Journal of Advanced Research Design 137, Issue 1 (2026) 33-46

Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

An Advanced Research Based on Machine Learning Techniques with Variant ARIMA Methods for Identifying Most Severity COVID-19 Data

You-Shyang Chen[1,*], Yi-Xuan Chen[1]

[1] College of Management, National Chin-Yi University of Technology, Taiping District, Taichung 411030, Taiwan

ARTICLE INFO

Article history:
Received 3 February 2025
Received in revised form 7 March 2025
Accepted 30 June 2025
Available online 20 July 2025

Keywords:
New COVID-19; confirmation case; time series model; autoregressive integrated moving average (ARIMA); machine learning technique

ABSTRACT

The COVID-19 epidemic has been a hot topic for a time-series forecasting in the last three years (2020-2022) data, and their national conditions and environmental backgrounds of different countries have different; thus, it is difficult but necessary to find a more suitable time-series model for prediction of relevant data within a short period of certain time. Therefore, tailored to countries to find out such an intelligent model to be able to predict the confirmation cases of coronavirus for different countries has an interesting and important challenge. COVID-19 studies have used traditional statistics-based research architectures, and there are several core problems: (1) it is lack of effectively organizing time-series forecasting models with machine learning techniques for COVID-19 data; (2) it may require different forecasting models depended on characteristics of different yearly COVID-19 data; (3) it is a valuable issue to find out a well-off forecasting model for modulating different parameters of COVID-19 data contexts. This study is trigged by such a research motivation to propose an integrated approach of different three time series-based models, such as autoregressive (AR), moving average (MA), and autoregressive integrated moving average (ARIMA), along with 12 machine learning models for addressing COVID-19 data applications. The study collects the data of confirmation cases from 2020-2022 and divides it into different sub-datasets from the most severities COVID-19 of countries to assess the effect of the proposed model. The evaluation standard is an effective indicator of mean absolute percentage error (MAPE). For the empirical results, it is found that different predictive techniques for the same country's epidemic data have different prediction error rates; it is proved that different COVID-19 data of different countries needs different forecasting methods. The empirical results provide valuable information as a useful reference for different interest considerations to the interested parties. It is a good issue to differentiate from the existing works, the study highlights an organization method for three time-series models and 12 machine learning techniques in matching a COVID-19 data application; thus, this study has a clear contribution of a technical and applicable innovation on benefiting COVID-19 data fields.

1. Introduction

This study aims to explore and address some emerged meaningful issues for COVID-19 data occurred in the recent years (2020–2023) to identify time-series forecasting models for number of

* *Corresponding author*
E-mail address: yschen@ncut.edu.tw

https://doi.org/10.37934/ard.137.1.3346

confirmed cases. Five core research subjects are defined, including the first research background, research motivation, research problem, research originality and its importance and research purpose at last, helpful to completely learn more about this study; they are respectively described, as follows.

First, on March 9, 2020, the World Health Organization (WHO) announced the new coronavirus, properly named "New Corona Virus Pneumonia 2019" (abbreviated as COVID-19) [1-3], which has become the world's largest epidemic virus start and continually, causing at least hundreds of millions of people to be diagnosed and millions of deaths worldwide lastly. Based on this major and terrible catastrophe, the world that has originally smooth life is instantly and completely changed; the damage for us is very heavy. Therefore, we can draw the historical retrospective back to the potential influence from the first case of actual COVID-19 abroad [4-6] with its subsequent development during 2020–2023. This serious COVID-19 causes countries to initiate vaccination policies, to have even closed border exchanges, and to enter the blocking phase, leading to global economic stagnation and even recession [7], and even the unemployment population has increased dramatically. More seriously, according to an updated statistical tracking from the reported coronavirus cases of Worldometer [8], there are currently 704,753,890 confirmed cases and 7,010,681 deaths worldwide from the coronavirus COVID-19 outbreak as of April 13, 2024. Thus, there is no doubt that the largest and most influential international news in recent times is the pneumonia epidemic of COVID-19 [9-12], and its harm has still continued to present. Due causing the world’s serious damage is very horrifying, it is too complex and difficult to predict and control for the continuously subsequent results. Thus, it will be a valuable issue for further addressing the great bodily injury for the COVID-19 data by taking and identifying the challenge from background of research in the last four years (2020 to 2023).

