

Transforming Native Epidemic Models by Using the Machine Learning Approach

Sathwik Amburi¹, Vanshika Jalan², Dr.S. Saravanan³

¹Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.E-mail: sathwik.amburi@gmail.com

²Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.E-mail:vanshikajalan26@gmail.com

³Assistant Professor (Sr.G), Computer Science and Engineering, SRM Institute of Science and Technology Chennai, India. E-mail: saravans2@srmist.edu.in

ABSTRACT

Since the dawn of civilization, disease outbreaks have always been a threat to humanity. Despite ongoing efforts to improve health care around the world, epidemic outbreaks remain a major health concern. The issue with an epidemic is that pathogen dissemination relies on numerous socio-environmental variables that are changing all the time, rendering epidemic models obsolete. A successful response to such outbreaks requires prompt action based on a variety of factors such as the influence of environmental conditions, incubation period, infection rate, etc. Conventional epidemic models often produce results that are essentially hand-coded to respond to inputs that way. In the event of an unforeseen circumstance, these models do not adapt well because they lack independent discernment. Furthermore, due to the variety of data-forms, issues, and strategies used to address these models- collecting, visualizing, and interpreting epidemic data is becoming more challenging. We hope to survey existing methodologies used to fight epidemics and intend to propose a data-driven approach by breaking down existing epidemic models, which will learn to react to a specific outcome and adapt to new data and varying variables. We conclude by emphasizing how this strategy can be optimized and adapted for potential epidemics, allowing governments and organizations to plan for the next outbreak and educate citizens against high-risk activities.

KEYWORDS

Epidemic Models, SEIRD MODEL, Compartmental Models, Deep Learning, Machine-learning, Susceptible, Infected, Recovered, Dead, Compartments,etc.

Introduction

Epidemic models reveal how infectious pathogens spread and serve as a useful tool for health professionals and governments to control the epidemic. [9]These models assist in the detection and quantification of the results of various initiatives such as mass vaccination programmes, social distancing steps, and lockdowns. They also assist in determining which strategies to prevent and which to test or forecast potential trends of spread, etc. Models are only as strong as the assumptions they are built on, and conventional epidemic models, sadly, depend on assumptions that only hold true in an ideal world. For example, these models assume a rectangular/ stationary age distribution and a homogeneous mixing of the population, which doesn't hold in countries with a diverse population. By introducing independent discernment to these models with the help of “machine learning and data science”, epidemic models will perform and adapt well in real-world scenarios.

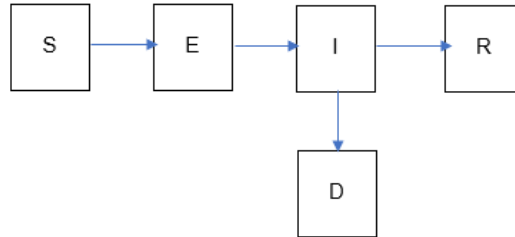
In this research paper, we discuss the [10] “machine-learning” approach aiming to base model selection, evaluation, and prediction on the analysis of the data acquired. We implement this approach on the current ongoing pandemic: Covid-19, as it seems the perfect opportunity to test our model.

We need to know how viruses spread from one individual to another when modelling epidemics; we should try to figure out how rapidly they spread, how much of the population they infect, who dies, and so on. One of the most powerful ways to comprehend this is to use a compartmental model. The compartmental model divides the population into multiple states or compartments. In this paper, we divide the population into the following compartments:

- Susceptible(S): This compartment is comprised of people who are at risk of being infected with the virus [20].
- Exposed(E): This compartment is comprised of people who have been exposed to the virus, but have not yet contracted it [20]. (This refers to the population in the incubation period of the virus.)
- Infected(I): This compartment contains a virus-infected population that is capable of infecting the susceptible

population[11] [20].

