemics. It provides valuable insights into public health and helps prepare for future health crises. As the situation evolves, continuous tracking and analysis of pandemic deaths will remain important for global health. This detailed breakdown helps us grasp the significance of tracking pandemic deaths by country and the various factors that influence these numbers. Machine learning (ML) is a type of artificial intelligence that allows computers to learn from data and make predictions or decisions without being explicitly programmed. In this research, three different ML algorithms are used: Linear Regression, Random Forest and Support Vector Machine (SVM). These are popular methods for making predictions based on data.

A) Linear regression

Linear regression is a statistical method used to understand the relationship between two or more variables. It helps us predict the value of one variable based on the value of another. Linear regression is used to predict how many deaths might occur based on certain factors.
Mean Squared Error (MSE): 567116.1296991897

The Mean Squared Error is a way to measure how well our model is performing. It calculates the average of the squares of the errors, which are the differences between the predicted values and the actual values. A lower MSE means our predictions are closer to the actual values. In this case, an MSE of 567116.1296991897 indicates that there is some error in our predictions, but we need to compare it with other models to understand if it is good or bad.

R-squared: 0.7779153973367876

R-squared is a statistic that tells us how well our model explains the variability of the data. It ranges from 0 to 1, where 0 means the model does not explain any variability, and 1 means it explains all the variability. An R-squared value of 0.7779 suggests that about 77.79% of the variability in the data can be explained by our model. This is generally considered a good fit, meaning our model is doing a decent job at predicting the outcomes.

Predicted Deaths: 30389.476377694515

This number represents the predicted number of deaths based on the model we created using linear regression. The prediction is approximately 30,389 deaths. This means that, according to our model, if we consider the factors we used, we expect around 30,389 deaths to occur. It’s important to remember that this is just a prediction. The actual number could be higher or lower depending on various factors that might not have been included in the model. If the MSE is high, it indicates that the model may need improvement. If the R-squared is low, it suggests that the model is not capturing the relationship well. The predicted deaths give a concrete number that can be used for planning and resource allocation, especially in fields like public health. The Mean Squared Error and R-squared values help us assess the accuracy and reliability of our predictions. The predicted number of deaths provides a specific estimate that can guide actions and decisions in relevant fields. Understanding these concepts is crucial for anyone looking to analyze data effectively.

B) Random Forest Algorithm:

Random Forest is a supervised learning algorithm, which means it learns from labeled data. It works by creating many decision trees and combining their results to make a final prediction.
The MSE for Random Forest is 567116.1296991897.
MSE measures how close the predicted values are to the actual values. A lower MSE indicates better performance. In this case, the value is quite high, suggesting there is some error in the predictions.
The R-squared value for Random Forest is 0.7779153973367876.
R-squared is a statistical measure that shows how well the model explains the variability of the data. It ranges from 0 to 1, where 1 means perfect prediction. A value of 0.77 indicates that the model explains about 77% of the variability in the data, which is considered good.

C) Support Vector Machine (SVM)

SVM is another supervised learning algorithm that works by finding the best boundary that separates different classes in the data. It is often used for classification tasks but can also be adapted for regression.
Mean Squared Error (MSE):
The MSE for SVM is the same as for Random Forest, 567116.1296991897. This indicates that both algorithms have similar levels of prediction error.
R-squared Value:
The R-squared value for SVM is also 0.7779153973367876, meaning it explains the same amount of variability in the data as the Random Forest model.

The note at the end states that the Random Forest algorithm works better for predicting deaths based on the given datasets. This conclusion is likely drawn from the fact that it provides a higher predicted death count, which may be more aligned with actual data trends. The choice of algorithm can depend on the specific characteristics of the data and the goals of the analysis. In this case, Random Forest seems to be the more reliable option for this particular prediction task. Understanding these algorithms and their performance metrics is crucial for making informed decisions in data analysis. The research highlights the importance of selecting the right algorithm based on the data and the desired outcomes, as different algorithms can yield different predictions even with the same data.