Next, the Global outbreaks were intensifying from 2020 to the end of the second year of the outbreak (i.e., end of 2021) as a period of time when we do an important review. For a clear presentation, we just compare the statistics of December 28, 2021, showing the world’s top 10 most severe outbreaking regions or countries and making a meaningful comparison, as shown in the Table 1 below. From the data of Table 1, we obtain the following four useful references:

i. The top three countries with the most seriously diagnosed orderly are the United States (USA), India, and Brazil, the highest number of deaths are also these three countries, and the world reaches more than 220 million confirmed cases and over 5.42 million deaths.
ii. The most severe outbreak situation is concentrated in countries in the three continents of America, Eurasia and Asia; comparatively, the Oceania countries are less severe.
iii. The three countries in order with the highest confirmation rate are the United Kingdom 18.762%, the United States 16.115% and France 14.194%.
iv. The three countries in order with the highest mortality rates are Russia 2,927%, Brazil 2.780% and Iran 2.125%. Given the above information, it is a serious challenge with an important and interesting topic that is worth looking at continually, and it is definitely that we can learn something from it.

Therefore, exploring such a global COVID-19 data for identifying sufficient knowledge of epidemic prevention is absolutely necessary and significantly motivates the study.

Table 1
Comparison of Covid-19 statistics for the top 10 major epidemic countries on December 28, 2021

No.	Country	Region	Confirmed cases	Total population	Confirmation rate (%)	Number of deaths	Death rate (%)
1	USA	America	53,659,688	332,970,000	16.115	822,892	1.534
2	India	Asia	34,808,886	1,395,000,000	2.495	480,592	1.381
3	Brazil	America	22,269,031	213,930,000	10.409	619,095	2.780
4	UK	Europe	12,585,924	67,081,000	18.762	148,202	1.178
5	Russia	Across Europe and Asia	10,279,009	145,760,000	7.052	300,886	2.927
6	France	Europe	9,579,277	67,486,000	14.194	123,434	1.289
7	Turkey	Across Europe and Asia	9,367,369	83,614,000	11.203	81,917	0.874
8	Germany	Europe	7,129,352	83,129,000	8.576	111,607	1.565
9	Iran	Asia	6,190,762	84,823,000	7.298	131,527	2.125
10	Spain	Europe	6,133,057	47,394,000	12.941	89,331	1.457

Accordingly, it is a truth that we can have always believed that there are some rules or relationships between the numbers in the relevant time-series data; there may exist a relationship between a certain sequence of data (e.g., COVID-19 data). We take a case, and Table 2 shows a partial data from the daily confirmed cases and daily deaths of COVID-19 diagnoses in the United States during January 1, 2020 to December 31, 2021; undoubtedly, it has the data phenomena of time-series, trends [13-15], auto-correlation [16,17], and other possible correlations. There are some effective forecasting models that can more accurately predict future data, and we can use useful data to model the better intelligent forecasting methods [18] as an effective tool for accurately predicting what will happen next. Therefore, we can excavate treasures from the hidden information of time-series data meaningfully. Given the above reason, this study explores and defines a research problem in order to identify international COVID-19 time-series data [19], seeking to establish guidelines of future effective forecasting decision-making model for accurately predicting country-specific COVID-19 series data, such as daily confirmed cases.

