ISSN 2616-5775 Vol. 6, Issue 11: 114-123, DOI: 10.25236/AJCIS.2023.061115

Prediction of confirmed cases of COVID-19 through time series models: A comparative study

Jinxuan Luo, Wenxuan Xue, Jin Zhao, Peiying Liu, Chenchen Yang

School of Science, Tianjin University of Commerce, Tianjin, 300134, China
jinxuan_luo2022@163.com

Abstract: *The widespread outbreak of Corona Virus Disease 2019 (COVID-19) poses a great risk to the lives and property of the world's citizens, especially in the USA, and Japan cases continue to increase dynamically. Several statistical models, machine learning models, and deep learning models were reported in the literature to forecast COVID-19 but there is no comprehensive report on all of them. This article analyzed several time series forecasting methods to predict the spread of COVID-19 during the pandemic wave in America and Japan (the period after 10 October 2022). The autoregressive moving average (ARIMA) model, support vector regression (SVR) model, and the long short-term memory (LSTM) model were employed to forecast the number of confirmed cases. In this study, machine learning and deep learning perform significantly better than traditional statistical models. The results show that SVR and LSTM are better single prediction models, which can help take precautions and policy formulation for this epidemic in other countries.*

Keywords: *ARIMA, SVR, LSTM, COVID-19, prediction, outbreak*

1. Introduction

COVID-19 is an acute respiratory syndrome that is transmitted through respiratory droplets and airways [1][2]. It is spreading very rapidly across the globe because of its highly contagious nature. As of 10 October 2022, the total number of global confirmed cases was 618521620 with 6534725 deaths. The World Health Organization (WHO) declared that the outbreak of COVID-19 still constitutes a "public health emergency of international concern" and is expected to last for a long time. Unfortunately, no full-proof drug is available to cure this deadly disease thus far [3]. To date, taking effective prophylactic measures could be the best method to decrease the cases. Besides, epidemic forecasting has also attracted ongoing research interest from both the academic and applied communities. Real-time and accurate prediction of potentially positive cases may guide local governments and societies to take more efficient actions in advance to prevent the spreading of COVID-19. Up to now, the time series model is one of the most popular methods to simulate the epidemiological process of diseases, which is based on the transmission characteristics and mechanisms of infectious diseases [4][5].

Different models were used to predict COVID-19 confirmed cases in recent studies. Alabdulrazzaq used Kuwait as a case study to test the accuracy of the ARIMA model to predict confirmed and recovered cases of novel coronavirus pneumonia, which showed that the ARIMA model can predict with reasonable accuracy[6][7]. Some researchers have also applied SVR models to pandemic prediction, such as Satyabrata Dash et al. who used support vector regression (SVR) and deep neural network methods to build prediction models with favorable results. Bandyopadhyay et al. have proposed the gated recurrent neural network and long short-term memory (LSTM) to evaluate the predictions with confirmed cases of COVID-19. Considering the many factors underlying the prevalence of COVID-19 and the complex nature of these factors leading to nonlinear, it is necessary to compare the effectiveness of different methods under time series forecasting models. Many researchers have missed this detail, it is a needed thing to conduct a sensitivity analysis on the model sample size and explore the effect of different sample characteristics on the experimental results.

As time-series data, COVID-19 cases are suitable for establishing a time-series model for prediction. Time-series forecasting is a classical class of problems that have been widely studied and applied in both academia and industry. Different models were used to predict COVID-19 confirmed cases in recent studies. Example, Autoregressive Integrated Moving Average (ARIMA) model, multiple linear regression[8], grey prediction [9], Long Short Term Memory(LSTM) model, Support Vector Regression(SVR) model, and Holt-Winters(HW) model [10]. Considering the many factors underlying

ISSN 2616-5775 Vol. 6, Issue 11: 114-123, DOI: 10.25236/AJCIS.2023.061115

the prevalence of COVID-19 and the complex nature of these factors leading to nonlinear, stochastic, and periodic data, it is necessary to compare the effectiveness of different models[11].

In this paper, daily new case data from 1 March 2020 to 10 October 2022 for America and Japan were selected and used to build the prediction model. We built ARIMA [12], SVR [13], and LSTM[14]models for predicting daily new cases in America and Japan in the next 20 days and evaluate the model's prediction accuracy to provide a further reference for the prediction and early warning of infectious diseases (As shown in Figure 1). Numerous studies have shown that different models have different predictive effects on different types of data, so we divided all datasets into six parts and performed sample size sensitivity analysis for each model. It is necessary to point out that the data interval between each part is six months. This work allows us to compare the applicability of ARIMA, SVR, and LSTM models in predicting the number of new patients per day in the US and Japan. It is also possible to compare the sample sensitivities of different models. More importantly, the prediction of the short-term virus trends provides a reference for the government's macroeconomic strategy and the allocation of emergency medical resources.

The remainder of the paper is organized as follows." Materials and Methods" present the data used in the analysis and the model employed." Evaluation Metrics" indicates the evaluation metrics used to measure the performance of the model. The "Results and Discussion" discusses the main results and findings. Finally, "Conclusion" provides some concluding considerations.