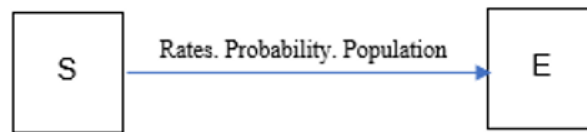
- Recovered(R): This compartment is comprised of people who were infected with the virus and have recovered.
- Dead(D): This compartment is comprised of people who were infected with the virus and have died as a result of it. [12]



To gain a deeper understanding of the compartments mentioned above. The following assumptions can be considered. We have $S = 6000$ people, $E = 2000$ people, $I = 1500$ people, $R = 1000$ people, and $D = 5000$ people in an assumed population of $N = 10000$ people. As a consequence, N (Total Population) = $S+E+I+R+D$. The population transitions from one compartment to the next compartment due to the following factors:

- Rate: The rate specifies how long it takes for the transition to take place.
- Probability: The probability of a transition occurring for a person is called probability.
- Population: The population is the group of people that will be affected by this change.

We think of compartments as boxes (or "states"), with arrows indicating transitions from one compartment to the next. For Example:



**Later in the paper, we consider the "Critical" compartment to better conform the model to the data. The "Critical" compartment consists of people in the intense care unit.

A. Parameters and Transition Rates

The following parameters and rates must be defined in order to derive the transition between compartments:

- N : Sample Size or Total Population.
- $S(t)$: On day t , the number of people who are at risk.
- $E(t)$: On day t , the number of people who were exposed to the virus.
- $I(t)$: On day t , the number of people who were infected.
- $R(t)$: The number of people who were able to recover on day t .
- $D(t)$: On day t , the number of people who died.
- β : The number of people an infected person is expected to infect per day.
- D : The number of days that a person who has been infected is contagious and may transmit the disease.
- γ : The percentage of infected people who recover per day. ($\gamma = 1/D$)
- R_0 : The total number of people infected by a single infected person. ($R_0 = \beta / \gamma$)
- δ : The amount of time it takes for a virus to incubate.
- α : % of deaths. (fatality rate)
- ρ : The rate at which people die (= 1/days from infected until death)

B. Deriving Equations

We now want to know how many people are infected, susceptible, exposed, recovered and dead for all days. It is difficult to find a definite formula for $S(t)$, $R(t)$, and other parameters. [13]The regular change in S , E , I , R , and D , on the other hand, is very simple to comprehend; we can use ordinary differential equations to work out how transitions occur based on existing data.[20] The following series of ordinary differential equations can be used to express them:

$$\begin{aligned}dS/dt &= -\beta.I.S/N \\dE/dT &= \beta.I.S/N - \delta.E \\dI/dT &= \delta.E - (1 - \alpha). \gamma.I - \alpha. \rho.I \\dR/dT &= (1 - \alpha). \gamma.I \\dD/dT &= \alpha. \rho.I\end{aligned}$$

C. Thesis Statement

In general, epidemiologists measure rate predictions and transition probabilities, then input the values into compartmental models to generate graphs and visualizations. [14]However, we propose to take advantage of recent advances in machine learning to forecast these rates and probabilities to obtain reliable results that can respond to minor changes in socio-environmental factors like lockdowns, social distancing measures etc.