In an in-depth literature review, it has been found that COVID-19 medical research topics are quite diverse [19], and researchers are more likely to use traditional statistical-based research architectures. Although they can result in different degrees of satisfaction, there are some major shortcomings in using traditional statistics methods. Except for the combination of conventional statistical methods with COVID-19 issues, there are some possible major problems:

i. A lack of effective predictive model related to COVID-19 indicators in the organization of time-series models and machine learning techniques, resulting in difficulties or setbacks in health policy and national resource-related decision-making.
ii. Due to certain national environmental differences, the forecasting model may be misleading, ineffective, or insufficient, so that the forecasting model needs to be opportunely revised for different countries.
iii. Not a forecasting model can be completely and suitably applied to all epidemiological data in different contexts, so that the forecasting model must be adjusted timely for different background parameters.

Table 2
Daily COVID-19 diagnoses and daily deaths in the United States from January 1 2020 to December 31 2021

Date	Daily diagnostic cases	Daily death cases
2020/1/1	0	0
2020/1/2	0	0
2020/1/3	0	0
2020/5/30	25,337	1,219
2020/5/31	23,297	945
2021/12/30	377,014	2,337
2021/12/31	489,267	2,184

By the way, this study expects to propose an advanced intelligent time-series forecasting model [20], focusing on solving different conditions, such as the country background environment, data characteristics, or specific time variables, expecting to individually find its best forecasting model, and improving prediction efficiency for time-series data of COVID-19 epidemic [21]. Moreover, this study tries to get rid of the shortcomings and major problems mentioned-above by developing a technical basis of an improved ARIMA-based model for predicting time-varying properties data, including autoregressive (AR(p)), moving average (MA(q)), autoregressive integrated moving average (ARIMA (p, d, q)) [22], based on their past excellent performance and research trend. Afterwards, due to the eminent evaluation performance for the mean absolute percentage error (MAPE) from the past research [23], it is adopted as the main performance metric to widely evaluate the effectiveness and applicability of the time-series forecasting models used in the field of machine learning techniques, providing valuable information for the empirical experiences from the data analysis and prediction.

More importantly, regarding the machine learning techniques, we good organize some effectively supervised algorithms [24,25], such as Gaussian Processes (abbreviated as GP), Linear Regression (LR), Multilayer Perceptron (MLP), SMO reg (SMO-R), Bagging, Decision Table (DT), M5 Rules (M5-R), Decision Stump (DS), M5P, Random Forest (RF), Random Tree (RT), and REP Tree (REP-T), which have a well-known and state-of-the-art study utilized for familiar learning methods in different domains, into the proposed model in order to well address the COVID-19 data. They are used as a key purpose of comparative study for measuring the prediction ability in this study. Thus, based on the above descriptions, this study particularly has the research originality and rich research architecture in terms of technicality and applicability.

Finally, we thus outline the three main research objectives:

i. Construct a set of time-varying and time-series forecasting methods, including AR(p), MA(q), and ARIMA (p, d, q) to predict the COVID-19 series data.
ii. Use the assessment indicator MAPE to objectively evaluate and identify the better advanced intelligent time-series models for different countries under different years' data.
iii. Use 12 machine learning techniques as a comparative analysis.

The remaining contents of this paper is organized in the following four sections. Section 2 reviews the related background introduction for COVID-19 and ARIMA. Section 3 is the related research processes and the processing flow for the proposed model. Section 4 makes the experimental results and data analysis results, and Section 5 is the conclusion and the future research direction.