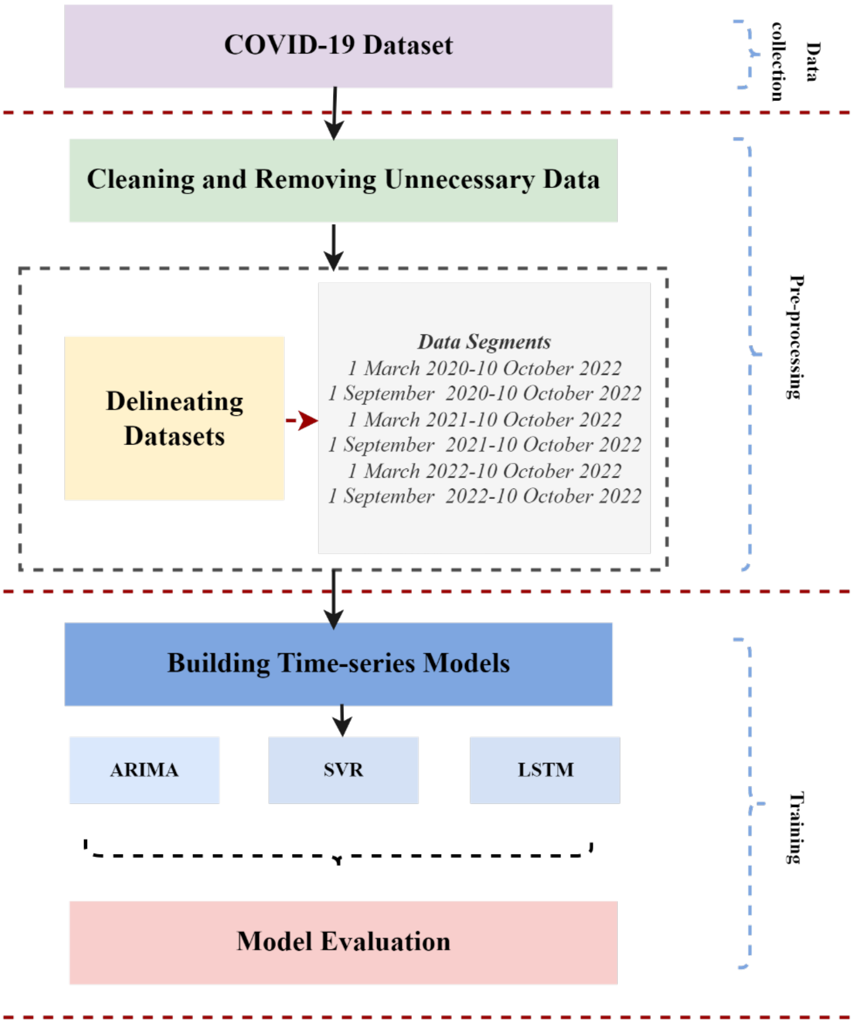

Figure 1: The proposed methodology of the COVID-19 forecasting model.

2. Materials and Methods

Data collection and pre-processing

The data array, which denotes the number of daily confirmed cases of COVID-19 from January 2020

ISSN 2616-5775 Vol. 6, Issue 11: 114-123, DOI: 10.25236/AJCIS.2023.061115

to July 2022, is abstracted from Johns Hopkins University. The confirmed cases are from the countries of America and Japan. For the robustness of the following time series models, we have detected and removed outlier points in the data. Experimental sample sizes, i.e., the beginning time T_i of utilized confirmed cases, have important influences on the prediction results. Therefore, we divided American and Japanese datasets into six parts as shown in Table 1.

Table 1: Description of the dataset start time

T_i	Beginning Time	T_i	Beginning Time
T_1	1 March 2020	T_4	1 September 2021
T_2	1 September 2020	T_5	1 March 2022
T_3	1 March 2021	T_6	1 September 2022

Fig. 2 presents the variation curves of daily new cases for both countries. Table 2 shows the statistical parameters of daily confirmed cases from the time T_i to Oct. 10, 2022, including the mean of values (Mean), the standard deviation of the values (Std), minimum value (Min), first quartile (25%), second quartile (50%), third quartile (75%), and maximum value (Max). We can find that the number of confirmed cases per day in America fluctuates greatly. Japan is recovering from a new wave of the outbreak, and the number of confirmed cases per day is on the decline.

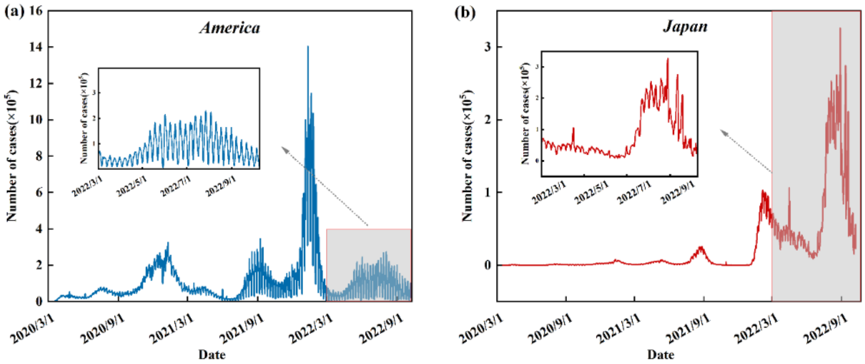

Figure 2: Time-series chart of the number of daily confirmed cases in America and Japan.

Table 2: Descriptive statistics on the daily confirmed cases of COVID-19 in America and Japan from the time Ti to 10 October 2022.

