

Deep learning application for real-time prediction of COVID-19 outbreak with susceptible-infected-recovered-deceased model

Hoang-Sy Nguyen¹, Thu-Ngan T. Phan², Cong-Danh Huynh³

¹Becamex Business School, Eastern International University, Binh Duong Province, Vietnam

²Faculty of Medical Laboratory, Hong Bang International University, Ho Chi Minh City, Vietnam

³Faculty of Economics, Thu Dau Mot University, Thu Dau Mot City, Vietnam

Article Info

Article history:

Received Jun 8, 2022

Revised Jun 28, 2022

Accepted Aug 2, 2022

Keywords:

COVID-19

Data visual

LSTM network

Real-time forecasting

SIRD model

ABSTRACT

Due to the complex nature of a pandemic such as COVID-19, forecasting how it would behave is difficult, but it is indeed of utmost necessity. Furthermore, adapting predictive models to different data sets obtained from different countries and areas is necessary, as it can provide a wider view of the global pandemic situation and more information on how models can be improved. Therefore, we combine here the long-short-term memory (LSTM) model and the traditional susceptible-infected-recovered-deceased (SIRD) model for the COVID-19 prediction task in Ho Chi Minh City, Vietnam. In particular, LSTM shows its strength in processing and making accurate numerical predictions on a large set of historical input. Following the SIRD model, the whole population is divided into 4 states (S), (I), (R), and (D), and the changes from one state to another are governed by a parameter set. By assessing the numerical output and the corresponding parameter set, we could reveal more insights about the root causes of the changes. The predictive model updates every 10 days to produce an output that is closest to reality. In general, such a combination delivers transparent, accurate, and up-to-date predictions for human experts, which is important for research on COVID-19.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Cong-Danh Huynh

Faculty of Economics, Thu Dau Mot University

06 Trần Văn Ôn, Phú Hoà, Thủ Dầu Một, Bình Dương, Vietnam

Email: danhhc@tdmu.edu.vn

1. INTRODUCTION

Since the first outbreak in Wuhan, Hubei, China (Nov. 2019), SARS-CoV-2, or COVID-19, has rapidly spread over the world with the total infected cases and total deaths of, respectively, about 490 and 6.15 million (Mar. 2022) [1]. Indeed, the COVID-19 disease was so contagious that it was declared "pandemic" (Mar. 2020) [2] shortly after WHO officially announce it the "Public Health Emergency of International Concern" (Jan. 2020) [3]. As Vietnam shares the border with China, COVID-19 rapidly became the most important public concern for Vietnamese society, government, and academia in early 2020 [4]. Studies in this early stage serve mainly as reports of infected cases [5], the reaction of the society reflected in the risk assessment of the pandemic [6]. Despite the early successes in containing the virus and maintaining economic growth [7], [8] with strict response, Vietnam was and has been severely affected by the disruption of global supply chains, causing a crisis and stagnation in hospitality, the national economy (including tourism). Among the most popular models used for predicting infectious disease outbreaks is the susceptible-infectious-recovered-deceased (SIRD) model [9]. As its name suggests, the model includes four states: susceptible (S) - people have the infection risk, infected (I) - people are infected, recovered (R) - people

recover from the disease, and deceased (D) - people pass away because of the disease. It should be noted that in a conventional SIRD, infected cases are assumed to gain immunity after recovery to decrease the computational complication when considering re-infected cases [10], [11]. Aside from the SIRD model, the COVID-19 pandemic has been studied by means of logistic model [12], confinement and quarantine model [13], Bayesian and stochastic techniques [14], [15]. However, predicting pandemics with mathematical models, which have been mainly developed from [16], could be ambiguous because it is hard to avoid bias in selection of the model parameters.

In recent years, self-learning algorithms categorized as deep learning (DL), a subbranch of machine learning (ML), have been developed for various tasks such as natural language processing, face and object detection, speech recognition. [17]. DL shows its strength in processing big data with relatively high accuracy in comparison to other traditional algorithms and can be fine-tuned to obtain even better results for particular applications. For instance, DL has been applied in various fields such as in the hospitality industry [18], legal field [19], stock market prediction [20], and traffic anomaly detection [21]. Furthermore, DL is a promising approach for processing the daily records of COVID-19 infected cases that increase massively amidst the pandemic [22], [23]. The infected cases over the course of time are fed into the DL models such as recurrent neural network (RNN) [24] and long short term memory (LSTM) [25], which are specialized for processing sequential history data. As a result, predictions based on historical inputs from these DL models are with remarkably high accuracy. However, the well-packed layer structures of the DL models often appear as black boxes, which provide no insight into how the predictions were obtained. Therefore, the analyzers face difficulties in interpreting the results to obtain the root causes and make appropriate counteractions. In view of this, interpretable machine learning has been developed aiming to offer the analyzers explanations for the results while maintaining computational power [26].

To mitigate the impact of COVID-19, prediction of infected cases, recoveries, and deaths is important, so that policy makers can prepare preventive measures for the worst scenarios in advance. This paper aims to fulfill this forecasting task utilizing the LSTM DL model in combination with the SIRD model. Over the course of the problem formulation, readers would understand how the two models are adopted and combined to yield accurate predictions with helpful insights, which can facilitate the making of public health policies.