2. Literature Review
2.1 COVID-19 Case with the Related Application Studies

At the end of 2019, the new coronavirus (COVID-19) originated in the city of Wuhan, mainland Chinese, and this virus is extremely infectious and has a high mortality rate at the beginning of its occurrence; thus, it has been declared a "severe specific infectious pneumonia" as a statutory infection virus. The incubation period of approximately 1 to 14 days (or even longer) is a severe virus with a longer incubatory period known today. So far, the virus has sparked a global economic crisis, causing a severe shock to the United States, the world's largest economic hub [26]. Specifically, all the economies of the world are greatly negatively affected by it. COVID-19 is a type of membrane RAN virus, and the symptoms are very similar to the common cold. The initial occurrence of cases with pneumonia becomes main symptom, but in serious, it will affect the entire lung infection, fever, cough, muscle pain, breathing difficulties and other symptoms. The main way of infection is through cough and spray sputum, infected to others, and it can even be transmitted through direct connection with the human body; currently, although there are related vaccines used, due to the evolution of the virus, it often causes many breakthrough infection diagnosis cases. Based on relevant statistics, although about 85% of COVID-19 patients are mild and common types of cases, it is clinically found that some patients are suddenly worsen, rapidly progress to severe or critical types, and increase mortality rate [5], and this particular situation deserves special attention and further exploration of possible reasons.

COVID-19 has spread globally, causing numerous deaths worldwide, leading governments around the world to implement relevant vaccination measures. According to Haussner *et al.,* [4] and Fauci *et al.,* [27], their researches showed that countries in order to control the spread of the epidemic, restrict people's travel abroad, and therefore enter the so-called "new locking the country" era, hoping to be able to control epidemics by reducing the displacement, gatherings and activities of personnel. Lauer *et al.,* [28] noted that scholars and experts' study clinical cases to understand the incubation period of the virus, reduce the number of infections and control the outbreak. Jiang *et al.,* [29] and Xu *et al.,* [30] also had noted that COVID-19 infection symptoms through sputum contact include nausea, cough, chest pain, dizziness, vomiting, fever, coughs, fatigue, diarrhoea and abdominal pain, and it can cause acute heart muscle damage and chronic damage to the cardiovascular system [31]. Studies had found that COVID-19 is a high-risk group of susceptible and severely ill patients, such as people with diabetes or other chronic diseases [32] or people with poor immune function (immunodeficient patients, infants and pregnant women) and smokers [33].

2.2 The ARIMA Model with its Applications

The ARIMA model is the first time series model proposed by Box *et al.,* [34], a predictive model consisting of three elements, such as AR, difference, and MA; its time sequencing data is taken simple averages, and excludes random fluctuations, making time sequential data smoother and combining with other models to make smooth predictions. The series data can be processed through "first difference", converting its sequence into a stable series, and then using the above ARMA model to make predictions. According to past paper review, the ARIMA model, because it has multiple adjusted parameters, can be more accurate to do the factual variables, and therefore has higher outstanding predictive performance, and it can be applied to a variety of different industry data applications, such as carbon reduction analysis [35], multiple QoS prediction [36], prediction of daily and monthly average global solar radiation [37] and predication of the indices of the consumer price index and the expected number of cancer patients [22]. From the aforementioned study of Hadwan *et al.,* [22], we found that ARIMA is applicable in the medical industry, so that the ARIMA model was selected to be used into the constructed structural model of this study to predict COVID-19-related epidemic indicator data. The ARIMA model is a widely used statistical method for time series forecasting, particularly effective for analysing and predicting data with temporal dependencies. It consists of three main components: the auto-regressive (AR) term, which represents the relationship between a value and its previous values, with p indicating the number of lagged observations included in the model; the integrated (I) term, which applies differencing to make the time series stationary, with d representing the number of differencing operations; and the moving average (MA) term, which captures the dependency between an observation and the residual errors from past observations, with q specifying the number of lagged error terms [38]. In this model algorithm, the ARIMA Eq. (1) is formatted and represented as follows; in ARIMA (p, d, q), AR is the above self-regression, p is the number of auto regression items, MA is the aforementioned moving average, q is the number of moving average terms, d is the differential times (order number) for stationary series, L is the lag operator, and it is d > 0 [39].

$$\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t=\left(1-\sum_{i=1}^{p}\emptyset_i\, L^i\right)(1-L)^d X_t. \tag{1}$$

