

Study and Identification of Various Factors for Outbreak Classification towards Sustainable Healthcare Sector

Ishan Saluja
Department of Computer Science & Engineering
Amity University
Uttar Pradesh
India
ishansaluja11@gmail.com

Shanu Sharma
Department of Computer Science & Engineering
ABES Engineering College,
Ghaziabad
India
shanu.sharma16@gmail.com

Abstract—Recently the world has witnessed a global outbreak, where the spread of Coronavirus was declared a pandemic situation due to its global spread. In past also, epidemic and the pandemic situation has been observed due to SARS, H1N1 influenza, etc. During a global health emergency i.e., during the pandemic situation, there is always a need for a huge number of medical facilities, etc. throughout the world. The shortage of these facilities can cause a loss of a plethora of lives. The spread of any virus or the phase of conversion of an epidemic or endemic situation into a pandemic situation can be tracked by measuring or tracking important parameters, and a timely classification and prediction of a pandemic situation can help the world in every sense. In this paper, an outbreak classification framework is presented to predict the pandemic situation from an endemic situation. To achieve this goal, a rigorous study has been done on various existing resources such as WHO, CDC, etc. to identify important factors for modeling the proposed system. Various factors such as Mode of transmission, Infectiousness, Contagiousness of disease, Incubation Period, Contagiousness of the Asymptomatic Carriers, Severity, Population Density, and Herd Immunity are identified. Furthermore, to test the efficacy of the identified factors, SVM and Random Forest-based classification models are developed for outbreak classification. The identified features are able to generate accuracy of 84.09 and 90.90% which shows their applicability for predicting outbreaks.

Keywords—pandemic, endemic, epidemic, outbreak, classification, machine learning

I. INTRODUCTION

The world has recently faced adversities due to the global outbreak of a novel coronavirus (SARS Cov-2 or Covid-19) [1]. Retrospectively, there have been various epidemics and pandemics also caused by diseases such as SARS, MERS, H1N1 Influenza, etc., [2][3]. An epidemic situation is declared when an infectious disease spreads over a very large area. An endemic refers to the presence of an infectious agent within a particular geographic area or population group. Whereas, a pandemic is declared when an infectious disease has crossed international boundaries and affected a large number of people [4][5]. Whenever there is a global public health emergency, there is a huge demand for medical equipment, testing kits, vaccinations, etc., and it takes a lot of time to meet the demands and supply of resources to various countries [6][7]. It also causes a loss of a plethora of lives thus; it becomes necessary to evade a predicament like this at all costs. Therefore, we need to have a prescience and must

be able to predict a pandemic whenever a disease is found in its nascent stage.

Over the past decades, researchers from all over the world have been studying the disease dynamics of various diseases like influenza viruses, viruses from the family of coronaviruses, etc. [8][9]. Advancing technologies such as Artificial Intelligence (AI) and Machine Learning (ML) have also performed various maneuvers in the healthcare sector [8]. These technologies have been used widely for disease identification and for making more accurate diagnoses. They are used to make smart records of patients and analyze large chunks of data to enhance the quality of health information [9]. AI and ML are also being used to analyze disease dynamics and predict pandemics, i.e., they are being used to scrutinize the disease outbreaks in their nascent stage and predict how fast and how many diseases could spread [10][11]. Various models and algorithms have also been developed to predict the potential hotspots of disease outbreaks [12].

This paper explicates a method to predict a pandemic from an endemic i.e., predicting whether an incipient disease has the capability to cause a global outbreak and become a pandemic. Having the power and capability to do so will help the world and various health organizations to take necessary precautions beforehand and a plethora of lives could be saved, this would also obviate the nations, businesses, and jobs of people from getting adversely affected by the recession. Motivated by this, in this paper, a rigorous study has been done on various existing resources such as World Health Organization (WHO), Centre for Disease Control and Prevention (CDC), etc. to identify various indispensable factors which affect the dynamics of a disease's spread. Furthermore, various machine learning models were inspected, out of which two models i.e., Random Forest and Support Vector Machine (SVM) produced promising results and were able to predict a pandemic with high accuracy.