2. METHOD

This study serves as a forecast for the COVID-19 situation in Ho Chi Minh City, Vietnam. First, sequential historical data are input into LSTM so that the main features of the data can be extracted. Second, the predicted outputs from LSTM are considered as optimal parameters for SIRD model, which is then used to predict the exact COVID-19 cases. The objections that we have to overcome are: processing massive amount of historical data, complicated computational model, long computational time, and making the predicted results interpretable while ensuring that they best reflect the real pandemic situation. These goals can be achieved using the approach illustrated in Figure 1.

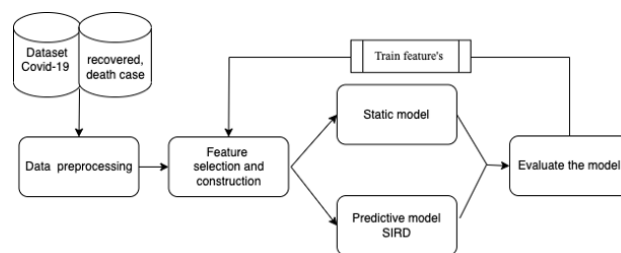


Figure 1. Proposed approach

2.1. Predictive modelling-SIRD model

The SIRD model was used to model the COVID-19 spread [26], [27]. The model starts with (S), where people are healthy. A population of (S) moves to (I) when they are infected with COVID-19. From (I), they can either move to (R) (people are cured), or to (D) (people pass away, deceased or death). As

mentioned above, the transition from (I) to (R) is irreversible because it was assumed that infected people would gain immunity to the disease after recovering from it. Besides, we assume three parameters for the SIRD model are (β, γ, μ) , which are respectively the infected rate per time unit, recovery rate per time unit, and death rate per time unit. The total population is denoted as $N = S + R + I + D$. Then, the SIRD model would include the first-order derivative, making up a system of four differential equations at the initial time (t) that epidemic was recognized, as (1.1) to (1.4).

$$\frac{dS}{dt} = -\beta \frac{S(t)I(t)}{N}, \tag{1.1}$$

$$\frac{dI}{dt} = \beta \frac{S(t)I(t)}{N} - \gamma I(t) - \mu I(t), \tag{1.2}$$

$$\frac{dR}{dt} = \gamma I(t), \tag{1.3}$$

$$\frac{dD}{dt} = \mu I(t), \tag{1.4}$$

At a particular time (t) and specified (β, γ, μ) , we are able to obtain a unique set of $S(t), I(t), R(t), D(t)$, which are the number of people respectively in each state [27]. Indeed, estimation of (β, γ, μ) is complicated and sometimes biased [28], but it is essential for the accuracy of the pandemic prediction with SIRD model. The three parameters (β, γ, μ) can be extracted from the historical data and by assessing the key parameters $\beta, \gamma, and \mu$, we can understand some insights of the pandemic. It should be noted that the time range is not given because each pandemic differs from one another.

$$S' = -\beta S I. \tag{2}$$

Subsequently, the variation of recovered cases per time unit and infected cases per time unit can be derived from (1.2) and (1.3) as (3) and (4).

$$R' = \gamma \times I. \tag{3}$$

$$I' = \beta \times SI - \gamma - \mu. \tag{4}$$

Eventually, we can obtain the variation of deceased cases per time unit from (1.4) as (5).

$$D' = \mu \times I. \tag{5}$$

Assessing these variations together with the key parameters can help us to get the insights of the pandemic situation. For example, if $D(t)$ has the tendency to increase while μ is high and β is small, it can be drawn that new deceased cases does not stem from the newly infected but from the serious health issues of the infected cases. On the other hand, it means that the contagion of COVID-19 has been under some control. Additionally, there may be some problems with the healthcare system, which is the reason for the increasing number of deaths. This would encourage governments to propose appropriate and up-to-date policies, such as lifting COVID restrictions and reviewing healthcare policies. By interpreting the obtained results in this manner, the SIRD model has become no longer inexplicable.

Remarkably, obtaining an accurate set of parameters (β, γ, μ) is inessential because it varies for cases, i.e., regions with different geography, economy, social policies, and capacity of healthcare systems. Because the parameters are variations of time, they can best describe the real situation only if the estimation is limited to a particular region in a short and particular time. Therefore, processing a massive amount of COVID-19 data from all around the world is nontrivial and indeed impossible because one cannot predict all potential changes that would happen to the COVID situation in the world. At the start of the pandemic, when the public is not aware of the disease, β describes the natural infection rate [28] in society. During the course of time, β would change as people perform hygiene practices and social distancing along the health regulations enforced by the government. On the other hand, γ reflects the number of recoveries within a time range, the inverse of which is the days an individual is infected, γ^{-1} . Thus, we can estimate the γ with the average recovery time of the individual from the medical records. Together with μ , which shows the rate of deceased cases within a time range, γ indicates how well the healthcare system of a region under consideration performs [29]. As explained, to describe the dynamic of an unknown disease, which has the potential to become a pandemic, estimation of parameters (β, γ, μ) is the utmost important task, which are

later presented in detail. According to [30], healthcare restriction is the most effective way to control the widespread of the pandemic and decrease the infected as well as the decreased cases.