3. The Proposed Model

3.1 Research Method and the Data Used

Based on the above clear descriptions, this study aims to overcome the major problems faced above by proposing three methods of related time-series models (i.e., AR(p), MA(q), and ARIMA(p, d, q)) and 12 machine learning techniques (including GP, LR, MLP, SMO-R, Bagging, DT, M5-R, DS, M5P, RF, RT and REP-T), in order to find out a forecasting model suitable for the confirmation cases data from different countries COVID-19 epidemic [21]. They are selected by this study as effective techniques of time-series forecasting models, because they have been found to have excellent performance and important research trends from past studies [20-22]. Furthermore, this study collects and uses the confirmed cases data of "different years 2020–2022" all over the world for various experimental datasets, which are divided into different subset data, and then separately evaluate the technical benefits of the proposed model in various sub-datasets under the effective evaluation standard MAPE for time-series data [40]. Importantly and interestingly, this study should be a pioneering and innovative topic on “field application analysis” and a new attempt on a newer model combination on a relatively new field of COVID-19 epidemiological data when compared with the existing researches. Therefore, this study has contributed to a certain extent innovative topic, model combination innovation, and application innovation in the topics of public health field.

3.2 Research Processes of the Proposed Model

The proposed model has the four main stages:

i. Pre-processing
ii. Modelling
iii. Forecasting
iv. Evaluating

The following statements briefly describes the four prediction processing stages for the technique combinations of some advanced forecasting models, describing the processing flow and algorithms of these predictive models for COVID-19 challenge, respectively, as follows:

i. Stage 1 - The Pre-processing Stage: This phase consists of the following three sub-stages.
- Stage 1-1: Select the subject and object data first on COVID-19 issue;
- Stage 1-2: Collect data related to the COVID-19 epidemic indicator (confirmed cases) in 2020–2022 into different sub-experimental datasets;
- Stage 1-3: Sort out the sub-experimental indicator data into CSV format to facilitate the experiments.

ii. Stage 2 - The Construction of Predictive Model Stage: There are two sub-stages to construct a predictive model.
- Stage 2-1: Select the organized forecasting methods from the three ARIMA-based models and 12 machine learning techniques as the proposed model, respectively;
- Stage 2-2: Build the prediction model for finding out the best parameter combination.

iii. Stage 3 - The Creation of a Predictive Value Stage: This stage has two sub-stages.
- Stage 3-1: Mainly input the actual epidemic indicator (confirmed cases) data into the predictive model
- Stage 3-2: To further generate the next value of epidemic prediction data, which can be later compared and analysed with the actual value helpful to the subsequent performance comparison.

iv. Stage 4 - The Performance Assessment Stage: It consists of three sub-stages.
- Stage 4-1: Calculate the evaluation indicator MAPE to make a comparison of experimental results and error analysis;
- Stage 4-2: Conducting in-depth assessments with a post hoc test.
- Stage 4-3: Do a main conclusion from empirical results.

4. Results and Discussion
4.1 Empirical Results of the Proposed Model

Accordingly, we followed several different COVID-19 sub-datasets in practice collected from the public website, with the first period of data scheduled for 2020–2022, and the main prediction feature is the related confirmation cases of countries; thus, the dataset actually includes instances from some original data from global countries and other areas or special districts. Afterwards, we then practically implement the proposed model with these collected datasets. Initially, just from the empirical results of reporting statistical data first in the descriptive statistics, we have three key points to define an important research outcome:

i. We first identify the top 10 countries or districts of the most severity for the confirmation cases by the day of December 31, 2022, and there are United States of America (USA), China, India, France, Germany, Brazil, Japan, Republic of Korea, Italy and The United Kingdom.

ii. Regarding the accumulative confirmation case rate, the top 10 countries, areas, or collectivity territorial unique are San Marino (69.76%), Faroe Islands (65.06%), Austria (64.82%), Slovenia (62.41%), Gibraltar (62.41%), Martinique (62.33%), Brunei Darussalam (61.17%), Andorra (59.75%), Jersey (59.38%) and France (59.37%).
iii. In particular, we are specifically based on the better MAPE of evaluation standard to show the top 10 countries or districts, they are USA, China, India, France, Germany, Brazil, Japan, Republic of Korea, Italy and The United Kingdom.