The work presented in this paper is structured as: Section 2 represents the background details related to the topic of interest along with the discussion on some of the related existing work. Section 3 presents the proposed approach which includes a description of various identified factors followed by the development of the outbreak classification model. Various results obtained during the development of the model along with the justification of achieved accuracy for outbreak classification are also presented here. The work presented in this paper is concluded in Section 4.

II. BACKGROUND

Recently, the world has faced an extreme pandemic situation due to the spread of coronavirus. Due to the extremely contagious nature of the pandemic, person-to-person contact hastens its transmission, resulting in the outbreak of COVID-19 [2]. Between December 2019 and early 2022, the outbreak infected more than 200 million people worldwide, posing a grave threat to human life and health [2]. The number of cases and deaths was far more than those that were seen in epidemics and pandemics like SARS, MERS, and H1N1 Influenza [3].

When COVID-19 first arose, the most important worry was how to limit the infection and safeguard billions of global citizens without jeopardizing the international economy, which would be severely impacted if nations adopted lockdown and quarantine laws. Numerous nations across the globe have undergone and continue to endure economic difficulties as a result of national/regional lockdown periods lasting a few days to hundreds of days [13]. For the economy to revive, unrestricted movement must be permitted. This implies that authorities should be able to promptly identify the possible contacts of any newly identified infected cases [14]. In this case, precise forecasting of such situations is required to swiftly contain the pandemic [15].

Before discussing the literature further, it is indispensable to understand the meaning of the terms: epidemic, endemic, and pandemic. An epidemic is the occurrence of a communicable disease in a community or region. It refers to a sudden increase in the spread of disease than what is normally expected. An endemic refers to the presence of an infectious agent within a particular geographic area or population group, without importation from outside when conditions are favourable, and an endemic may burst into an epidemic [16]. Whereas, a pandemic is an epidemic occurring over a very wide area, crossing international boundaries and affecting a large number of people. For a disease to be pandemic it must be infectious and contagious, i.e., it can infect humans and can spread easily from one person to another [4][5][16].

A disease outbreak is said to happen when there is a rise in the number of endemic cases. Whenever a disease outbreak rises really fast or exponentially, it may cause an epidemic or a pandemic and therefore, the demands for medical equipment and necessary healthcare products also increase. The affected regions are hit by various adversities, people lose their lives, nations face recession, businesses stumble, and various other repercussions that common people have to face. Due to this, it is important to predict a pandemic so that the world can be saved from the repercussions of public health emergencies [6][7].

Extensive new studies show that population movement and geographic distribution are key factors in the transmission of infectious diseases [13][14][15]. A measure in epidemiology, known as the Basic Reproduction Number R_0 can be used to determine how fast a disease can spread [17]. It was adopted for use by epidemiologists, from the field of demography where it was first introduced. It tells the average number of people an infected person can transmit the disease to, considering a population where everyone is susceptible to the disease. The '0' in the subscript of R_0 signifies that no one in the region is immune to the disease or has been vaccinated for the disease [17][18]. Various factors

are employed to calculate the value of R_0 such as average contact rate, duration of infectiousness, and mode of transmission. If $R_0 > 1$ the disease will spread at a high rate and might cause a global outbreak, but if $R_0 < 1$ then the disease will spread at a slow rate and might die out. The value of R_0 is calculated using mathematical models, but these calculations are performed on the basis of various assumptions and guesses because only a paucity of information is available during the initial stage of a disease outbreak [16-17].