To consider this dependence of the parameters, we employ probability, p_0 as studied in [31] to represent the number of infected cases caused by an infected case (reproduction number) as (6):

$$p_0 = \begin{cases} \frac{\beta}{\gamma+\mu} S(t), & (i), \text{ if } p_0 > 1 \\ (ii), & \text{ otherwise} \end{cases} \quad (6)$$

in which, (i) indicates that there is a reduction in the newly infected cases and a tendency of the pandemic to be moderated. (ii) indicates the exponential and uncontrollable growth of the pandemic. The formula emphasizes the importance of social distancing, since for a given β in a particular time of pandemic, p_0 can only be decreased if we reduce the number of people that are potentially infected. The interrelation between the key parameters inspired us to employ the DL model for a more accurate estimation of the three parameters (β, γ, μ) from the historical data. In addition, utilizing the DL approach in this manner also makes the predictive DL model more transparent for medical experts.

Finally, we proposed a hybrid model architecture by combining it with ideas of employing interpretative machine learning techniques to automatically find the best approximated parameters as presented in Figure 2. The descriptions are discussed in detail as follows. Step 1: extracting the key features from the historical data. Given a historical data and population N as the input for the LSTM layers, we first produce a collection of feature vectors $\Delta \bar{S}_t, \Delta \bar{I}_t, \Delta \bar{R}_t, \Delta \bar{D}_t$, respectively, which describe the concentration of COVID-19 on a set of local contaminated surfaces as (7) to (10).

$$\Delta \bar{S}_t = \frac{S(k+1)}{N} - \frac{S(k)}{N}, \quad (7)$$

$$\Delta \bar{I}_t = \frac{I(k+1)}{N} - \frac{I(k)}{N}, \quad (8)$$

$$\Delta \bar{R}_t = \frac{R(k+1)}{N} - \frac{R(k)}{N}, \quad (9)$$

$$\Delta \bar{D}_t = \frac{D(k+1)}{N} - \frac{D(k)}{N}. \quad (10)$$

It should be noted that indirect infection is considered, but only at a random time every day for each individual case, because the indirect infection rate is relatively small.

Step 2: encoding and decoding for SIRD training. The output of LSTM is flattened to a 1D vector so that it could be encoded with the variational autoencoder (VAE) [32]. One of the most typical encoding technique namely Multilayer Perceptron is employed [33]. The decoder is designed to generate the output, which is a matrix size of $n \times 4$, the same as the input. Since the matrix in the output records the data from 2^{nd} to $(n+1)^{th}$ day from the set of historical data, it enables automatic labeling for the encoding-decoding process. Euler method [34] is utilized to decode the matrix output according to the parameters (β, γ, μ) yielded from the previous learning process. Specifically, based on the k^{th} day, the decoder estimates the number of cases in (S), (I), (R), (D) states at $(k+1)^{th}$ day as $S(k+1), I(k+1), R(k+1), D(k+1)$, respectively:

$$\frac{S(k+1)}{N} = \frac{S(k)}{N} - \beta \frac{S(k) I(k)}{N}. \quad (11)$$

$$\frac{I(k+1)}{N} = \frac{I(k)}{N} + \beta \frac{S(k) I(k)}{N} - \gamma \frac{I(k)}{N} - \mu \frac{I(k)}{N}. \quad (12)$$

$$\frac{R(k+1)}{N} = \frac{R(k)}{N} + \gamma \frac{I(k)}{N}. \quad (13)$$

$$\frac{D(k+1)}{N} = \frac{D(k)}{N} + \mu \frac{I(k)}{N}. \quad (14)$$

Additionally, during training, after the data is fed through the SIRD's layers, we employ mean squared error (MSE) for loss function to determine the true error rate for our prediction [35]. Besides, for such a historical data set with lots of noises, Adam optimizer is appropriate for weight updating task [36]. Subsequently, we can calculate the MSE of a given epoch for loss function as (15):

$$MSE = \frac{1}{n} \sum_{k=1}^n (W_k - \widehat{W}_k)^2, \tag{15}$$

where W_k, \widehat{W}_k are k -th vector respectively from the historical data and the prediction, considering the output matrix sizing of $n \times 4$ discussed previously. Remark: by means of loss function, it is possible to update the system's weights following the back propagation method. It is important to mention that the decoder is not updated throughout the training process because it is not assigned with trainable weights. The training process works so that when the results of the loss function are stable, the prediction of (β, γ, μ) from any historical data is announced successful.