Literature Survey

A. Literature Review

[1] Fokas AS et al. proposes a method for forecasting the time progression of the total number of people infected with SARSCoV-2 in a given country. The analytical formulae include a few steady parameters decided from the accessible information utilizing an error-minimizing calculation. The methodology is based on deep learning networks and mathematical calculations that yield results whose input is the existing information within the given nation. They've addressed the SIR Model and haven't looked at the perspectives of the class of people who have been subjected to the infection. [6] Jocelina Lega et al. focuses on identifying a basic property that appears to be normal in the epidemiological bends of several flare-ups and investigating the modelling effects of this finding. It simply allows us to depict the progression of each outbreak using a very simple model whose two parameters can be derived from epidemiological data. It doesn't assess infection transmission rate because the knowledge on the model isn't sufficient to recover the boundaries of a re-enacted epidemic that follows SIR elements. [3] Baoquan Chen et al. discuss the evaluation of control measures using a new epidemic dynamic model (C-SEIR demonstration) to better consider and modulate epidemic dynamic behaviour. But it doesn't utilize profound deep learning models to assess the simulation and forecast of the spread designs that can be used to develop the model. [2] Binti Hamzah FA et al. using SEIR modelling to predict COVID-19 outbreak within and outside of China based on daily perceptions[15]. Real-time data querying and visualization are performed, and the resulting information is used for Susceptible-Exposed-Infectious-Recovered pattern recognition. They have analysed the queried news and have classified them into negative and positive opinions to understand its impact on people's political and financial behaviour. The model introduced in the paper is hand-coded and doesn't adjust well within the occasion of an unforeseen circumstance. Also, contamination rate and recuperation rate are not determined progressively. [4] Alberto Godio et al. used a modified variant of the traditional SEIR model by dealing with the model in a stochastic manner and monitoring the spread of the model arrangement's flaws. The methodology presents time-subordinate model boundaries. They have used the Computational Swarm Intelligence technique to give a set of model arrangements as likely situations determined by methods for adensity distribution. But it is sensitive, for example, a little change in the straightforward guidelines brings about distinct group-level behaviour. [5] They discuss significant challenges associated with the transmission of productive data analysis pipelines. They have concentrated on laying out the context of outbreak response and surveying the common components, their interdependencies, and how each component can be used to educate various stages of an outbreak response. [8] Yingxue Li et al. have used a RNN based transfer learning approach to predict the spread patterns of various nations. [4]The introduction of lockdown measures and the use of the consideration instrument into the models indicated critical improvement in expectation performance. [7] Can Hou, Jiabin et al. have attempted to ascertain the efficacy of isolating Wuhan from the epidemic using models based on person-to-person transmission [16]. They did not take into consideration the rate of change in individual

movements, such as the wearing of masks, expanding of social distancing, and the cessation of travel to Wuhan.

B. Inference from the Survey

[2]Outbreak analytics is a dynamic field, and there is a big gap in terms of information collection, investigation, and detailing tools. Models are just comparable to the presumptions on which they are based. On off the chance that a model makes expectations that are off the mark with noticed outcomes and the science is right, the underlying suspicions must change to make the model helpful, but this isn't the situation for existing models (Jonathan A. et al.,2018, Alberto Godio et al.,2020). The models can be developed by using profound deep learning models to assess the simulation and forecast of the spread designs (Baoquan Chen et al.,2020, Binti Hamzah FA et al.,2020). Therresults produced by the epidemic models are hand-coded to reply to inputs that way. Within the occasion of an unanticipated circumstance being experienced, these models do not adapt well since they need autonomous insight. Moreover, the collection, visualization, and translation of spread information are getting to be more challenging due to the assortment of information forms, issues, and strategies utilized to address them.

C. Analytical Insights

By predicting the trajectory of the disease spread, epidemic models can help predict the likely outcome of an epidemic and direct public health initiatives. Models utilize fundamental assumptions or collected insights alongside arithmetic to discover boundaries for various infectious diseases and utilize those parameters to calculate the impacts of distinctive mediations, like mass vaccination programs. The modelling can offer assistance, choose which intervention/s to dodge and which to implement or can anticipate future growth patterns, etc.

D. Demerits

Although the models and methods we've considered for the literature survey are intuitive, they don't acknowledge certain elements that impact the prediction and spread of the virus. Models surveyed do not consider factors like lockdown, social distancing implications on the infection spread, etc. With AI and Data Science advancements we believe there can be improvements in epidemic models and predictions.