Artificial Intelligence (AI) and Machine Learning (ML) technologies are being used widely nowadays in the healthcare sector. They have also been rigorously used to predict a pandemic and the effects of various diseases on different regions [8]. Researchers have been studying and explicating various methods to apprehend the dynamics of the spread of various diseases. Many scholars have tried to predict a pandemic using messages and tweets on social media and applying various machine learning and deep learning algorithms to the big flow of data [9][10]. In [21], authors used prediction markets and Twitter to predict a swine flu pandemic. They aggregated the convictions of people on Twitter about the world using a prediction market. The results show that information available on social media has the potential to be used as an envoy for public opinion [19]. Authors have used machine learning to predict a pandemic in [20], where the data from social media (Twitter) was collected and various machine learning techniques were applied to analyze the big flow of messages. Three ML models were applied to predict a pandemic.

The work done in the field of analyzing the age distribution of deaths in the United States caused by influenza is presented in [21]. Here, the authors observed decreasing pattern for the mortality rate of people <65 years of age during this course of pandemics and epidemics and they have also predicted the mortality rate for the next influenza pandemic based on age group. In [22], the authors discussed population consequences due to Universal Influenza Vaccines (UIVs). They use mathematical modelling to explicate the adverse consequences of vaccination programs that are focused on specific types of immunity. The researchers have also attempted to predict the Covid-19 outbreak using machine learning and soft computing techniques [23]. There is a high degree of uncertainty and only a paucity of information available, due to which standard models of epidemiology do not manifest much accuracy for long-term prediction. Due to this reason, here authors manifest the efficacy of two machine learning models to model the time series of the Covid-19 outbreak [23].

To date, no such method has been proposed which could predict a pandemic with utmost accuracy. Due to this reason, the world needs a method or a model which can accurately predict if a disease has the potential to cause a global outbreak. The methodology of the proposed outbreak classification and prediction framework is presented in the next section.

III. PROPOSED WORK

A pandemic is an extremely ravaging ordeal. Therefore, it becomes necessary to predict a pandemic so that an incipient disease can be detected timely anywhere in the world. To achieve this goal, in this paper a systematic flow

presented in Fig 1 is adopted. First, a study was initiated by understanding the previous work done in this field. Afterward, various articles and other written materials were obtained from various sources such as the WHO, CDC, and numerous other research institutes, etc. The various factors which affect and influence the disease dynamics were studied and analyzed [23][24]. After a rigorous study, various factors are finalized which are mentioned in the subsequent subsection. After the finalization of factors, the corresponding dataset is constructed and modeled to apply machine learning algorithms to test the efficacy of these factors for outbreak prediction.

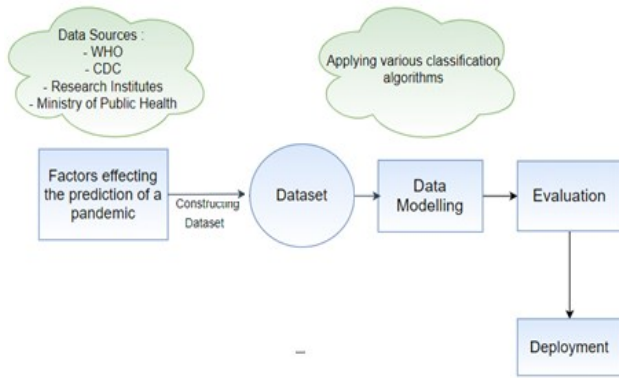


Fig. 1. Proposed Approach

A. Factors for Outbreak Classification

The degree of spread of any disease can be classified as an endemic, epidemic, or pandemic. This classification can be made using the following factors:

- Mode of Transmission
- Incubation Period of the disease
- Severity
- Infectiousness and Contagiousness
- Contagiousness of Asymptomatic Carriers
- Herd Immunity and Population Density

It should be noted that these factors are exclusive to this study and have contributed to the construction of an entirely original dataset in order to fulfill the objective of the research. Below are a few examples that have been quoted in order to emphasize the individualistic importance of each of the above factors, alongside the basis on which they have been characterized and selected.

a) *Mode of Transmission*: It is the manner in which a disease may be transmitted from one individual to the other, the mode of transmission manifests itself in a wide variety of voluntary and non-voluntary actions, this includes sexual contact, pregnancies, mucous, micro-droplets from cough or sneeze, the saliva of the infected person, etc. This factor can be broadly classified into three attributes that have been assigned a particular digit, a diagrammatic representation of which has been shown in Fig 2.