Furthermore, through the implementing process of "The proposed model" with the collected datasets, different research results were obtained from the different datasets. We summarize and consolidate all outcomes of the given data derived from the experimental results into two parts:

i. One is to show the empirical results of the experiments for the cases of the highlighted importance of parts recorded by the three ARIMA-based models.
ii. Another is the results obtained from using the implementation of 12 supervised predictive algorithms from machine learning techniques in detail described in Section 1.

For the case of USA, the related information for empirical results of the three ARIMA-based models is determined. Figure 1 shows The line chart of the confirmation cases of COVID-19 data.

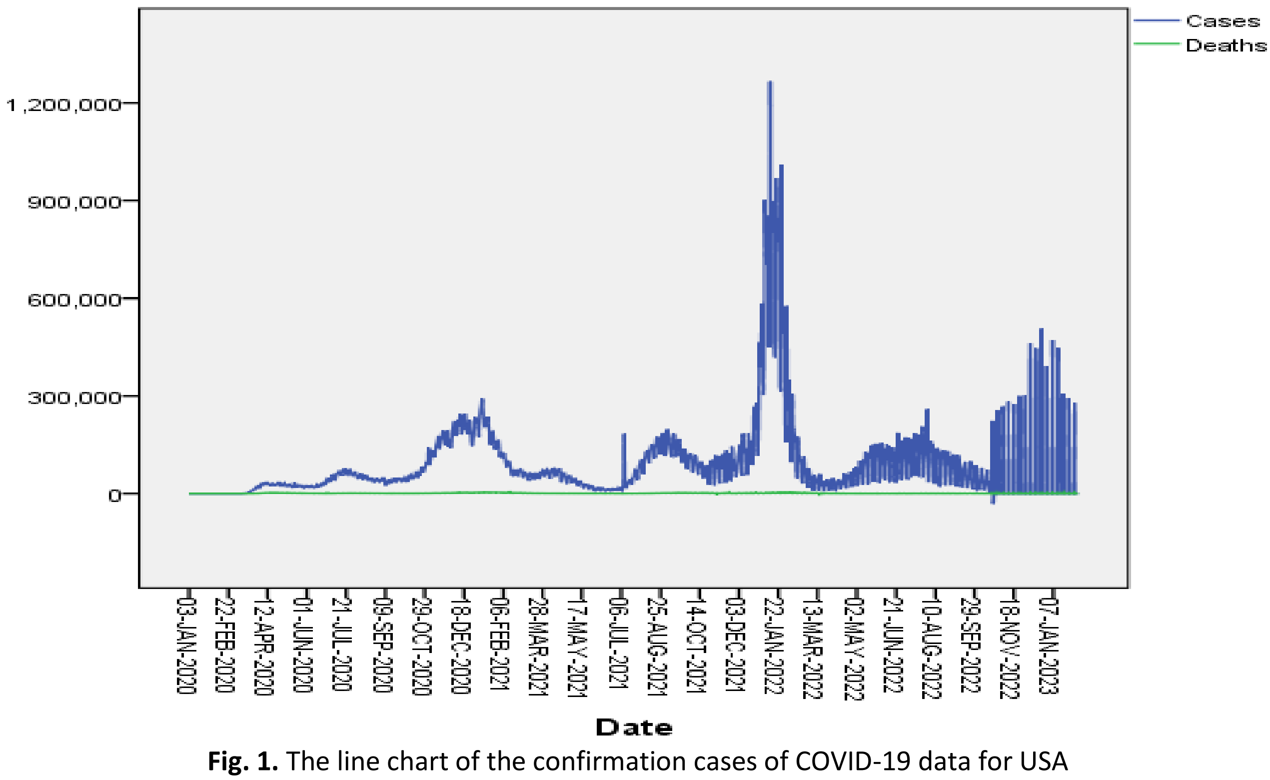

Fig. 1. The line chart of the confirmation cases of COVID-19 data for USA

Table 3 lists Parameters of ARIMA model for the confirmation cases to the most severities of countries.