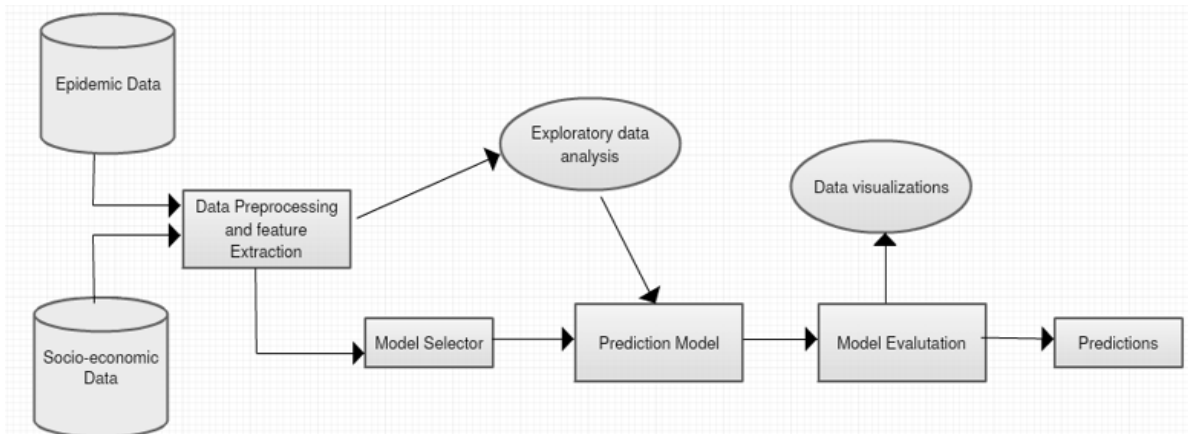
Proposed Work

A. Description

Our proposed system will develop a streamlined approach from collecting to implementing epidemic models, which can be scaled and adjusted for any epidemic. The ability to pre-process datasets and select the epidemic model required to address pathogen dissemination is a crucial attribute of our proposed system. The system would also understand the consequences of socio-environmental factors such as lockdowns, social distancing measures, and other social activities like protests, elections, rallies. Our system would adapt to the current trends of the virus and predict accordingly. This system will base its predictions on a more diverse and elaborate dataset. It will examine strategies and make recommendations for social measures to prevent the virus from spreading. We hope to build a model that will predict infection rates for different age/gender/occupation categories along with other factors that will help governments assess the risk populations and enforce measures to protect the inhabitants at risk.

We plan to include features like lockdown measures, social distancing measures, and public view on the epidemic while training our model. The model will use the current advanced methods and practices that would enable accurate prediction rates and suggest interventions to control the spread of the virus.

In conclusion, our model should be able to predict how socio-environmental factors affect the virus spread and intimate government officials and health professionals to take necessary precautions and advise the public against risky actions.



B. Merits

Our proposed system would accurately predict how the virus would spread and layout interventions needed to control it. It would consider features such as lockdown measures, social activities, and other aspects of society to understand why and how the virus will affect the population. It will also predict how the virus will affect different denominations of the population by age, gender, race, and occupation category to determine the best strategy to protect people at risk.

Implementation

A. Data Collection and Data Pre-processing

[17] This module collects all the necessary data and pre-processes them for feature extraction. The model collects data about the daily infected, recovered, and deceased as it would help predict infection, recovery, and mortality rates. It also gathers information like lockdown status, social distancing measures, and other activities that might affect the prediction. Aside from pre-processing data and extracting features, the model also produces a new dataset that can be used to predict infection, recovery, and mortality rates for various population groups (age, gender, race, pre-existing conditions).

B. Exploratory Data Analysis

This module makes inferences from the data collected so that it would select the right epidemic model. We hope that the model will be able to determine epidemic models like SIR, SEIR, SIS, SIRD, MSEIR, and other compartmental epidemic models. [18]The model also analyzes newspaper articles and reports like situation updates- to understand the influence it has on infection rates and assess the spread of the pathogen to determine interventions needed to control the virus.