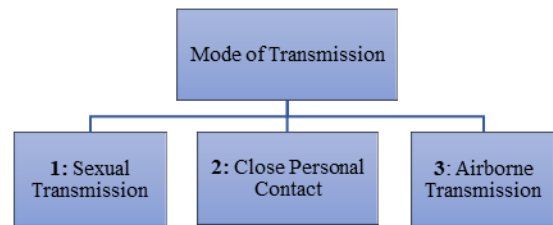


Fig. 2. Type of Mode of Transmission

If a particular disease has been assigned the number '1' under the mode of transmission column in the dataset, it suggests that it spreads through sexual contact. Similarly, if it has been assigned the number '2', it suggests that close personal contact with foreign mucous membrane, broken skin etc. can lead to the transmission of the disease under observation. Number '3' describes the subset of diseases that are characterized by airborne transmission and spread through micro-droplets due to coughing or sneezing or viruses/bacteria living on surfaces for prolonged time periods. The above classification allows us to account for the vast majority of contagious diseases that are or once were in existence, covering a whole spectrum of diseases that can be transmitted in several distinctive ways.

b) *Mode of Transmission*: The possibility and degree of spread of a particular disease, not all diseases are deemed contagious, for e.g., cancer is not contagious however viruses such as the common cold, coronavirus etc are communicable through micro-droplets. This factor states whether a disease can be transmitted or not. There are some diseases that cannot be directly transmitted or transmitted at all, the subset of these diseases has been assigned the number '0' in the dataset, however, if a disease is capable of being transmitted, it has been assigned the number '1'. This factor has been selected because it directly offers insight into whether a disease can be transmitted from an infected individual to a healthy person or not. For example, the value of this factor for cancer would be '0' as it cannot be transmitted from one individual to another, whereas, for Covid-19, its value would be '1', as presented in Fig 3.

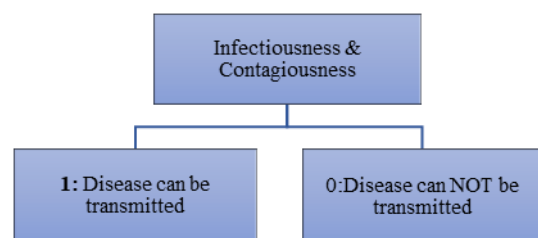


Fig. 3. Possible types of Infectiousness and Contagiousness

c) *Contagiousness of Asymptomatic Carriers*: Whether or not an asymptomatic (an infected person displaying no symptoms) carrier holds the capability to infect a healthy individual. An asymptomatic carrier, for descriptive purposes refers to the host of a particular disease that displays no signs and symptoms of that disease. This factor is based on a Boolean condition that simply states whether a disease can be contracted through a carrier that is in the asymptomatic phase. This has been described by assigning the number '0' to the disease if it does not hold the capability of infecting another individual whilst being in

its asymptomatic phase, and ‘1’ should it hold the capability of being transmitted during the same period, as shown in Fig 4. This factor is considered in this research because transmission of a disease from an asymptomatic carrier plays a crucial role in determining the status of pandemic.

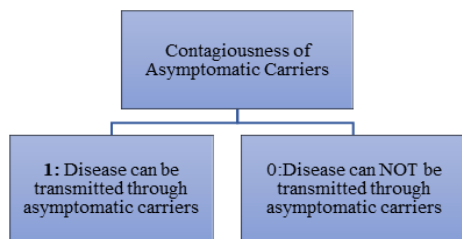


Fig. 4. Classification and assigned values for Contagiousness of Asymptomatic Carriers

d) *Incubation Period*: This is the time period between the contraction of the disease and the actual manifestation of its symptoms, i.e., The time elapsed between an individual contracting a disease and them showcasing symptoms of the disease. For e.g. An individual infected with the coronavirus may take 1-2 weeks before displaying any symptoms of the virus. This is simply mentioned in the dataset in the form of the number of days. The reason behind the selection of this factor is related to the degree of spread through asymptomatic carriers, in simple words, an individual may not be aware that they have contracted a particular disease when it is in its incubation period, which may deceive him or her into believing that they are well enough to establish regular social contact with other individuals, in such cases, every individual that establishes any form of contact with the asymptomatic carrier is at risk of contracting the disease. This poses a massive threat to large regions.