Country	T_i	Mean	Std	Min	25%	50%	75%	Max
America	T_1	103583	139738	9	30632	62197	130808	1398242
	T_2	120362	150520	4161	39825	76325	152773	1398242
	T_3	118113	167121	4161	33056	68678	135643	1398242
	T_4	144695	192611	4772	39781	96902	154537	1398242
	T_5	80301	62530	4772	29258	63153	123597	273512
	T_6	60151	45859	5826	20027	52197	86755	198955
Japan	T_1	22841	47454	9	506	2266	21368	326090
	T_2	28289	51441	48	1192	4223	34845	326090
	T_3	36518	56488	48	1743	12028	47377	326090
	T_4	50939	63342	48	2004	32911	65196	326090
	T_5	77862	70739	8905	31542	47409	100729	326090
	T_6	83858	56600	15167	41478	68718	105010	274599

Methods

Time-series forecasting, whose main task is to predict the future value of an indicator based on its historical data. The theory related to the time-series forecasting problem is more extensive, and the following three mainstream solutions are currently available [15].

First, time-series forecasting methods are based on traditional parametric models. The more classic models are AR, MA, ARMA, ARIMA [16], and the Prophet model introduced by Meta [17]. Second, time-series forecasting methods based on the machine learning model, commonly known as the decision tree (DT) model [18], support vector regression (SVR) model [19], and hidden markov (HMM) model developed from the markov (MM) model with multiple implicit states added on top of them [20]. Finally, time-series forecasting methods are based on deep learning models. The mainstream application scenarios of deep learning belong to two major fields, CV and NLP, the latter of which is specifically

ISSN 2616-5775 Vol. 6, Issue 11: 114-123, DOI: 10.25236/AJCIS.2023.061115

used to solve the problem of modeling sequence problems. Common models are the long short-term memory (LSTM) model, autoencoder (AE) model, and restricted boltzmann machines (RBM) model [21].

Traditional parametric models require researchers with sufficient domain knowledge and extensive practical experience to model accurately and obtain more precise predictions. Traditional machine learning models also require certain a priori knowledge to assist in model prediction. Deep learning models, on the other hand, do not require any domain knowledge because of their powerful feature learning capability, which can automatically capture the correlation between data. Deep learning models also have some limitations and the interpretability of the prediction results is poor [22].

In this paper, the ARIMA model, SVR model, and LSTM model are selected to predict univariate time-series data for America and Japan, respectively.

The ARIMA model

ARIMA model, which was first proposed by Box and Jenkins, represents one of the most widely used frameworks for pandemic and disease time-series predictions. The idea of the $ARIMA(p, d, q)$ model is to learn the patterns over time from historical data and summarize the patterns to predict the future. The model comprises the autoregressive (AR) model and moving average (MA) with integration based on the decomposition method [23]. The model parameters can be defined as follows: The parameter p is the corresponding order of the autoregressive model; the parameter d is the differencing order to obtain a smooth series, and the parameter q is the corresponding order of the moving average model. The following equation generalizes the p^{th} order AR model and q^{th} order MA model respectively [24].

$$y_t = A + \beta_1 y + \beta_p y_{t-p} + \cdots + \beta_n y_{t-n} + \theta_1 \varepsilon_{t-1} + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{1}$$

Where A is the intercept, $\beta_i (i = 1,2. \dots p)$ is an autoregressive parameter, y_t is the current time-series value, $y_{t-1}, y_{t-2}, y_{t-p}$ are past values and $\varepsilon_t = y_t - y_{t-1}$.

The SVR model

A support vector machine (SVM) is a type of supervised ML algorithm used for both regression and classification. The SVM regression algorithm is called support vector regression (SVR) [25]. The SVR model is a supervised learning algorithm that is used to predict discrete values. It is a nonlinear generalization and can be used as an algorithm for learning purposes. The SVR model is suitable for dealing with nonlinear small sample data by inputting fused values, mapping the two-dimensional input space to the high-dimensional space, and constructing an optimal decision function in the high-dimensional space to deal with the predictive regression problem in two dimensions. The linear function can be depicted as:

$$f(x) = x'\alpha + b \tag{2}$$

The objective is to make it as flat as possible to find the value of $f(x)$ with $(\alpha'\alpha)$ as the minimum parametrization. Therefore, the problem fits into the minimization function as follows:

$$J(\alpha) = \frac{1}{2}\alpha'\alpha \tag{3}$$

We make a condition for all residual values ε as in the following equation:

$$\forall_n : |y_n - (x'\alpha + b)| \leq \varepsilon \tag{4}$$

The LSTM model

To overcome the vanishing gradients and exploding gradients problem of the recurrent neural network (RNN) model, Hochreiter and Schmidhuber have proposed the LSTM model [26]. Unlike the RNN model, the LSTM model is designed with special storage units. This gate is composed of a Sigmoid neural network and matrix point-by-point multiplication operations. The input gate controls whether to allow new inputs, the forget gate ignores the unimportant information, and finally, the information is output through the output gate. The network can learn well about the long-term dependence on input data and remember long time historical data information.