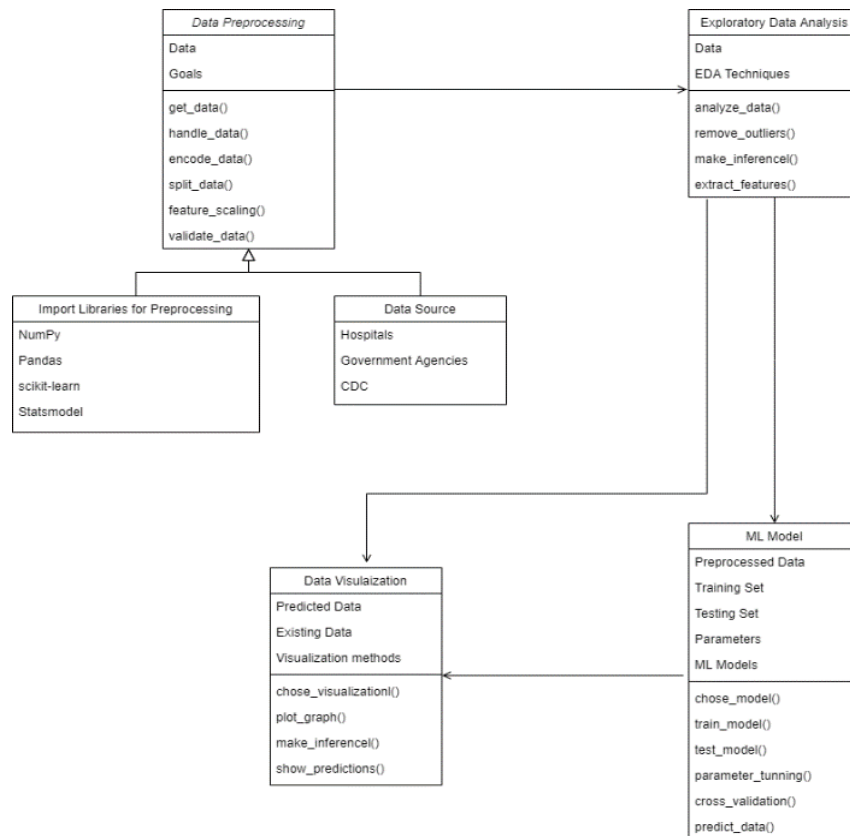
C. Prediction Model

This module is the cornerstone of predictions and decision-making. We plan to use Recurrent Neural Networks (RNN) to design proposed models. [21] The long short-term memory (LSTM) variants of RNN architecture help us analyze time series and predict infection and fatality rates accordingly. We also plan to compare Multivariate LSTM Models and Multi-Step LSTM Models with Univariate LSTM Models to study and predict the number of people who will be infected and which category of the population will be at most risk. We also compare different networks to analyze which is the best model to use for the prediction of infection rates.

D. Data Visualization

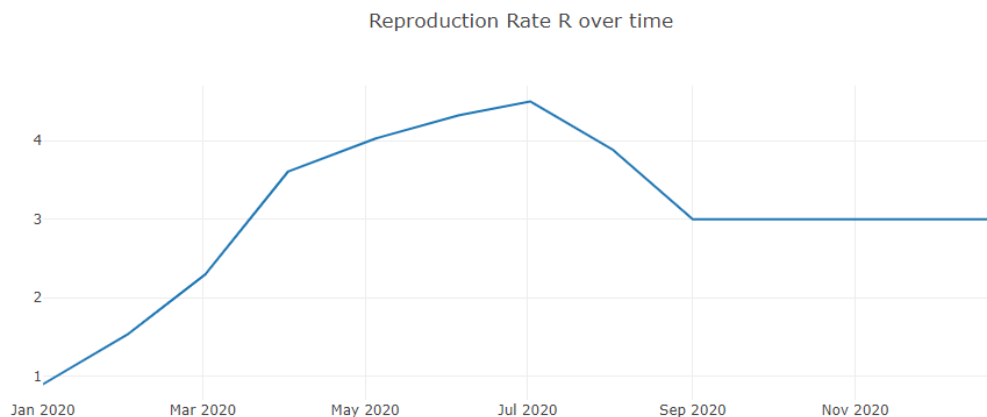
We've all seen models forecasting the spread of the novel coronavirus—most notably "flatten the curve" graphics that's been widely publicized. This module helps us understand and make human inferences easier as it would be helpful for the general public to understand the situation better and work in unity towards eradicating pathogen

dissemination. Visualizing through heatmaps, bar-graphs, and distributions by categories like susceptible, exposed, recovered, infected, critical, along with age, occupation distributions, etc. are an integral part of this module.