e) *Severity*: Severity represents the degree of lethality of a disease. This factor refers to the spectrum of adverse effects a disease can have on an individual, ranging from mild to moderate to severe virulence that has been correspondingly assigned the numbers ‘1’, ‘2’, and ‘3’ respectively, as shown in Fig 5. The reason behind the selection of this factor is to acquire some insight into the lethality of the disease and how adversely it will affect the infected population. It is to be noted that, the more the severity of the disease, the less likely it would be for the disease to inflict a pandemic and vice versa; the reason is that if the severity of a disease is very high, the infected individual would tend to stay at home and rest or get hospitalized (this was the case with Ebola Virus), whereas, less the severity of a disease, more the infected individual is likely to go out, and that would ultimately spread the disease (this was the case with H1N1 Influenza Virus).

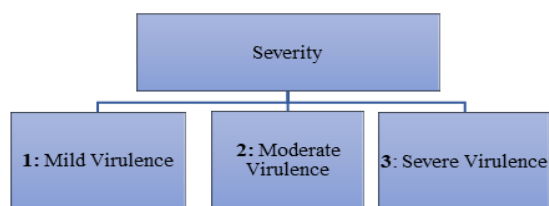


Fig. 5. Classification and assigned values for Severity Factor

f) *Population Density (Pd)*: This refers to the number of individuals per square kilometer in a region, this factor was taken into consideration in order to include the number of people that can be infected due to a particular disease as well as to estimate how many other individuals can be infected due to one sole individual. This offers great insight into determining how fast a disease can spread.

g) *Herd Immunity (Hi)*: This refers to the immunity level of the population in a region to a particular disease. It represents the percentage of the population who is immune to a disease or the resistance to the contraction of any disease, arising due to a large chunk of the population being infected previously or vaccinated for a particular disease. This factor has been selected in order to estimate the effects of a particular family of viruses and to what degree is it capable of affecting an entire region’s population; if the population in a particular region acquires herd immunity to a particular virus, chances are that the virus will die out, which diminishes the probability of a pandemic.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The factors presented in the previous section were taken into consideration in order to fulfill the construction of an entire original dataset of diseases that were evaluated on the basis of the above factors.

TABLE I. IDENTIFIED FACTORS AND POSSIBLE VALUES

Factors	Possible Values		
	1	2	3
Mode of Transmission (MT)	Sexual Transmission	Close Personal Contact with Mucous Membranes, Broken Skin, Urine/Saliva of the infected	Airborne Transmission, Micro-droplets due to coughing or sneezing, viruses/bacteria living on surfaces for prolonged time periods
	Infectiousness & Contagiousness (IC)	1 The disease under observation is contagious and infectious.	0 The disease under observation is neither contagious nor infectious.
Contagiousness of Asymptomatic Carriers (CAC)	1 Asymptomatic carriers can cause transmission	0 Asymptomatic carriers can not cause transmission.	
	Incubation Period (IP)	Time elapsed (in days) between the contraction of disease and the display of symptoms	
Severity (S)	1 Mild Virulence: Common Cold, H1N1	2 Moderate Virulence: Covid-19	3 Severe Virulence: Ebola Virus
	Population Density (PD)	Population density of the region where the disease was found in its nascent stage	
Herd Immunity (HI)	Percentage of the population immune to the disease		