Figure 3 shows the general architecture of the LSTM cell, which mainly consists of an input gate, output gate, and forget gate. The following formula is a mathematical expression for the forget gate, which determines which information can be ignored:

$$f_t = sigmoid\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{5}$$

ISSN 2616-5775 Vol. 6, Issue 11: 114-123, DOI: 10.25236/AJCIS.2023.061115

The following equation is a mathematical expression of the input gate:

$$i_t = sigmoid(W_t \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\tilde{C}_t = sigmoid(W_C \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

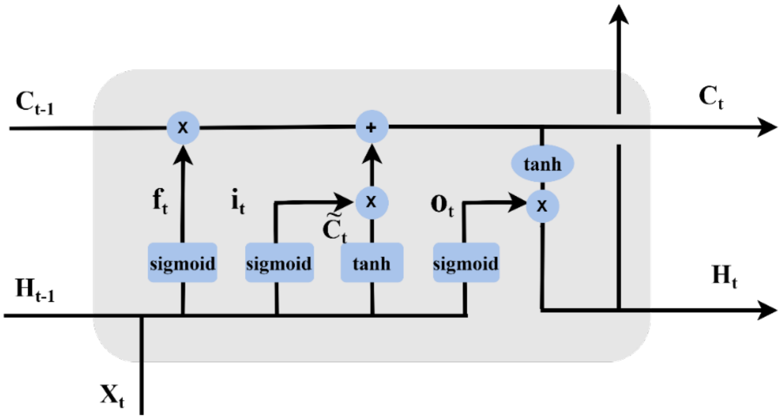

Figure 3: Schematic representation of Long-Short Term Memory (LSTM)

The following formula is a mathematical expression for the output gate:

$$o_t = sigmoid(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{8}$$

The final updated results are as follows:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{9}$$

$$h_t = o_t \cdot tanh(C_t) \tag{10}$$

The above symbols are interpreted as follows: $sigmoid$ are used to determine the update range of the value, $tanh$ are used to create a new input, C_{t-1} for the long-term memory of the previous period, h_{t-1} for the last period of short-term memory, x_t for the event information, f_t as the forgetting factor, it is the long-term memory output from the forgetting gate, $\tilde{C}_t$ for short-term memory with learned output, $W_{(f,i,c,o)}$ are the weight parameters, $b_{(f,i,c,o)}$ for deviations.

3. Evaluation Metrics

The main metrics used to compare the performances of the models were MAE, and RMSE.MAE is easier to interpret because minimizing it leads to predictions of the median while minimizing RMSE leads to predictions of the mean. The formulae used to calculate each of these metrics appears below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{12}$$

where n represents the number of observations, y_i denotes the actual values, and $\hat{y}_i$ indicates the predicted values [27].

The dataset has been preprocessed and divided into two subsets: the training set and the testing set (the last 20 days). The performance evaluation has been done in terms of important measures including mean absolute error (MAE), and root means square error (RMSE). The daily new confirmed cases for the next 20 days are reported in the Appendix.

4. Results and Discussions

Fig. 4 shows the variation of daily new confirmed cases with respect to the number of days for America and Japan. In Fig. 4 (a), we can see that the predicted value is close to the real value under the beginning time of T_1- T_5. In contrast, as shown in Figure 4 (b), there is a large gap between the predicted result and the actual value when the initial time is T_1- T_3, and the development trend of the predicted result is relatively gentle compared with the actual result. Under the setting of T_6 in Fig. 4 (a) and T_4- T_6

ISSN 2616-5775 Vol. 6, Issue 11: 114-123, DOI: 10.25236/AJCIS.2023.061115

in Fig. 4 (b), the prediction result becomes a straight line, which is caused by the fact that the ARIMA model cannot recognize the seasonal/cyclical characteristics in this timing data. Instead of using that simple ARIMA model, the SARIMA(Seasonal Auto Regressive Integrated Moving Average) model can be considered for further study.

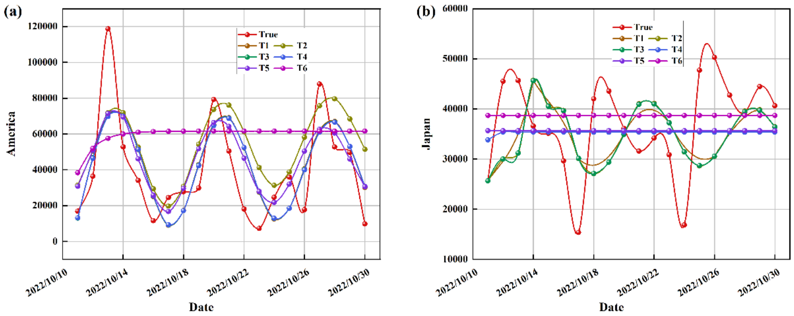

Figure 4: Daily new confirmed cases under different beginning times Ti relative to the number of days by the ARIMA model. (a) and (b) are for America and Japan, respectively.

To further analysis, we calculate the parameters of mean absolute error (MAE) and root means square error (RMSE) for the ARIMA model with different beginning times T_i , as shown in Fig. 5. Fig. 5 shows that the Japanese data are more suitable for the ARIMA model compared to the American data and have a higher accuracy of prediction. The change of T_i has a significant impact on the error of the America model, which is not monotonously increasing or decreasing. The model with the beginning time of T_5 has the best prediction results. There are two explanations for this. One is that more historical data means that the coefficients may be better optimized to describe what happens with the variability from more years of data. Another view is that the timeliness of the data is more important. Therefore, it is necessary to consider the sample size in our prediction. For the Japanese data, as the sample size decreases, the model error also decreases, although the effect is not significant.