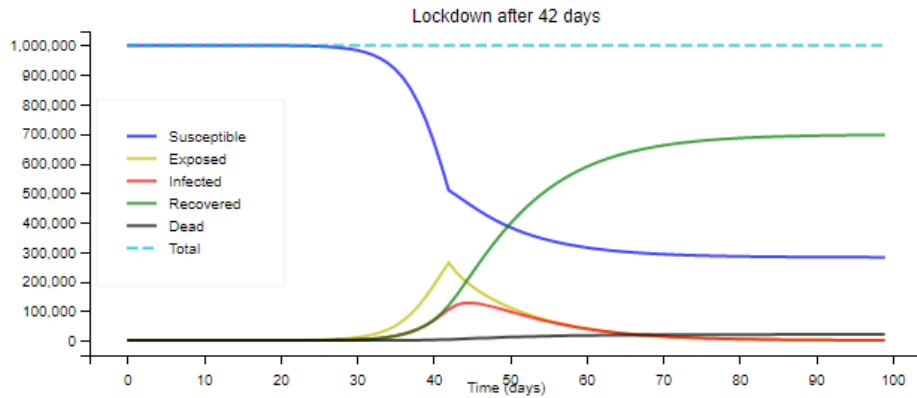


Results Discussion

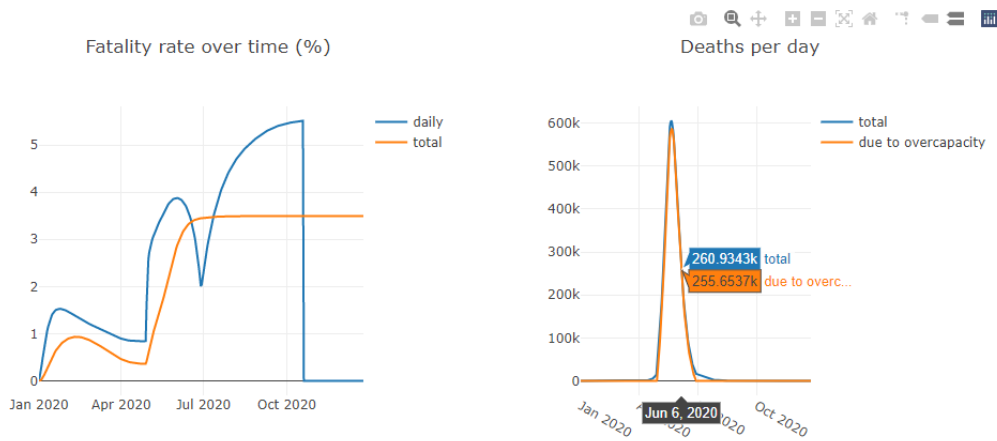
We simulated infectious outbreaks in several population samples. Our system was able to forecast future values and respond to changes in reproduction rates. This demonstrates that by using machine-learning techniques, epidemic models can be made more efficient and be given independent discernment.