6. DISCUSSION

In Figure 4, "Death distribution" refers to how the number of deaths from COVID-19 is spread across different countries. This means looking at how many people died in each country due to pandemic. It helps us understand which countries were most affected by the virus. Knowing how many deaths occurred in each country can help governments and health organizations respond better to health crises. It can also inform the public about the severity of the situation in different areas. This information can guide decisions on travel, healthcare resources, and vaccination efforts.

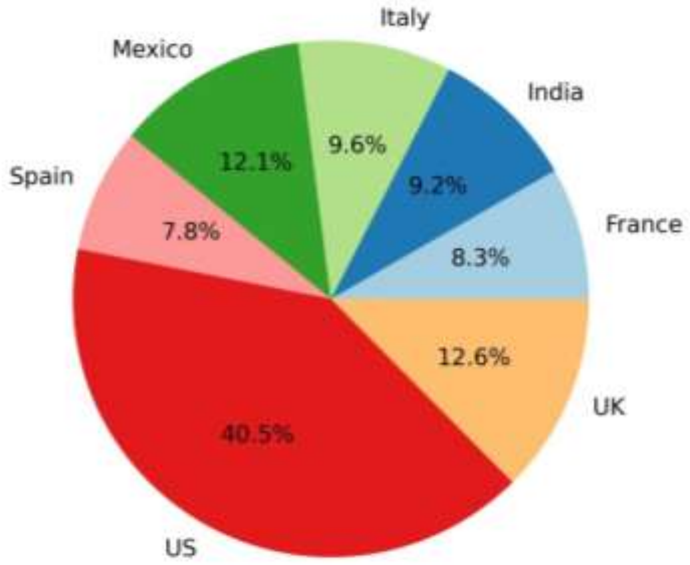

Figure 4. Pandemic death distribution

Generally, the several factors (figure 6) can affect the number of deaths in a country. Larger countries may have more deaths simply because they have more people. Second, Countries with better healthcare systems can treat patients more effectively, potentially reducing deaths. Third, Government responds to the outbreak can impact death rates. Finally, the age and health of the population matter. Older people or those with pre-existing health conditions are at higher risk of severe illness and death.

The distribution of pandemic deaths by country is a crucial aspect of understanding the pandemic. It helps identify trends, inform public health strategies, and ultimately save lives. By studying this data, we can learn valuable lessons for future health emergencies. Tracking deaths helps governments and health organizations respond effectively to the pandemic. It allows them to allocate resources, such as medical supplies and personnel, to areas that need them most. It also helps researchers study the virus and its impact on different populations. By comparing COVID-19 deaths across countries, we can see which countries have been more affected by the pandemic. This comparison can highlight the effectiveness of different strategies used to combat the virus.

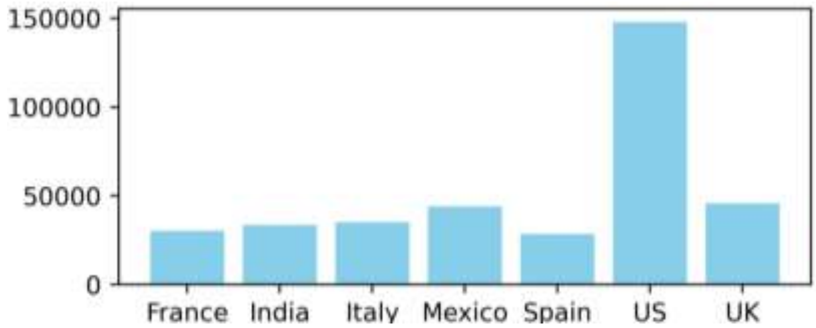

Figure 5. No. of Death

7. PANDEMIC PREDICTION FACTORS

Historical data reveals a remarkable decline in the death rate within the human population following the introduction of vaccination programs, with fatalities from preventable diseases plummeting to below 1%. This substantial reduction serves as a powerful testament to the efficacy of vaccines in not only curbing the incidence of severe illnesses but also in significantly lowering mortality rates associated with these diseases. The implications of this decline are profound, emphasizing the critical role that widespread immunization campaigns play in safeguarding public health and enhancing community resilience against infectious diseases. Moreover, the landscape of vaccine research is continually evolving, with dedicated efforts focused on the development of innovative vaccines aimed at combating emerging pathogens. This ongoing research is vital in addressing new health threats and ensuring that public health initiatives remain proactive and effective. By prioritizing the advancement of vaccination strategies, health authorities can better prepare for future outbreaks and protect vulnerable populations, thereby reinforcing the overarching goal of maintaining and improving global health standards.