To construct the dataset, various diseases like SARS, MERS, Ebola virus, etc. were considered. Then, the values to

the following factors: Infectiousness and Contagiousness, Incubation Period, Contagiousness of the Asymptomatic Carriers, Herd Immunity for each of these diseases, were gathered from various authenticated sources like WHO, CDC, numerous research institutes, ministries of public health, etc. The values for Mode of Transmission and Severity of disease were given as: '1', '2', or '3' depending upon the classification they fall into. The values for a population density of the region were gathered from the internet separately. After the collection of the data, the dataset was prepared, and the modalities to assign values to all the factors, corresponding to each disease, are explained in Table 1. Furthermore, a glimpse of the constructed dataset can be seen in Fig 6.

SNo.	Disease	Mode of Transmission(H/M/L)	Population Density	Severity	Incubation Period (in days)	Contagiousness of Asymptomatic Carriers	Infectious and Contagious	Herd Immunity	Pandemic
1	Covid-19	3	1306.74	2	5	1	1	0	2
2	SARS	2	46.67	2	2	0	1	0	1
3	MERS	2	14	2	5	0	1	0	1
4	H1N1	3	31.16	1	1	1	1	0	2
5	Ebola Virus	2	50.5	3	2	0	1	0	1
6	Mumps	3	60	1	4	1	1	0	2
7	HIV-AIDS	1	70	2	15	1	1	0	1
8	H1N1	3	40	1	2	1	1	0	2
9	Covid-19	3	50	2	2	1	1	0.2	2
10	Covid-19	3	38	2	3	1	1	0.26	2
11	Covid-19	3	112	2	7	1	1	0.54	1
12	Covid-19	3	200	2	10	1	1	0.11	2
13	Covid-19	3	1306	2	6	1	1	0.9	0
14	Covid-19	3	1250	2	9	0	0	0.5	0
15	Covid-19	3	800	2	4	0	1	0	1
16	Covid-19	3	1000	2	5	0	0	0	0
17	Covid-19	3	1306.74	2	5	1	1	0.82	0
18	Covid-19	3	1000	2	3	1	1	0.76	0
19	SARS	2	100	2	1	1	1	0.09	1
20	SARS	2	58	2	2	0	1	0	1
21	SARS	2	46.67	2	5	1	1	0.12	2
22	SARS	2	50	2	2	0	1	0	1
23	SARS	2	46.67	2	4	1	1	0	2
24	SARS	2	50	2	6	1	1	0.78	0
25	SARS	2	150	2	5	0	0	0	0

Fig. 6. A Glimpse of Constructed Dataset

To test the efficacy of the identified factors, two Machine Learning Algorithms are chosen i.e., Support Vector Machine (SVM) and Random Forest. The support vector machine algorithm leverages the power of the polynomial kernel, support vector classifiers and soft margins to classify data points. Classifications can be made using a threshold that lies on the midpoint of the distance between the edges of the different clusters of data, as shown through the red and green data points example below. The Random Forest Algorithm leverages the power of decision trees, as well as appends the new feature of flexibility to them, increasing the algorithm's power tenfold. It works on the principle of constructing decision trees on the basis of a bootstrapped dataset.

The Random Forest and SVM algorithms manifested accuracies of 84.09 % and 90.90 % respectively as shown in Figure 7 and 8, these results can be extrapolated to several diseases in order to draw out accurate disease classifications. The confusion matrices illustrate the accuracy of the algorithms in use, as shown below. This broaches the discussion regarding the usefulness of these results, and their applicability for outbreak classification.