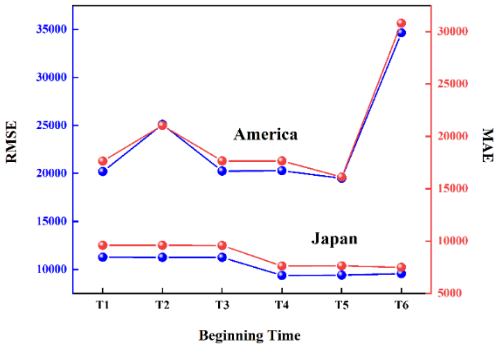

Figure 5: Parameters of mean absolute error (MAE) and root means square error (RMSE) for the ARIMA model from the different beginning times T_i.

The SVR model has an advantage when dealing with nonlinear data, compared with Fig. 4 (a), the forecast results of America in Fig. 6 (a) are more similar to the actual situation. In Fig. 6 (a), the predicted trend with different beginning times T_i is generally consistent with the real trend, but there is still a large value gap. In Fig. 6 (b), it can be found that there is a large gap between the prediction results corresponding to different beginning times T_0, the specific model error can be seen in Fig. 6.

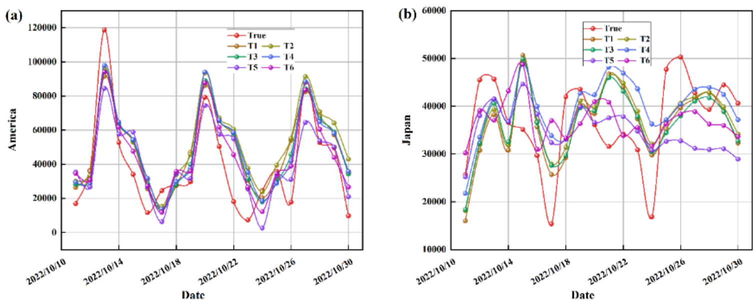

Figure 6: Daily new confirmed cases under different beginning times T_i relative to the number of days by the SVR model. (a) and (b) are for America and Japan, respectively.

ISSN 2616-5775 Vol. 6, Issue 11: 114-123, DOI: 10.25236/AJCIS.2023.061115

In the SVR model, the Japanese data are not sensitive to the sample size. When the beginning time of American data is from T_2 to T_6, as the sample size decreases, the model error also decreases. The two error indicators for the Japanese data show different variations. The RMSE indicator reaches its maximum at the beginning time of T_4 and then starts to decrease, while the MAE indicator decreases step by step, as shown in Figure 7.

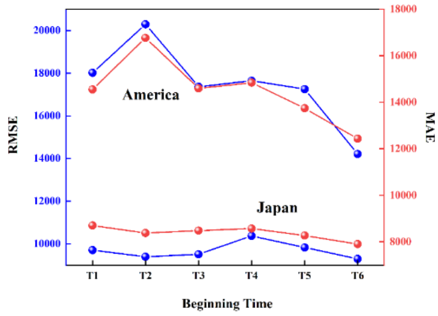

Figure 7: Parameters of mean absolute error (MAE) and root means square error (RMSE) for the SVR model from the different beginning times T_i.

To avoid the influence of human factors on the data set and fully exploit the nonlinear relationship between the time-series data, this paper constructed an LSTM neural network prediction model [28]. In Fig. 8, we show the real values and the predicted values from October 11, 2022 to October 30, 2022. The disparity between the model prediction results at different beginning times T_i is significant, especially in the performance of the Japanese dataset. From the image morphology, the prediction performance of the LSTM model is worse than the SVR model. The most important difference between deep learning and traditional machine learning manifests itself as the size of the data increases. Deep learning algorithms do not perform well when the data is very small. This is because deep learning algorithms require a large amount of data to understand it perfectly.

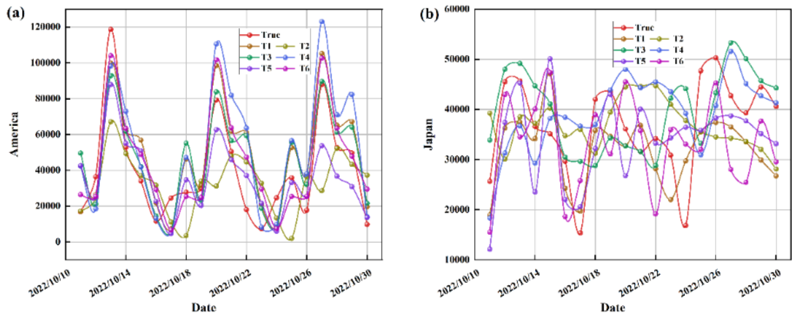

Figure 8: Daily new confirmed cases under different beginning times T_i relative to the number of days by the LSTM model. (a) and (b) are for America and Japan, respectively.

As seen in Fig. 9, the LSTM model is sensitive to the sample size, which shows Irregular variation of error Indicators, both for the American data and the Japanese data. There is a large gap between the highest and lowest value of the model error indicators. Therefore, we should focus on the sample size factor when using the model for forecasting.