Figure 6. Pandemic factors

In Figure 7, the death rate is a measure that tells us how many people die in a specific population over a certain period of time, usually expressed per 1,000 individuals. It helps us understand the health of a population and can indicate how well a society is doing in terms of healthcare, living conditions, and overall quality of life. A high death rate may indicate problems such as disease outbreaks, poor healthcare, or unsafe living conditions.

Governments use death rate data to create policies aimed at improving health services, addressing social issues, and planning for future healthcare needs. Several factors can affect the death rate in a population:

- Age Distribution
- Healthcare Access
- Lifestyle Choices
- Environmental Factors
- Medical Advances
- Socioeconomic Changes

Outbreaks of diseases, like the COVID-19 pandemic, can cause sudden spikes in death rates. Understanding the death rate is crucial for assessing the health of a population. It provides valuable insights into the effectiveness of healthcare systems, the impact of lifestyle choices, and the overall well-being of a community. By analyzing death rates, we can work towards improving health outcomes and ensuring a better quality of life for everyone.

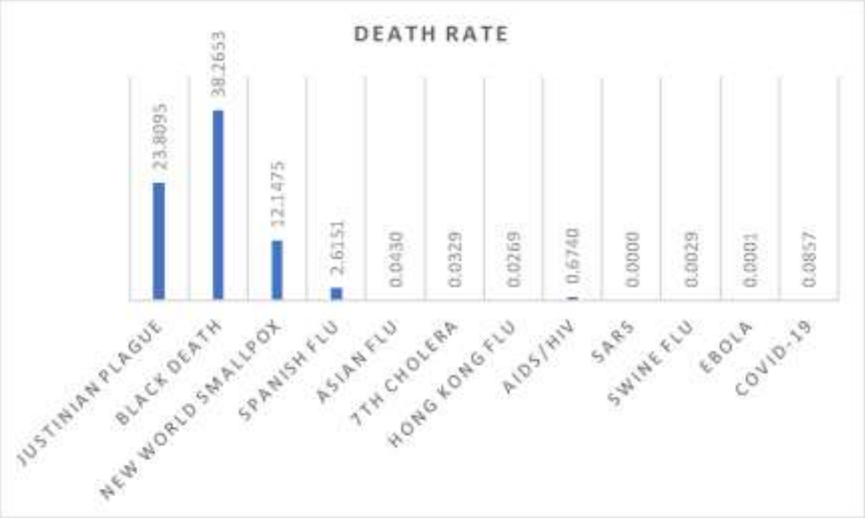

Figure 7. Death Rate before and after vaccination history

Throughout history, pandemics have been closely linked to substantial societal, economic, and political transformations. Events such as the bubonic plague, the Spanish flu, and more recently, the COVID-19 pandemic, have had profound repercussions on societies, healthcare systems, and economies. To prevent future pandemics, a comprehensive strategy encompassing the surveillance of animal health, promotion of sustainable environmental practices, enhancement of public health infrastructure, and fostering global collaboration to address emerging infectious diseases is imperative. The One Health approach, recognizing the interconnectedness of human, animal, and environmental health, is increasingly acknowledged as indispensable for tackling these intricate challenges.