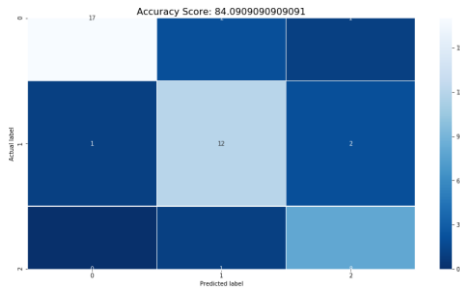


Fig. 7. Model accuracy using Random Forest Classifier

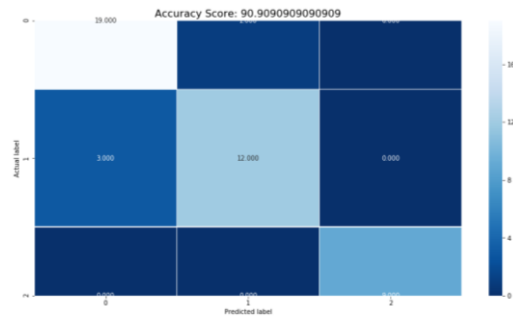


Fig. 8. Model accuracy using SVM Classifier

V. CONCLUSION

In this paper, an attempt is made to classify the outbreak situation using Machine learning approaches. This was done by building a custom dataset through various authentic resources available on the WHO and CDC websites, as well as the websites of other authentic research institutes. The results obtained are indicative of a clear classification being made when new data-points belonging to the test-set are introduced. We tested several different algorithms, which narrowed the candidates down to the Random Forest and SVM algorithms, which yielded accuracies of 84.09 % and 90.90 %, respectively. These results serve the essential purpose of classifying a disease into one of the three disease-spread classifications; endemic, epidemic or pandemic. The usefulness of these results showcases itself during the present-day scenario; where humanity walks alongside a deadly virus, if accurate classifications akin to this, are made corresponding to most diseases, the biological threats against humanity may dwindle away to a large extent, otherwise, innumerable devastating consequences may ripple through humanity for generations to come.

REFERENCES

- [1] World Health Organization, "Coronavirus disease (COVID-19)," World Health Organization, 2021. https://www.who.int/health-topics/coronavirus#tab=tab_1 (accessed Dec 5, 2022)
- [2] World Health Organization, "Coronavirus Disease (COVID-19) - Events as They Happen," www.who.int, Jul. 31, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen> (accessed Nov 15, 2022)
- [3] World Health Organization, "Coronavirus disease (COVID-19) pandemic," World Health Organization, 2022. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed Dec 15, 2022)
- [4] L. V. Allen Jr. "PreScriptio: Endemics, Epidemics, and Pandemics. International journal of pharmaceutical compounding", vol. 25, no. 4, pp. 268, 2021.