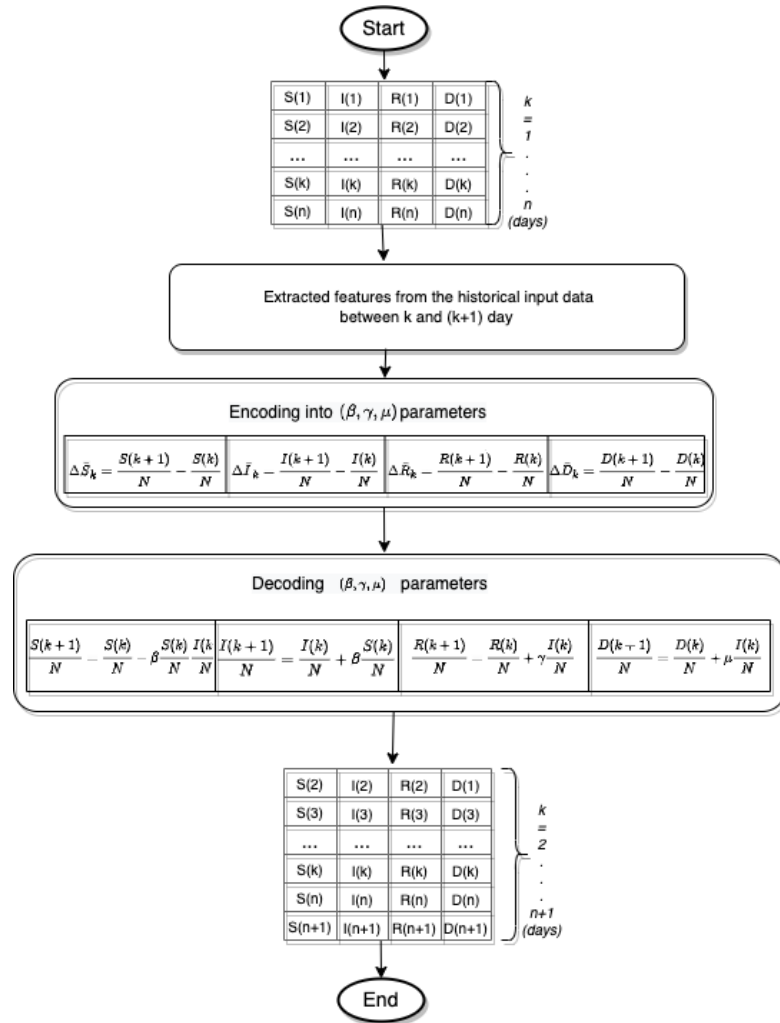


Figure 2. Proposed a combine SIRD-LSTM model

2.2. Data source and visualization

We investigate in this study two sets of data, i.e., from COVID-19 updates of the Vietnamese Ministry of Health (MOH) [37] and from coronavirus resource center of John Hopkins University [38] (with data set in [39]). Figure 3 illustrates the total infectious (T_I) case in Ho Chi Minh city (including the infected, recovered, and deceased cases) from Jul. 2021 to Feb. 2022. Python programming language was used to collect and process the data. The historical data input is updated once every 10 consecutive days. The predictive model requires inputs that are classified into (S)-(I)-(R)-(D) states. However, from the available sources, only (R), (D), and (S_I : including (I), (R), and (D)) can be extracted. Thus, we need to consider the whole population in Ho Chi Minh city to calculate the other inputs. Then, the inputs are normalized by dividing all the classified data by the population in Ho Chi Minh City, to obtain $C_S = N - S_I, C_I = S_I - R - D, C_R = R, C_D = D$, which are the percentages of S-I-R-D cases, respectively. The total population is N and other parameters were aforementioned.

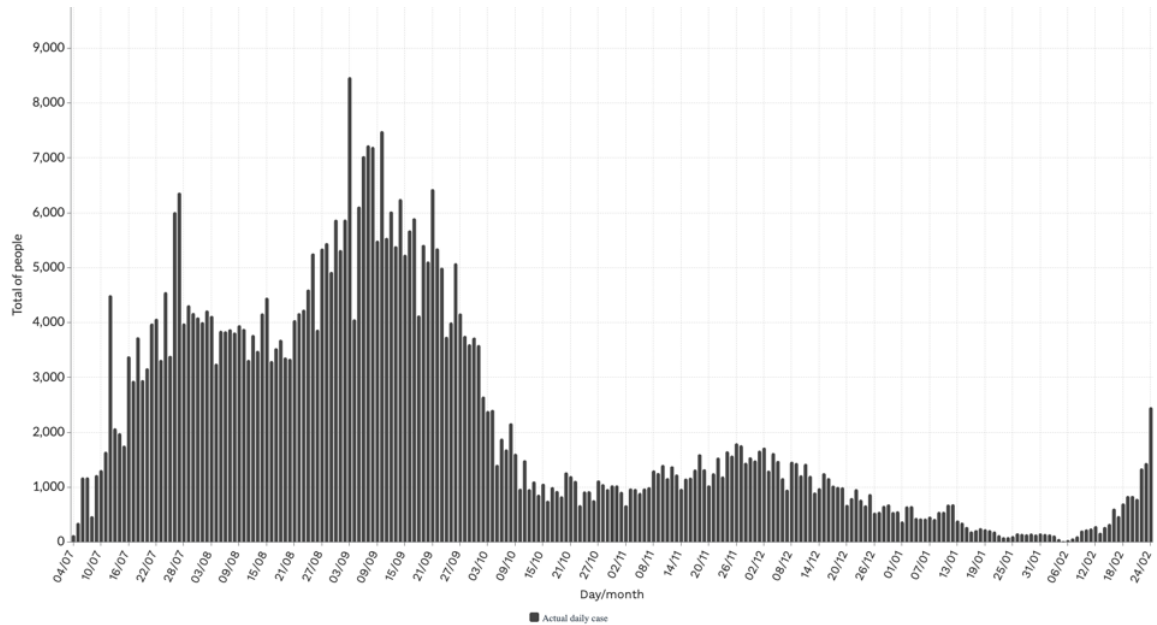


Figure 3. Total infectious cases in Ho Chi Minh City (Jul. 2021 to Feb. 2022)

3. RESULTS AND DISCUSSION

The model works so that data in the previous days is combined with parameters (β, γ, μ) to predict the S, I, R, D cases in the upcoming days. Notably, the key features selected from the input data are important for the estimation of parameters (β, γ, μ) , thus, must be chosen carefully. The goal of the proposed hybrid predictive model is to accurately approximate the parameters (β, γ, μ) based on which future cases can be calculated (with lowest error in comparison with reality) and their characteristics can be inferred. Another factor that would affect the prediction is day lag n , which is in the range of 5 to 31 in [40]. We utilize recursive strategy [41] for k -step forecasting to determine the most optimal n lag days for the proposed model. As for the accuracy assessment, the error R^2 between the predicted and the real cases can be calculated as $R^2 = 1 - \frac{\varepsilon_1}{\varepsilon_2}$, where $\varepsilon_1, \varepsilon_2$ denotes the residual sum of squares, and total sum of squares, respectively. Ideally, $R^2 < 1$, and the smaller the error is, the more explainable the SIRD model is. Subsequently, we can calculate the mean of the total percentage errors for k -step forecasting metric as $MAPE = \frac{1}{n} \sum_{k=1}^n |W_k - \widehat{W}_k| / W_k$, where W_k, \widehat{W}_k denote for the actual cases and predicted cases. The better result is with the smaller $MAPE < 0$.