8. CONCLUSION

The potential emergence of Pandemic X, as indicated by advanced Machine Learning algorithms, poses significant concerns for humanity. Key factors influencing the timeline and impact of such a pandemic include population density, climate, economic activity, healthcare resource availability, preventive measures, historical data, global travel and trade, and the effectiveness of existing pandemic models. High population density in urban areas can facilitate rapid disease spread, while climate affects pathogen viability and population susceptibility. Economic activity influences human mobility, impacting virus dissemination. Strong healthcare systems are crucial for managing outbreaks, while effective preventive measures can significantly alter pandemic trajectories. Historical insights from past pandemics can inform current strategies, and global interconnectedness necessitates coordinated international responses. Continuous evaluation and refinement of pandemic models are essential, along with ongoing monitoring and updating of risks associated with Pandemic X. A comprehensive approach integrating these factors is vital for effective preparedness and response strategies in the future. We have developed and implemented a range of cutting-edge scientific methodologies and innovative technologies designed to effectively manage and mitigate the impact of future pandemics. These advancements not only enhance our capacity to respond swiftly to emerging health threats but also significantly bolster our overall public health infrastructure. Furthermore, our commitment to interdisciplinary collaboration among researchers, healthcare professionals, and policymakers ensures a robust framework for global health security. Through these comprehensive strategies, we are better equipped to safeguard populations worldwide, ultimately fostering a healthier and more resilient global community.

Ethics approval

The study involved no human participants only open data was used for this research.

Consent to participate

No human participants were involved in this research study.

Consent for publication

The authors affirm that no human research participants provided informed consent for publication.

Competing interests

The authors declare that they have no conflicts of interest to report regarding the present study.

Data and materials

Researchers are bound by rules of ethics, such as collecting data from open sources only.

Author contributions

All authors contributed to the study's conception and design. Material preparation, data collection, and analysis.

REFERENCES

1. Jocelyne, Piret., Guy, Boivin. Pandemics Throughout History. Frontiers in Microbiology, (2021). doi: 10.3389/FMICB.2020.631736
2. Mostafa, Javanian., Mohammad, Barary., Sam, Ghebrehewet., Veerendra, Koppolu., Veneela, Krishna, Rekha, Vasigala., Soheil, Ebrahimpour. A brief review of influenza virus infection. Journal of Medical Virology, (2021). doi: 10.1002/JMV.26990
3. Samuel, V., Scarpino., Giovanni, Petri. On the predictability of infectious disease outbreaks. arXiv: Physics and Society, (2017).
4. Pooja, Bhadoria., Gaurisha, Gupta., Anubha, Agarwal. Viral Pandemics in the Past Two Decades: An Overview.. Journal of family medicine and primary care, (2021). doi: 10.4103/JFMPC.JFMPC_2071_20
5. Jean-Marc, Cavaillon., Marcin, F., Osuchowski. COVID-19 and earlier pandemics, sepsis, and vaccines: A historical perspective. (2021). doi: 10.1016/J.JOINTM.2021.04.003
6. K., Nagaraju, Shivaprakash., Sandeep, Sen., Seema, Paul., Joseph, M., Kiesecker., Kamaljit, S., Bawa. Mammals, wildlife trade, and the next global pandemic.. Current Biology, (2021). doi: 10.1016/J.CUB.2021.06.006
7. Mulugeta, Kiros., Henok, Andualem., Teklehaimanot, Kiros., Wasihun, Hailemichael., Sisay, Getu., Alene, Geteneh., Derbie, Alemu., Woldaregay, Erku, Abegaz. COVID-19 pandemic: current knowledge

about the role of pets and other animals in disease transmission. Virology Journal, (2020). doi: 10.1186/S12985-020-01416-9