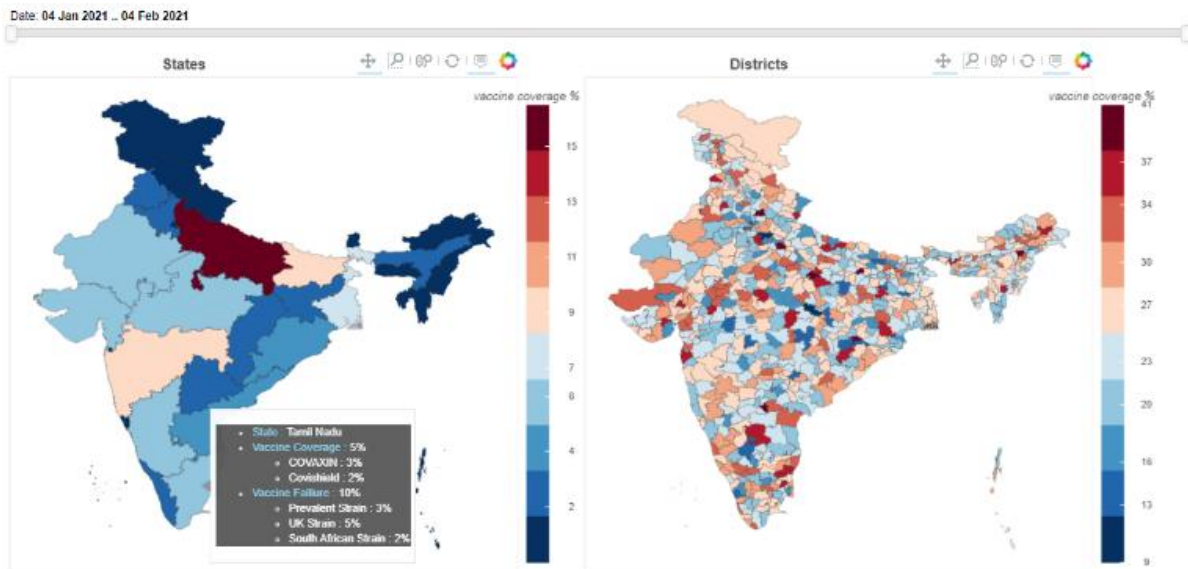
We were also able to model how socio-environmental factors affect pathogen dissemination.



We also investigated how fatality rates varied by age and resource distribution.



Finally, we explained how population compartments change over time using data visualizations. We also visualized graphs depicting vaccine distribution figures.



Conclusion

This paper looks at current methodologies for fighting epidemics and attempts to propose a machine-learning approach by breaking up existing models of epidemics that can respond to a specific outcome and adjust to new data and changing variables. We conclude by highlighting how we can optimize and adjust this approach so that governments and organizations can prepare for the next outbreak and warn people against high-risk activities.