The historical data obtained from the time range between Jul. 2021 and Feb. 2022 was used for training the DL model. The prediction is realized in a way that 28 days in the past are used to forecast the next 28 days. The below figures illustrate the actual historical data (from late Jan. 2022 to late Feb. 2022) and predicted daily cases (from late Feb. 2022 to late Mar. 2022). The predictions were realized for three scenarios (normal, best and worst) that would happen with the COVID situation in Ho Chi Minh City, Vietnam. The trends over the predicted days reflect different insights, with regard to different parameter values (β, γ, μ) . In the case of prediction made by black-box models, for high number of deceased cases, the most common conclusion would be that the current health system could not handle the rising number of infectious cases. Regarding the same prediction made with SIRD model, if we assess the parameters (β, γ, μ) and would see that the high deceased number is accompanied by the increase of recovery rate and the decrease of transmission as well as the deceased rate, we could conclude that the pandemic situation has been under control in the area under study. Other reasoning can be made when we assess the raise or fall of cases in combination with the parameters (β, γ, μ) .

From Figure 4 to Figure 7, the number of predicted cases for the three different scenarios are presented. All the predicted scenarios share one thing in common, i.e., the number of cases stabilize over time. In Figure 4 from the recent historical data set, the new daily cases are predicted to decrease over time for all three scenarios. This is a positive sign which indicates that the healthcare measures taken so far by the government in Ho Chi Minh City is effective in the fight against the pandemic. Accompanying this, in Figure 5, how well the healthcare system has been navigated is also shown by the increase in daily recovered cases in the next days for all three scenarios. In Figure 6, the number of deceased cases in best and normal

scenarios are predicted to be relatively small. This is reasonable because at this stage of the pandemic situation, a majority of the population has gain immunity either by vaccination or recovering from previous infection. Moreover, it could be as well because the new COVID-19 variant has been less severe. As can be observed in Figure 7, in the worst scenario, the deceased cases grow exponentially to a high number. Even though this may appear to be unrealistic, it still can warn the government about how severe the situation could be so that appropriate plans can be prepared in advance to contain the pandemic in the worst case.