8. Amr, El-Sayed., Mohamed, Kamel. Coronaviruses in humans and animals: the role of bats in viral evolution.. Environmental Science and Pollution Research, (2021). doi: 10.1007/S11356-021-12553-1
9. Seyyed-Mahdi, Hosseini-Motlagh. Design of control strategies to help prevent the spread of COVID-19 pandemic. European Journal of Operational Research, (2023). doi: 10.1016/j.ejor.2021.11.016
10. Rachel, J., Davies. A systematic review of observational methods used to quantify personal protective behaviours among members of the public during the COVID-19 pandemic, and the concordance between observational and self-report measures in infectious disease health protection. BMC Public Health, (2022). doi: 10.1186/s12889-022-13819-0
11. Association of Public Health Measures During the COVID-19 Pandemic With the Incidence of Infectious Conjunctivitis. JAMA Ophthalmology, (2022). doi: 10.1001/jamaophthalmol.2021.4852
12. Emerging Technologies Used in Health Management and Efficiency Improvement During Different Contact Tracing Phases Against COVID-19 Pandemic. IEEE Reviews in Biomedical Engineering, (2023). doi: 10.1109/rbme.2022.3219433
13. Sanaz, Sohrabizadeh., Shiva, Yousefian., Amirhosein, Bahramzadeh., Mohammad, Hossein, Vaziri. A systematic review of health sector responses to the coincidence of disasters and COVID-19.. BMC Public Health, (2021). doi: 10.1186/S12889-021-10806-9
14. Jan, C., Semenza., Joacim, Rocklöv., Kristie, L., Ebi. Climate Change and Cascading Risks from Infectious Disease. Infectious Diseases and Therapy, (2022). doi: 10.1007/s40121-022-00647-3
15. Rachel, E., Baker., Ayesha, S., Mahmud., Ian, F., Miller., Ian, F., Miller., Malavika, Rajeev., Fidisoa, Rasambainarivo., Benjamin, L., Rice., Saki, Takahashi., Andrew, J., Tatem., Caroline, E., Wagner., Lin-Fa, Wang., Lin-Fa, Wang., Amy, Wesolowski., C., Jessica, E., Metcalf. Infectious disease in an era of global change.. Nature Reviews Microbiology, (2021). doi: 10.1038/S41579-021-00639-Z
16. Erick, Huarcaya, Castilla., Fiorella, Rossi, Leyva., Alejandro, Llanos-Cuentas. Influencia de factores climáticos sobre las enfermedades infecciosas. Revista Médica Herediana, (2013). doi: 10.20453/RMH.V15I4.776
17. Faburay B, LaBeaud AD, McVey DS, Wilson WC, Richt JA. Current status of Rift Valley fever vaccine development. Vaccines. 2017 Sep 19;5(3):29.
18. Wanyoike F, Mtimet N, Bett B. Willingness to pay for a Rift valley fever (RVF) vaccine among Kenyan cattle producers. Preventive veterinary medicine. 2019 Nov 1;171:104763.
19. Mahmoud MA, Ghazy AA, Shaapan RM. Advances in Diagnosis and Control of Rift Valley Fever in Small Ruminants; Sheep and Goats-A Review. Egyptian Journal of Veterinary Sciences. 2024 Mar 1;55(2):487-98.
20. Uwishema O, Chalhoub E, Torbati T, David SC, Khoury C, Ribeiro LL, Nasrallah Y, Bekele BK, Onyeaka H. Rift Valley fever during the COVID-19 pandemic in Africa: A double burden for Africa's healthcare system. Health science reports. 2022 Jan;5(1):e468.
21. Fatima I, Ahmad S, Abbasi SW, Ashfaq UA, Shahid F, ul Qamar MT, Rehman A, Allemailem KS. Designing of a multi-epitopes-based peptide vaccine against rift valley fever virus and its validation through integrated computational approaches. Computers in biology and medicine. 2022 Feb 1;141:105151.
22. Grossi-Soyster EN, LaBeaud AD. Rift Valley fever: Important considerations for risk mitigation and future outbreaks. Tropical medicine and infectious disease. 2020 Jun 2;5(2):89.
23. Nielsen SS, Alvarez J, Bicout DJ, Calistri P, Depner K, Drewe JA, Garin-Bastuji B, Rojas JL, Schmidt CG, Michel V, Chueca MÁ. Rift Valley Fever–epidemiological update and risk of introduction into Europe. EFSA Journal. 2020 Mar;18(3):e06041.
24. Sarker S, Jamal L, Ahmed SF, Irtisam N. Robotics and artificial intelligence in healthcare during COVID-19 pandemic: A systematic review. Robotics and autonomous systems. 2021 Dec 1;146:103902.

25. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. Chaos, Solitons & Fractals. 2020 Oct 1;139:110059.
26. Pham QV, Nguyen DC, Huynh-The T, Hwang WJ, Pathirana PN. Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: a survey on the state-of-the-arts. IEEE access. 2020 Jul 15;8:130820-39.
27. Kaur I, Behl T, Aleya L, Rahman H, Kumar A, Arora S, Bulbul IJ. Artificial intelligence as a fundamental tool in management of infectious diseases and its current implementation in COVID-19 pandemic. Environmental Science and Pollution Research. 2021 Aug;28(30):40515-32.
28. Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. International journal of health geographics. 2020 Dec;19(1):1-2.
29. Ibrahim NK. Epidemiologic surveillance for controlling Covid-19 pandemic: types, challenges and implications. Journal of infection and public health. 2020 Nov 1;13(11):1630-8.
30. Mubareka S, Amuasi J, Banerjee A, Carabin H, Copper Jack J, Jardine C, Jaroszewicz B, Keefe G, Kotwa J, Kutz S, McGregor D. Strengthening a One Health approach to emerging zoonoses. Facets. 2023 Jan 12;8(1):1-64.
31. Chatterjee P, Nair P, Chersich M, Terefe Y, Chauhan AS, Quesada F, Simpson G. One health,"disease X" & the challenge of "unknown" unknowns. Indian Journal of Medical Research. 2021 Mar 1;153(3):264-71.
32. Janik E, Ceremuga M, Niemcewicz M, Bijak M. Dangerous pathogens as a potential problem for public health. Medicina. 2020 Nov 6;56(11):591.
33. Ijaz MK, Sattar SA, Rubino JR, Nims RW, Gerba CP. Combating SARS-CoV-2: leveraging microbicidal experiences with other emerging/re-emerging viruses. PeerJ. 2020 Sep 8;8:e9914.
34. Chala B, Tilaye T, Waktole G. Re-Emerging COVID-19: Controversy of its zoonotic origin, risks of severity of reinfection and management. International journal of general medicine. 2023 Dec 31:4307-19.
35. Buheji M, Ahmed D. Foresight of Coronavirus (COVID-19) opportunities for a better world. American Journal of Economics. 2020;10(2):97-108.
36. Zhang L, Zhao W, Sun B, Huang Y, Glänzel W. How scientific research reacts to international public health emergencies: a global analysis of response patterns. Scientometrics. 2020 Jul;124(1):747-73.
37. Shiwlani A, Khan M, Sherani AM, Qayyum MU. Synergies of AI and Smart Technology: Revolutionizing Cancer Medicine, Vaccine Development, and Patient Care. International Journal of Social, Humanities and Life Sciences. 2023;1(1):10-8.
38. Majid S, Khan MS, Rashid S, Niyaz A, Farooq R, Bhat SA, Wani HA, Qureshi W. COVID-19: Diagnostics, therapeutic advances, and vaccine development. Current Clinical Microbiology Reports. 2021 Feb 15:1-5.
39. Gupta N, Singh B, Kaur J, Singh S, Chattu VK. COVID-19 pandemic and reimagination of multilateralism through global health diplomacy. Sustainability. 2021 Oct 19;13(20):11551.
40. Gostin LO, Habibi R, Meier BM. Has global health law risen to meet the COVID-19 challenge? Revisiting the International Health Regulations to prepare for future threats. The Journal of Law, Medicine & Ethics. 2020 Jun;48(2):376-81.
41. Gostin LO, Habibi R, Meier BM. Has global health law risen to meet the COVID-19 challenge? Revisiting the International Health Regulations to prepare for future threats. The Journal of Law, Medicine & Ethics. 2020 Jun;48(2):376-81.
42. Gupta N, Singh B, Kaur J, Singh S, Chattu VK. COVID-19 pandemic and reimagination of multilateralism through global health diplomacy. Sustainability. 2021 Oct 19;13(20):11551.
43. Roberts M, Dobson A, Restif O, Wells K. Challenges in modelling the dynamics of infectious diseases at the wildlife–human interface. Epidemics. 2021 Dec 1;37:100523.