- [5] P. P. Bassareo, M. R. Melis, S. Marras, and G. Calcaterra, "Learning from the past in the COVID-19 era: rediscovery of quarantine, previous pandemics, origin of hospitals and national healthcare systems, and ethics in medicine," *Postgraduate Medical Journal*, vol. 96, no. 1140, pp. 633–638, Oct. 2020, doi: 10.1136/postgradmedj-2020-138370.
- [6] H. M. Chung, "Blockchain-based Value Creation for Supply Chain in COVID-19 Drug Distribution," 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), Madrid, Spain, 2022, pp. 1-6, doi: 10.23919/CISTI54924.2022.9820529.
- [7] L. Alla, B. Bentalha and N. Bouhtati, "Assessing Supply Chain Performance in the Covid 19 Context: A Prospective Model," 2022 14th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA), EL JADIDA, Morocco, 2022, pp. 1-6, doi: 10.1109/LOGISTIQUA55056.2022.9938083.
- [8] S. Srivastav, P. K. Singh, and D. Yadav, "An Approach for fast Compressed text Matching and to avoid false Matching using WBTC and Wavelet Tree," *EAI Endorsed Trans. Scalable Inf. Syst.*, vol. 30, pp. 1–10, 2021.
- [9] S. Agrebi and A. Larbi, "Use of artificial intelligence in infectious diseases," *Artificial Intelligence in Precision Health*, pp. 415–438, 2020, doi: 10.1016/B978-0-12-817133-2.00018-5.
- [10] Y. Cheng, Y. Bai, T. Xu and H. Hu, "Global COVID-19 development trend forecast based on machine learning," 2022 International Conference on Big Data, Information and Computer Network (BDICN), Sanya, China, 2022, pp. 128-131, doi: 10.1109/BDICN55575.2022.00032.
- [11] R. Setiawan, Z. E. Rasjid and A. Effendi, "The Effect of World Happiness Aspects with COVID-19 Infected using Machine Learning Multiple Regression Model," 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa, 2021, pp. 1-4, doi: 10.1109/ICECET52533.2021.9698628.
- [12] V. K. Gupta, A. Gupta, D. Kumar and A. Sardana, "Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model," in *Big Data Mining and Analytics*, vol. 4, no. 2, pp. 116-123, June 2021, doi: 10.26599/BDMA.2020.9020016.
- [13] B. Murugesan, S. Karuppanan, A. T. Mengistie, M. Ranganathan, and G. Gopalakrishnan, "Distribution and trend analysis of COVID-19 in India: Geospatial approach," *J. Geographical Stud.*, vol. 4, no. 1, pp. 1–9, Apr. 2020, doi: 10.21523/gcj5.20040101.
- [14] J. M. Read, J. R. E. Bridgen, D. A. T. Cummings, A. Ho, and C. P. Jewell, "Novel coronavirus 2019-nCoV (COVID-19): Early estimation of epidemiological parameters and epidemic size estimates," *Phil. Trans. R. Soc.*, 2021, Art. no. B3762020026520200265.
- [15] B. Tang, X. Wang, Q. Li, N. L. Bragazzi, S. Tang, Y. Xiao, and J. Wu, "Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions," *J. Clin. Med.*, vol. 9, no. 2, p. 462, Feb. 2020, doi: 10.3390/jcm9020462.
- [16] M. E. K. Niemi, M. J. Daly, and A. Ganna, "The human genetic epidemiology of COVID-19," *Nature Reviews Genetics*, May 2022, doi: 10.1038/s41576-022-00478-5.
- [17] "What Is R0? Gauging Contagious Infections," *Healthline*, Apr. 20, 2020. <https://www.healthline.com/health/r-nought-reproduction-number/#prevention>
- [18] P. L. Delamater, E. J. Street, T. F. Leslie, Y. T. Yang, and K. H. Jacobsen, "Complexity of the Basic Reproduction Number (R0)," *Emerging Infectious Diseases*, vol. 25, no. 1, pp. 1–4, Jan. 2019, doi: 10.3201/eid2501.171901.
- [19] J. Ritterman, M. Osborne, and E. Klein. "Using prediction markets and Twitter to predict a swine flu pandemic." 1st international workshop on mining social media. Vol. 9. 2009.
- [20] T. Saba, I. Abunadi, M. N. Shahzad, and A. R. Khan, "Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types," *Microscopy Research and Technique*, Feb. 2021, doi: 10.1002/jemt.23702.
- [21] L. Simonsen, M. J. Clarke, L. B. Schonberger, N. H. Arden, N. J. Cox, and K. Fukuda, "Pandemic versus Epidemic Influenza Mortality: A Pattern of Changing Age Distribution," *Journal of Infectious Diseases*, vol. 178, no. 1, pp. 53–60, Jul. 1998, doi: 10.1086/515616.
- [22] N. Arinaminpathy, S. Riley, W. S. Barclay, C. Saad-Roy, and B. Grenfell, "Population implications of the deployment of novel universal vaccines against epidemic and pandemic influenza," *Journal of The Royal Society Interface*, vol. 17, no. 164, p. 20190879, Mar. 2020, doi: 10.1098/rsif.2019.0879.
- [23] S. F. Ardabili et al., "COVID-19 Outbreak Prediction with Machine Learning," *Algorithms*, vol. 13, no. 10, p. 249, Oct. 2020, doi: 10.3390/a13100249.
- [24] S. S. Morse, "Factors in the Emergence of Infectious Diseases," *Emerging Infectious Diseases*, vol. 1, no. 1, pp. 7–15, Mar. 1995, doi: 10.3201/eid0101.950102.