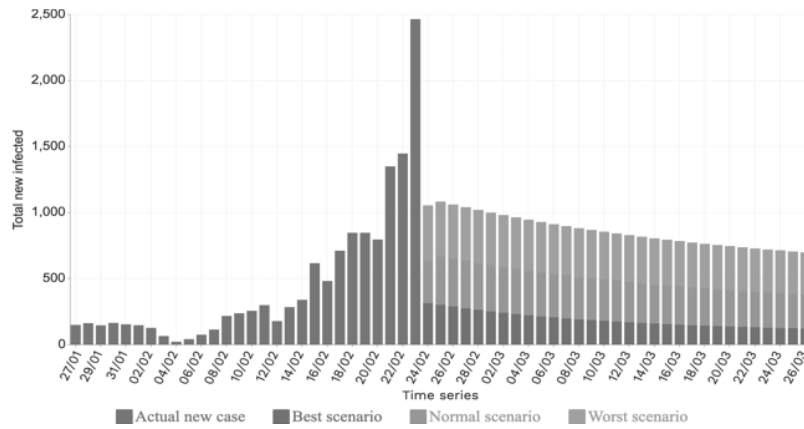


Figure 4. Forecast of new daily infected cases

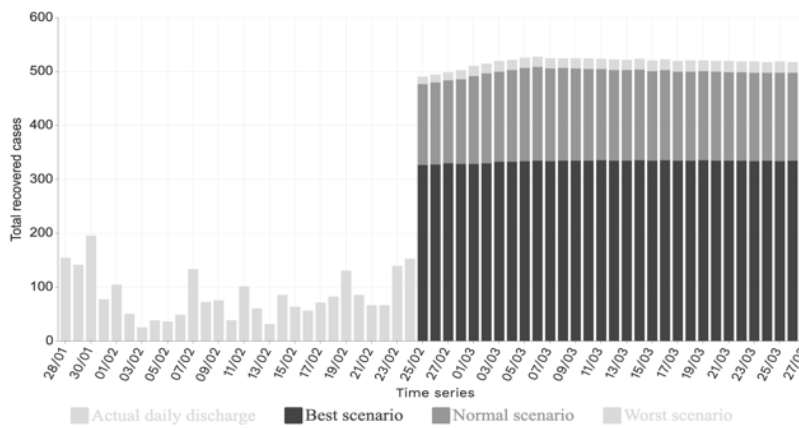


Figure 5. Forecast of new daily recovered cases

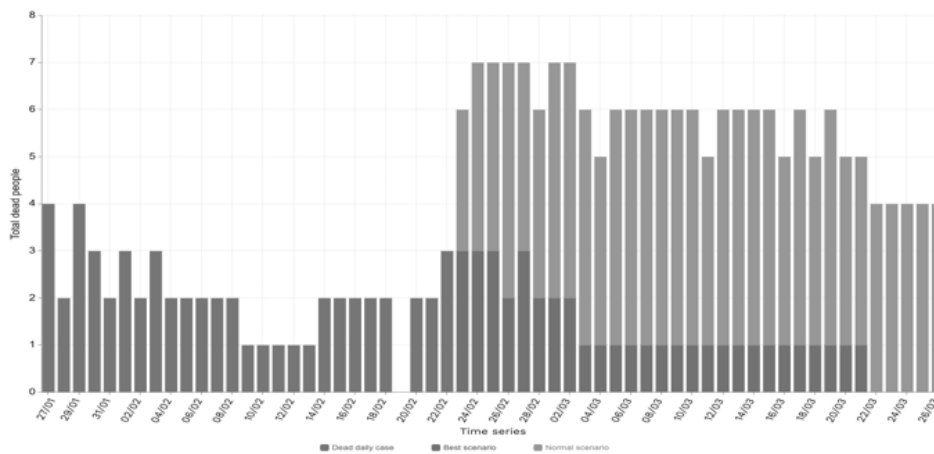


Figure 6. Forecast of new daily deceased causes (normal case and best case)

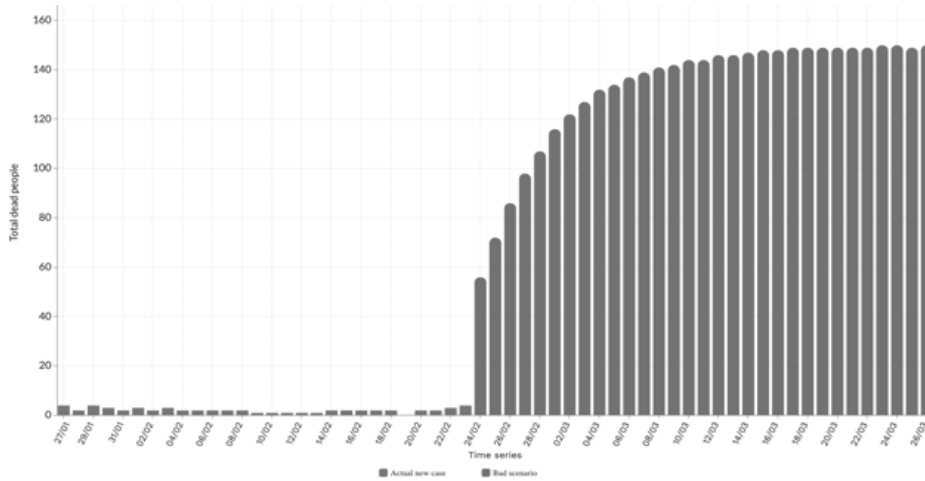


Figure 7. Forecast of new daily deceased causes (worst case)

In general, considering the presented results and the insights they provide, we can proactively cope with the pandemic changes. For a fixed training data set and stable pandemic situation, the SIRD model can forecast well the situation. Otherwise, in the case of sudden changes in the pandemic, more time needs to be taken for the model to predict accurately. To overcome this, the SIRD model is often updated with the help of reinforcement learning to stay up-to-date with reality. We recommend to update the historical data input on more or less weekly basis for the best prediction. Impact of COVID-19 on the economics and politics have been surprisingly severe and long lasting. According to report from World Health Organization (WHO), from Dec. 2019 up to Mar. 2022, there have been approximately 494.5 million infected cases and over 6.17 million deaths were recorded all over the world. To reduce the life loss and mitigate the impacts of COVID-19 on many aspects of a country or region, policy makers must be provided with the most accurately predicted scenarios to propose appropriate moves, which is why predictive models for pandemic situations are important. Regarding the pandemic prediction, it is essential to take into consideration the model's ability to rapidly adapt to pandemic changes and reflect them to the predicted results. Therefore, it is a promising approach to utilize the SIRD model for pandemic description as well as DL approach to process the historical data for the key parameters (β, γ, μ) . This helps the predictive model to perform well and maintain itself for long-term predicting task.

We compare the accuracy of the results obtained from the proposed combined SIRD-LSTM model and from other popular forecasting models (both statistical and machine learning models) in Table 1, i.e., least absolute selection Shrinkage operator (LASSO) [42], support vector machine for regression (SVR) [43], decision tree regression (DTR) [44]. We use the same n-day lag for all the models. It can be observed that similar to other popular predictive methods, SIRD model can predict with high accuracy only when it is fed with a sufficient amount of input data. The predicted results are of relatively high accuracy in comparison with other methods.

Table 1. Comparison of the forecasting model

Rate Method	Total infectious		Recovered		Deceased	
	R^2	MAPE	R^2	MAPE	R^2	MAPE
SIRD-LSTM	0.9979	0.0051	0.9878	0.0187	0.9890	0.0068
LASSO [42]	0.9991	0.0011	0.9666	0.0115	0.5365	0.0402
SVR [43]	-18.1162	0.6208	-13.4640	0.0187	-46.3249	0.5586
DTR [44]	-2.8516	0.2324	-3.0777	0.2922	-2.8120	0.1325

4. CONCLUSION

In conclusion, due to the complex nature of the pandemic, it is not sufficient for a predictive model to provide only the closest prediction to reality. Indeed, the prediction must be interpretable so that human experts can base on that to make appropriate decisions. In view of this, we investigated the combination SIRD-LSTM of SIRD model to describe the pandemic and DL approach to identify key parameters (β, γ, μ) to best fit the SIRD model to reality. Implementation of other semi-supervised learning models could make the training process become more efficient. Besides reviewing the COVID-19 situation in Vietnam and

conducting a predictive case study in Ho Chi Minh City, we also explain how to build the predictive model and interpret the results. This study can be applied for other regions or prediction of other diseases as long as the historical data is sufficiently available, thanks to the adaptability that DL offers. This paper would serve as a starting point for further in-depth research into the many questions related to this pandemic.