44. Harvey E, Holmes EC. Diversity and evolution of the animal virome. Nature Reviews Microbiology. 2022 Jun;20(6):321-34.
45. El-Sayed A, Kamel M. Coronaviruses in humans and animals: the role of bats in viral evolution. Environmental Science and Pollution Research. 2021 Apr;28(16):19589-600.
46. Harris JE. The repeated setbacks of HIV vaccine development laid the groundwork for SARS-CoV-2 vaccines. Health policy and technology. 2022 Jun 1;11(2):100619.
47. Hedman HD, Krawczyk E, Helmy YA, Zhang L, Varga C. Host diversity and potential transmission pathways of SARS-CoV-2 at the human-animal interface. Pathogens. 2021 Feb 8;10(2):180.
48. Khanal P, Bento F, Tagliabue M. A scoping review of organizational responses to the COVID-19 pandemic in schools: A complex systems perspective. Education Sciences. 2021 Mar 10;11(3):115.
49. Jovanović A, Klimek P, Renn O, Schneider R, Øien K, Brown J, DiGennaro M, Liu Y, Pfau V, Jelić M, Rosen T. Assessing resilience of healthcare infrastructure exposed to COVID-19: emerging risks, resilience indicators, interdependencies and international standards. Environment Systems and Decisions. 2020 Jun;40:252-86.
50. Zoumpourlis V, Goulielmaki M, Rizos E, Baliou S, Spandidos DA. [Comment] The COVID-19 pandemic as a scientific and social challenge in the 21st century. Molecular medicine reports. 2020 Oct 1;22(4):3035-48.
51. Sood S, Aggarwal V, Aggarwal D, Upadhyay SK, Sak K, Tuli HS, Kumar M, Kumar J, Talwar S. COVID-19 pandemic: from molecular biology, pathogenesis, detection, and treatment to global societal impact. Current pharmacology reports. 2020 Oct;6:212-27.
52. Javaid M, Haleem A, Singh RP, Suman R, Rab S. Significance of machine learning in healthcare: Features, pillars and applications. International Journal of Intelligent Networks. 2022 Jan 1;3:58-73.
53. Nadella GS, Satish S, Meduri K, Meduri SS. A Systematic Literature Review of Advancements, Challenges and Future Directions of AI And ML in Healthcare. International Journal of Machine Learning for Sustainable Development. 2023 Sep 8;5(3):115-30.
54. Sabry F, Eltaras T, Labda W, Alzoubi K, Malluhi Q. Machine learning for healthcare wearable devices: the big picture. Journal of Healthcare Engineering. 2022;2022(1):4653923.
55. Waring J, Lindvall C, Umeton R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artificial intelligence in medicine. 2020 Apr 1;104:101822.
56. Alzahrani SI, Aljamaan IA, Al-Fakih EA. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. Journal of infection and public health. 2020 Jul 1;13(7):914-9.
57. N. Kumar and S. Susan, "COVID-19 Pandemic Prediction using Time Series Forecasting Models," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225319.
58. Dr. Saeed Q Al-Khalidi Al-Maliki, Dr. Prakash Kuppuswamy**,** Dr. Rajan John, Dr. Nithya Rekha Sivakumar, "COVID-19: An Efficient Big Data Analytics for SARS-CoV-2 Mutations Prediction: A Machine Learning Approach", Indian Journal of Science and Technology 15(21): 1022-1031. https://doi.org/ 10.17485/IJST/v15i21.374, June 2022.
59. Albzeirat M, Zulkepli N, Qaralleh H. A Vision to Face Covid-19 pandemic and Future Risks Through Artificial Intelligence. J. basic appl. Res. Biomed. [Internet]. 2022 Jan. 25 [cited 2024 Aug. 17];6(1):15-20. Available from: https://jbarbiomed.com/index.php/home/article/view/19
60. Maher MC, Bartha I, Weaver S, Di Iulio J, Ferri E, Soriaga L, Lempp FA, Hie BL, Bryson B, Berger B, Robertson DL. Predicting the mutational drivers of future SARS-CoV-2 variants of concern. Science translational medicine. 2022 Jan 11;14(633):eabk3445.
61. Charlotte, J., Houldcroft., Simon, Underdown. Infectious disease in the Pleistocene: Old friends or old foes?. American journal of biological anthropology, (2023). doi: 10.1002/ajpa.24737