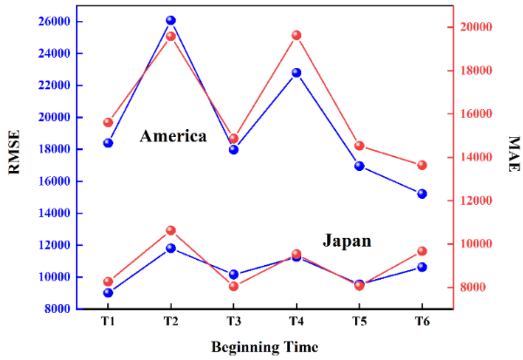

Figure 9: Parameters of mean absolute error (MAE) and root means square error (RMSE) for the LSTM model from the different beginning times T_i.

ISSN 2616-5775 Vol. 6, Issue 11: 114-123, DOI: 10.25236/AJCIS.2023.061115

Discussion

In the model fitting and prediction phase, the LSTM model and the SVR model performed better than the ARIMA model when comparing models for predicting new cases in different countries. Because of the sample size, the difference between the accuracy of LSTM and SVR models is not significant. It must be noted that the above test results only illustrate the performance of the three schemes on this dataset and are not representative of the performance of this class of models when used for time-series prediction problems. The ARIMA model is very simple and requires only endogenous variables without the help of other exogenous variables, but it can only capture linear relationships but not nonlinear relationships. The SVM and LSTM models can handle nonlinear relationships in the data better, but the SVM model requires higher parameter settings and the LSTM model requires a higher sample size for the training set. In conclusion, the experiments show that time-series can be used well to predict the trend of COVID-19. Much of the value of models in the COVID-19 pandemic is in informing immediate policy decisions. To povide the government with a clear picture of developments so that they can respond effectively, as shown in Figure 10.

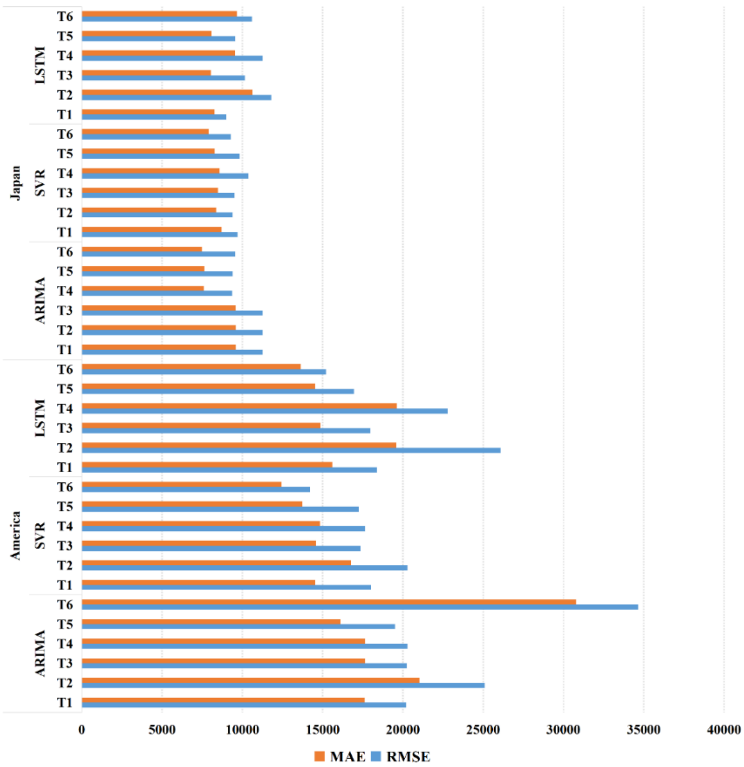

Figure 10: Accuracy evaluation of ARIMA, SVM, and LSTM on fitting and forecasting COVID-19 in America and Japan with different beginning times T_i. The blue color represents the indicator RMSE and the orange color represents the indicator MAE.

5. Conclusion

This study briefly describes the prevalence of COVID-19 in America and Japan. We attempted to predict the short-term dynamics of patients with confirmed COVID-19 in America and Japan, using a single time-series prediction method. The prevalence development was estimated using ARIMA, SVR, and LSTM models. The results show that the best single model for this prediction is the SVR model. Although time-series prediction models proved to be accurate, it still needs to be emphasized that this approach may lead to unavoidable uncertainties and biases, such as some governmental prevention and control initiatives that are not taken into account. Combining time-series prediction models with infectious disease dynamics models will be a direction for our further research. The current work can provide a reference for the government to formulate epidemic prevention and control policies and economic development policies, and provide more basis and value for the prediction of the COVID-19 epidemic.