REFERENCES

- [1] "COVID-19 Map - Johns Hopkins Coronavirus Resource Center," Johns Hopkins Coronavirus Resource Center, 2022. [Online]. Available: <https://coronavirus.jhu.edu/map.html> (Accessed: 02- Apr- 2022).
- [2] "Coronavirus disease (COVID-19) – World Health Organization," Who.int, 2022. [Online]. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (Accessed: 02- Apr- 2022)
- [3] "WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020," Who.int, 2022. [Online]. Available: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (Accessed: 02- Apr- 2022).
- [4] T. H. Nguyen and D. C. Vu, "Summary of the covid-19 outbreak in vietnam – lessons and suggestions," *Travel Medicine and Infectious Disease*, vol. 37, p. 101651, 2020, doi: 10.1016/j.tmaid.2020.101651.
- [5] L. T. Phan et al., "Importation and human-to-human transmission of a novel coronavirus in vietnam," *New England Journal of Medicine*, vol. 382, no. 9, pp. 872–874, 2020, doi: 10.1056/NEJMc2001272.
- [6] T. Huynh, "The covid-19 risk perception: A survey on socioeconomics and media attention," *Economics Bulletin*, vol. 40, no. 1, pp. 758–764, 2020.
- [7] D. V. Huynh, T. T. K. Truong, L. H. Duong, N. T. Nguyen, G. V. H. Dao, and C. N. Dao, "The covid-19 pandemic and its impacts on tourism business in a developing city: insight from Vietnam," *Economies*, vol. 9, no. 4, p. 172, 2021. doi: 10.3390/economies9040172.
- [8] E. B. Postnikov, "Estimation of covid-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions?," *Chaos, Solitons & Fractals*, vol. 135, p. 109841, 2020, doi: 10.1016/j.chaos.2020.109841.
- [9] J. Fernandez-Villaverde and C. I. Jones, "Estimating and simulating a SIRD model of covid-19 for many countries, states, and cities," *National Bureau of Economic Research, Working Paper 27128*, 2020, doi: 10.3386/w27128.
- [10] O. Iyiola, B. Oduro, T. Zabilowicz, B. Iyiola, and D. Kenes, "System of time fractional models for covid-19: Modeling, analysis and solutions," *Symmetry*, vol. 13, no. 5, p. 787, 2021, doi: 10.3390/sym13050787.
- [11] R. Cherniha and V. Davydovych, "A mathematical model for the covid-19 outbreak and its applications," *Symmetry*, vol. 12, no. 6, p. 990, 2020, doi: 10.3390/sym12060990.
- [12] M. M. De la Sen, A. Abeas, and R. Agarwal, "On confinement and quarantine concerns on an seiar epidemic model with simulated parameterizations for the covid-19 pandemic," *Symmetry*, vol. 12, no. 10, p. 1646, 2020, doi: 10.3390/sym12101646.
- [13] Berihuete, M. S. Sanchez, and A. S. Llorens, "A bayesian model of covid-19 cases based on the gompertz curve," *Mathematics*, vol. 9, no. 3, p. 228, 2021, doi: 10.3390/math9030228.
- [14] L. Taghizadeh, A. Karimi, and C. Heitzinger, "Uncertainty quantification in epidemiological models for the covid-19 pandemic," *Computers in Biology and Medicine*, vol. 125, p. 104011, 2020, doi: 10.1016/j.combiomed.2020.104011.
- [15] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A. Containing Papers of a Mathematical and Physical Character*, vol. 115, no. 72, pp. 700–721, 1927, doi: 10.1098/rspa.1927.0118.
- [16] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013, doi: 10.1109/TPAMI.2013.50.
- [17] P. Block et al., "Social network-based distancing strategies to flatten the covid-19 curve in a post-lockdown world," *Nature Human Behaviour*, vol. 4, pp. 588–596, 2020, doi: 10.1038/s41562-020-0898-6.
- [18] E. Mathew and S. Abdulla, "The LSTM technique for demand forecasting of e-procurement in the hospitality industry in the UAE," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 4, pp. 757–765, 2020, doi: 10.11591/ijai.v9.i4.pp757-765.
- [19] N. Rosili, R. Hassan, N. Zakaria, S. Kasim, F. Rose and T. Sutikno, "A systematic literature review of machine learning methods in predicting court decisions," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 4, pp. 1091–1102, 2021, doi: 10.11591/ijai.v10.i4.pp1091-1102.
- [20] A. Saud and S. Shakya, "Directional movement index based machine learning strategy for predicting stock trading signals," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 4185–4194, 2022, doi: 10.11591/ijece.v12i4.pp4185-4194.
- [21] C. Do, N. Q. Dam, and N. T. Lam, "Optimization of network traffic anomaly detection using machine learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 3, pp. 2360–2370, 2021, doi: 10.11591/ijece.v11i3.pp2360-2370.
- [22] P. Almasinejad et al., "Predicting the status of COVID-19 active cases using a neural network time series," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 3, pp. 3104–3117, 2022, doi: 10.11591/ijece.v12i3.pp3104-3117.
- [23] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986, doi: 10.1038/323533a0.
- [24] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp.1735–1780, 11 1997, doi: 10.1162/neco.1997.9.8.1735.
- [25] A. Holzinger, "From machine learning to explainable AI," *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, 2018, pp. 55-66, doi: 10.1109/DISA.2018.8490530.
- [26] G. Ellison, "Implications of heterogeneous sir models for analyses of covid-19," *National Bureau of Economic Research*, 2020, doi: 10.3386/w27373.
- [27] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000, doi: 10.1137/S0036144500371907.
- [28] N. G. Davies, A. J. Kucharski, R. M. Eggo, A. Gimma, and W. J. Edmunds, "Effects of non-pharmaceutical interventions on covid-19 cases, deaths, and demand for hospital services in the UK: a modelling study," *The Lancet Public Health*, vol. 5, no. 7, pp. 375–385, 2020, doi: 10.1016/S2468-2667(20)30133-X.
- [29] T. I. for Disease Modelling, "SEIR and SEIRS models," *The Institute for Disease Modelling, NBER Working Papers*, 2019. [Online]. Available: <https://institutefordiseasemodeling.github.io/Documentation/general/model-seir.html> (Accessed: 02- Apr- 2022).