References

- [1] Fokas, A.S., Dikaios, N., & Kastis, G.A. (2020). Mathematical models and deep learning for predicting the number of individuals reported to be infected with SARS-CoV-2. *Journal of the Royal Society Interface*, 17(169), 20200494. <http://dx.doi.org/10.1098/rsif.2020.0494>
- [2] Hamzah, F.B., Lau, C., Nazri, H., Ligot, D.V., Lee, G., Tan, C.L., & Chung, M. H. (2020). CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. *BullWorld Health Organ*, 1(32).<https://doi.org/10.2471/BLT.20.255695>
- [3] Chen, B., Shi, M., Ni, X., Ruan, L., Jiang, H., Yao, H., & Ge, T. (2020). Data visualization analysis and simulation prediction for covid-19. *arXiv preprint arXiv:2002.07096*
- [4] Godio, A., Pace, F., & Vergnano, A. (2020). SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence. *International Journal of Environmental Research and Public Health*, 17(10), 3535.<https://doi.org/10.20944/preprints202004.0073.v1>
- [5] Polonsky, J.A., Baidjoe, A., Kamvar, Z.N., Cori, A., Durski, K., Edmunds, W.J., & Jombart, T. (2019). Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Philosophical Transactions of the Royal Society B*, 374(1776), 20180276. <https://doi.org/10.1098/rstb.2018.0276>
- [6] Lega, J., & Brown, H.E. (2016). Data-driven outbreak forecasting with a simple nonlinear growth model. *Epidemics*, 17, 19-26.<https://doi.org/10.1016/j.epidem.2016.10.002>
- [7] Hou, C., Chen, J., Zhou, Y., Hua, L., Yuan, J., He, S., & Jia, E. (2020). The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis. *Journal of medical virology*, 92(7), 841-848.<https://doi.org/10.1002/jmv.25827>
- [8] Li, Y., Jia, W., Wang, J., Guo, J., Liu, Q., Li, X., & Wang, F. (2021). ALeRT-COVID: Attentive lockdown-aware transfer learning for predicting COVID-19 pandemics in different countries.<https://doi.org/10.1101/2020.07.09.20149831>
- [9] Stutt, R.O., Retkute, R., Bradley, M., Gilligan, C.A., & Colvin, J. (2020). A modelling framework to assess the likely effectiveness of facemasks in combination with 'lock-down' in managing the COVID-19 pandemic. *Proceedings of the Royal Society A*, 476(2238), 20200376.<http://doi.org/10.1098/rspa.2020.0376>
- [10] Chu, C., & Lee, S. (2015). Assessment of intensive vaccination and antiviral treatment in 2009 influenza pandemic in Korea. *Osong public health and research perspectives*, 6(1), 47-51.<http://doi.org/10.1016/j.phrp.2014.11.007>
- [11] Guzman, M.G., Gubler, D.J., Izquierdo, A., Martinez, E., & Halstead, S.B. (2016). Dengue infection. *Nature reviews Disease primers*, 2(1). <https://doi.org/10.1038/nrdp.2016.55>
- [12] Dixit, N.M., & Perelson, A.S. (2004). Complex patterns of viral load decay under antiretroviral therapy: influence of pharmacokinetics and intracellular delay. *Journal of Theoretical Biology*, 226(1), 95-109. <https://doi.org/10.1016/j.jtbi.2003.09.002>
- [13] ElBagoury, M., Tolba, M.M., Nasser, H.A., Jabbar, A., & Hutchinson, A. (2021). The find of COVID-19 vaccine: Challenges and opportunities. *Journal of Infection and Public Health*, 14(3), 389-416. <https://doi.org/10.1016/j.jiph.2020.12.025>
- [14] Yu, H., Li, C., Wang, X., Duan, J., Yang, N., Xie, L., & Li, Y. (2020). Techniques and Strategies for Potential Protein Target Discovery and Active Pharmaceutical Molecule Screening in a Pandemic. *Journal*

of Proteome Research, 19(11), 4242-4258.<https://doi.org/10.1021/acs.jproteome.0c00372>

- [15] Wangping, J., Ke, H., Yang, S., Wenzhe, C., Shengshu, W., Shanshan, Y., & Yao, H. (2020). Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *Frontiers in medicine*, <https://doi.org/10.3389/fmed.2020.00169>
- [16] Verma, V., Vishwakarma, R.K., Verma, A., Nath, D.C., & Khan, H.T. (2020). Time-to-Death approach in revealing Chronicity and Severity of COVID-19 across the World. *PloS one*, 15(5), e0233074.<https://doi.org/10.1371/journal.pone.0233074>
- [17] Taghrir, M.H., Akbarialiabad, H., & Marzaleh, M.A. (2020). Efficacy of mass quarantine as leverage of health system governance during COVID-19 outbreak: a mini policy review. *Archives of Iranian medicine*, 23(4), 265-267.<https://doi.org/10.34172/aim.2020.08>
- [18] Cheynet, E. (2020). Generalized SEIR Epidemic Model (Fitting and Computation). Zenodo. <https://doi.org/10.5281/ZENODO.3911854>
- [19] Khailaie, S., Mitra, T., Bandyopadhyay, A., Schips, M., Mascheroni, P., Vanella, P., & Meyer-Hermann, M. (2021). Development of the reproduction number from coronavirus SARS-CoV-2 case data in Germany and implications for political measures. *BMC medicine*, 19(1), <https://doi.org/10.1186/s12916-020-01884-4>
- [20] Smith, D., & Moore, L. (2004). The SIR model for spread of disease: the differential equation model. *Convergence*.
- [21] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.<https://doi.org/10.1162/neco.1997.9.8.1735>