References

[1] Fauci AS, Lane HC, Redfield RR. Covid-19-navigating the uncharted[J]. New England Journal of Medicine, 2020, 382(13): 1268-1269.
[2] Wu Z, McGoogan J M. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention [J]. Jama, 2020, 323(13): 1239-1242.
[3] Bavel J J V, Baicker K, Boggio P S, et al. Using social and behavioral science to support COVID-19 pandemic response[J]. Nature human behavior, 2020, 4(5): 460-471.
[4] Kumar N, Susan S. COVID-19 pandemic prediction using time series forecasting models[C]//2020 11th international conference on computing, communication and networking technologies (ICCCNT). IEEE, 2020: 1-7.
[5] Zeroual A, Harrou F, Dairi A, et al. Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study[J]. Chaos, Solitons & Fractals, 2020, 140: 110121.
[6] Rahimi I, Chen F, Gandomi A H. A review on COVID-19 forecasting models[J]. Neural Computing and Applications, 2021: 1-11.
[7] López L, Rodo X. A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics[J]. Results in Physics, 2021, 21: 103746.
[8] Rath S, Tripathy A, Tripathy A R. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model[J]. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2020, 14(5): 1467-1474.
[9] Saxena A. Grey forecasting models based on internal optimization for Novel Corona virus (COVID-19)[J]. Applied Soft Computing, 2021, 111: 107735.
[10] Panda M. Application of ARIMA and Holt-Winters forecasting model to predict the spreading of COVID-19 for India and its states[J]. medRxiv, 2020.
[11] Perone G. Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy[J]. The European Journal of Health Economics, 2022, 23(6): 917-940.
[12] Sahai A K, Rath N, Sood V, et al. ARIMA modelling & forecasting of COVID-19 in top five affected countries [J]. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2020, 14(5): 1419-1427.
[13] Sapankevych N I, Sankar R. Time series prediction using support vector machines: a survey[J]. IEEE computational intelligence magazine, 2009, 4(2): 24-38.
[14] Han Z, Zhao J, Leung H, et al. A review of deep learning models for time series prediction[J]. IEEE Sensors Journal, 2019, 21(6): 7833-7848.
[15] Liu Z, Zhu Z, Gao J, et al. Forecast methods for time series data: a survey[J]. IEEE Access, 2021, 9: 91896-91912.
[16] Talkhi N, Fatemi N A, Ataei Z, et al. Modeling and forecasting number of confirmed and death caused COVID-19 in IRAN: A comparison of time series forecasting methods[J]. Biomedical Signal Processing and Control, 2021, 66: 102494.
[17] Satrio C B A, Darmawan W, Nadia B U, et al. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET[J]. Procedia Computer Science, 2021, 179: 524-532.
[18] Rady E H A, Fawzy H, Fattah A M A. Time series forecasting using tree based methods[J]. J. Stat. Appl. Probab, 2021, 10: 229-244.
[19] Tealab A. Time series forecasting using artificial neural networks methodologies: A systematic review [J]. Future Computing and Informatics Journal, 2018, 3(2): 334-340.
[20] Li S T, Cheng Y C. A stochastic HMM-based forecasting model for fuzzy time series[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2009, 40(5): 1255-1266.
[21] Sezer O B, Gudelek M U, Ozbayoglu A M. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019[J]. Applied soft computing, 2020, 90: 106181.
[22] Torres J F, Hadjout D, Sebaa A, et al. Deep learning for time series forecasting: a survey[J]. Big Data, 2021, 9(1): 3-21.
[23] Satrio C B A, Darmawan W, Nadia B U, et al. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET[J]. Procedia Computer Science, 2021, 179: 524-532.
[24] Ning Y, Kazemi H, Tahmasebi P. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet[J]. Computers & Geosciences, 2022, 164: 105126.
[25] Rustam F, Reshi A A, Mehmood A, et al. COVID-19 future forecasting using supervised machine learning models[J]. IEEE access, 2020, 8: 101489-101499.
[26] Manaswi N K. Rnn and lstm[M]//Deep Learning with Applications Using Python. Apress, Berkeley,

ISSN 2616-5775 Vol. 6, Issue 11: 114-123, DOI: 10.25236/AJCIS.2023.061115

CA, 2018: 115-126.

[27] Wang W, Lu Y. Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model[C]//IOP conference series: materials science and engineering. IOP Publishing, 2018, 324(1): 012049.

[28] Sagheer A, Kotb M. Time series forecasting of petroleum production using deep LSTM recurrent networks [J]. Neurocomputing, 2019, 323: 203-213.

Appendix

The results of the models for confirmed cases in America.