- [30] S. Flaxman *et al.*, “Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe,” *Nature*, vol. 584, pp. 257–261, 2020, doi: 10.1038/s41586-020-2405-7.
- [31] A. Roddam, “Mathematical epidemiology of infectious diseases: model building, analysis and interpretation of diekmann and japh heesterbeek” *International Journal of Epidemiology*, vol. 30, no. 1, pp. 186–186, 2001, doi: 10.1093/ije/30.1.186.
- [32] M. Tschannen, O. Bachem, and M. Lucic, “Recent advances in autoencoder-based representation learning,” Dec. 2018, [Online]. Available: <http://arxiv.org/abs/1812.05069> (Accessed: 02- Apr- 2022).
- [33] Y. Bengio, “Learning deep architectures for AI,” *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009, doi: 10.1561/2200000006
- [34] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, “Mean absolute percentage error for regression models,” *Neurocomputing*, vol. 192, pp. 38–48, 2016, doi: 10.1016/j.neucom.2015.12.114.
- [35] T. Chai and R. R. Draxler, “Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding rmse in the literature,” *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014, doi: 10.5194/gmd-7-1247-2014
- [36] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” Dec. 2014, [Online]. Available: <http://arxiv.org/abs/1412.6980> (Accessed: 02- Apr- 2022).
- [37] “Information Portal of the Ministry of Health on Pandemic Covid-19”. [Online]. Available: <https://ncov.moh.gov.vn/web/guest/trang-chu> (Accessed: 02- Apr- 2022).
- [38] “GitHub - CSSEGISandData/COVID-19: Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE”, GitHub, 2022. [Online]. Available: <https://github.com/CSSEGISandData/COVID-19> (Accessed: 02- Apr- 2022).
- [39] E. Dong, H. Du, and L. Gardner., “An interactive web-based dashboard to track covid-19 in real time,” *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, 2020, doi: 10.1016/S1473-3099(20)30120-1.
- [40] McAloon *et al.*, “Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research,” *BMJ Open*, vol. 10, no. 8, p. e039652, 2020, doi: 10.1136/bmjopen-2020-039652.
- [41] N. H. An and D. T. Anh, “Comparison of strategies for multi-step-ahead prediction of time series using neural network,” *2015 International Conference on Advanced Computing and Applications (ACOMP)*, 2015, pp. 142–149. Doi: 10.1109/ACOMP.2015.24
- [42] F. Santosa and W. W. Symes, “Linear inversion of band-limited reflection seismograms,” *SIAM Journal on Scientific and Statistical Computing*, vol. 7, no. 4, pp. 1307–1330, 1986, doi: 10.1137/0907087.
- [43] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, pp. 273–297, 1995, doi: 10.1007/BF00994018.
- [44] J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, pp. 81–106, 1986, doi: 10.1007/BF00116251.

BIOGRAPHIES OF AUTHORS






Hoang-Sy Nguyen    received the B.S. and MS.c degree from the Department of Computer Science from Ho Chi Minh City University of Information Technology (UIT-HCMC), Vietnam in 2007, 2013, respectively. He received his Ph.D. degree in computer science, communication technology and applied mathematics from the VSB-Technical University of Ostrava-Czech Republic, in 2019. He is a senior researcher at Becamex Business School, Eastern International University, Vietnam. His research interests include applied mathematical, artificial intelligence, big data, and data analyst. He can be contacted at email: sy.nguyen@eiu.edu.vn.



Thu-Ngan T. Phan    is a student isssn Master of medical laboratory, Faculty Medical Laboratory, Hong Bang International University, Ho Chi Minh city, Vietnam. Her research interests are in fields of Molecular Biology, medicine, physiology, and data visual. She can be contacted at email: nganttt2110069@student.hiu.vn.



Cong-Danh Huynh    is a lecturer at the Faculty of Economics, Thu Dau Mot University, Binh Duong Province, Vietnam. His research interests are in computer science, economic law, data mining, and data analysis. He can be contacted at email: danhhc@tdmu.edu.vn.