Country	**America**																	
Model	**ARIMA**						**SVR**						**LSTM**					
T_i	T_1	T_2	T_3	T_4	T_5	T6	T_1	T_2	T_3	T_4	T_5	T6	T_1	T_2	T_3	T_4	T_5	T6
2022/10/11	13067	31418	13038	13190	30789	38377	26467	29973	28239	29579	35411	34598	42426	17166	49611	42410	42586	26488
2022/10/12	46873	50560	46682	46482	50920	51933	35728	33398	30336	28832	26772	30813	26461	24431	21153	18520	24886	25502
2022/10/13	70286	71767	69957	69753	71456	57578	91577	96023	94206	97857	84451	94121	99873	67136	92796	98139	87894	104032
2022/10/14	70236	71893	69998	69975	69386	59929	63865	63723	63649	64793	57663	61805	62160	49391	61804	72949	55134	63358
2022/10/15	51290	52630	51333	51382	46143	60908	53049	54358	53998	54416	58844	47645	56888	37057	42164	41798	50903	48888
2022/10/16	25132	29423	25345	25394	25817	61316	25682	30968	27934	31765	26799	27292	21843	31668	14510	13202	22699	29209
2022/10/17	9065	19806	9222	9365	16719	61485	14053	15264	12827	13764	6368	11855	7397	11168	5810	5067	4672	6715
2022/10/18	17261	30776	17205	17228	29613	61556	27554	33987	28166	29814	34365	35748	46256	3570	55195	47174	34710	25465
2022/10/19	42864	54293	42542	42344	51735	61586	46995	45811	39969	37555	31833	36046	31690	34003	23307	24744	20397	24779
2022/10/20	65419	73715	64961	64725	66377	61598	86109	93534	89001	94015	74388	87763	98458	31263	83867	110493	62705	101719
2022/10/21	69080	76091	68786	68688	64284	61603	65739	67207	65452	65847	57576	61706	61722	48116	56578	81902	45922	63753
2022/10/22	52382	61232	52451	52501	46490	61605	56301	60520	57971	58573	54949	45602	60872	44692	59330	63594	37095	47353
2022/10/23	27438	41257	27768	27939	27973	61606	30384	37653	31513	35272	25499	26593	20747	32851	18761	8254	21590	29549
2022/10/24	12522	31364	12809	12984	21836	61606	20330	23854	17923	18557	2643	12286	8475	13414	7796	9828	5953	7510
2022/10/25	18597	38831	18567	18567	32023	61606	31586	39484	29537	28874	33199	34572	52714	2065	55636	56571	33192	25417
2022/10/26	40604	58191	40192	39976	50389	61606	54081	55151	45797	41991	31102	38869	37813	36496	32105	38128	25507	25550
2022/10/27	61860	75724	61286	61008	62713	61606	82434	91364	84280	88320	64517	83504	105270	28608	89667	123062	53713	102626
2022/10/28	67026	79646	66664	66511	60494	61607	68007	70740	66527	64951	53769	60493	65199	52325	61286	71117	36767	63494
2022/10/29	52858	68430	52938	53002	45989	61607	57263	64218	58786	58107	50114	44006	67187	43371	64091	82286	30855	47849
2022/10/30	30115	51421	30533	30763	30541	61607	35151	43061	34339	35774	20979	26703	19715	37331	21594	14091	13626	29539

The results of the models for confirmed cases in Japan.

Country	**Japan**																	
Model	**ARIMA**						**SVR**						**LSTM**					
T_i	T_1	T_2	T_3	T_4	T_5	T6	T_1	T_2	T_3	T_4	T_5	T6	T_1	T_2	T_3	T_4	T_5	T6
2022/10/11	25656	25661	25665	33836	35656	38694	16016	18153	18376	21751	25230	30243	18968	39245	33871	18348	12135	15554
2022/10/12	30035	30024	30005	35367	35656	38694	30733	32003	32275	33523	38024	39101	36290	30108	48002	31299	37392	42971
2022/10/13	31249	31228	31199	35367	35656	38694	38220	39311	40592	41510	41524	37053	37461	38528	49168	36700	45259	34525
2022/10/14	45716	45676	45616	35367	35656	38694	30810	31898	32535	36828	36730	43207	34191	37201	44685	29290	23527	40082
2022/10/15	40584	40553	40507	35367	35656	38694	50636	49566	49892	48793	44638	48736	47112	40254	41087	38214	50096	47368
2022/10/16	39663	39647	39621	35367	35656	38694	35653	36893	36569	39957	38368	30994	24298	34765	30482	38432	22045	18634
2022/10/17	30111	30119	30133	35367	35656	38694	25719	27894	27609	33851	32379	36985	19754	35975	29645	36675	20625	25814
2022/10/18	27062	27081	27111	35367	35656	38694	29262	31407	29675	33180	33133	33365	35793	31246	28788	36948	32212	38923
2022/10/19	29384	29388	29397	35367	35656	38694	40198	41108	39647	42686	39903	36337	34636	39425	34298	43861	42994	31122
2022/10/20	34966	34944	34912	35367	35656	38694	38438	39804	38991	42380	36576	40929	32744	44494	32719	47959	26794	45489
2022/10/21	41020	40973	40900	35367	35656	38694	46764	46333	45957	48195	37560	40835	36814	44584	31590	44323	40063	35745
2022/10/22	41089	41044	40973	35367	35656	38694	44067	44829	43125	46923	37823	33857	28172	44745	28793	45460	33256	19191
2022/10/23	37269	37249	37216	35367	35656	38694	37366	38957	37792	43611	34786	35544	22024	41001	42197	43542	34325	35962
2022/10/24	31393	31405	31425	35367	35656	38694	29772	32097	30456	36293	30759	31673	29728	37781	44143	39215	36454	33078
2022/10/25	28640	28668	28713	35367	35656	38694	35165	36897	34457	37183	32624	36295	35541	35555	33240	30932	35837	32028
2022/10/26	30543	30555	30577	35367	35656	38694	39636	40638	37996	40415	32763	38371	37394	34489	43284	40747	38282	45297
2022/10/27	35391	35368	35335	35367	35656	38694	41543	41954	41054	43529	31180	38852	36504	34257	53292	51582	38718	28045
2022/10/28	39553	39501	39421	35367	35656	38694	42825	42871	41781	43928	30931	36286	33503	33592	50104	45145	37714	25479
2022/10/29	39828	39777	39695	35367	35656	38694	39023	39877	38852	42436	31076	35961	29911	32023	45727	42707	35152	37691
2022/10/30	36439	36417	36381	35367	35656	38694	32286	34162	32729	37165	28935	33653	26740	28111	44286	41314	33162	29535