Table 3
Parameters of ARIMA model for the confirmation cases to the most severities of countries

Country \ ARIMA	Type of best model	Lagging periods of AR	Lagging periods of MA
Brazil	ARIMA(7,1,6)	3 & 7 periods	1 & 2 & 3 & 5 & 6 periods
China	ARIMA(0,1,7)	-	1 & 2 & 3 & 4 & 5 & 7 periods
France	ARIMA(7,1,11)	3 & 5 & 7 periods	1 & 11 periods
Germany	ARIMA(3,0,8)	1 & 2 & 3 periods	1 & 2 & 6 & 7 & 8 periods
India	ARIMA(0,1,8)	-	1 & 3 & 7 & 8 periods
Indonesia	ARIMA(0,1,14)	-	1 & 2 & 4 & 6 & 7 & 14 periods
Italy	ARIMA(0,1,18)	-	1 & 5 & 6 & 7 & 8 & 9 & 18 periods
Japan	ARIMA(0,1,14)	-	1 & 3 & 7 & 8 & 14 periods
Mexico	ARIMA(0,1,7)	-	1 & 2 & 5 & 7 periods
Peru	ARIMA(0,1,14)	-	1 & 2 & 6 & 7 & 9 & 14 periods
Republic of Korea	ARIMA(0,1,7)	-	1 & 2 & 4 & 5 & 7 periods
Russian Federation	ARIMA(0,1,8)	-	1 & 2 & 3 & 4 & 6 & 7 & 8 periods
UK	ARIMA(0,1,7)	-	1 & 3 & 5 & 6 & 7 periods
USA	ARIMA(0,1,8)	-	1 & 7 & 8 periods

Table 4 lists statistical information of ARIMA model for the confirmation cases to the top most severities of countries.

Table 4
Statistical information of ARIMA model for the confirmation cases to the most severities of countries

Country \ ARIMA	MAPE	Statistical data	DF	Significance
Brazil	34.774	56.148	11	0.000***
China	263.194	56.798	12	0.000***
France	46.710	26.287	13	0.016**
Germany	64.031	824.237	10	0.000***
India	18.640	340.589	14	0.000***
Indonesia	20.218	115.671	12	0.000***
Italy	51.002	466.377	11	0.000***
Japan	320.583	80.985	13	0.000***
Mexico	37.904	906.931	14	0.000***
Peru	36.532	91.008	12	0.000***
Republic of Korea	173.192	413.169	13	0.000***
Russian Federation	6.274	456.590	11	0.000***
UK	15.872	386.557	13	0.000***
USA	33.820	565.078	15	0.000***

Note: *, **, and *** refer to different significance levels for 0.1, 0.05 and 0.01, respectively

Regarding the analysis results of Table 3, four directions are defined. First, from Table 3, it is clear that the best parameters of ARIMA (p, d, q) model are achieved, and the lagging periods of AR(p) and MA(q) are also identified. Second, the best parameters can have seven combinations, ARIMA (7,1,6) for Brazil, ARIMA (0,1,7) for China, Mexico, Republic of Korea and UK, ARIMA (7,1,11) for France, ARIMA (3,0,8) for Germany, ARIMA (0,1,8) for India, Russian Federation and USA, ARIMA (0,1,14) for Indonesia, Japan and Peru, and ARIMA (0,1,18) for Italy. Third, it is found that only three countries (Brazil, France and Germany) have lagging periods for AR model. Fourth, for MA model, every country has different lagging periods, and all logging periods are different. As for Table 4, three key points are defined. First, different significance levels are determined, and the lowest MAPE 6.274 on Russian

Federation is defined. The lowest MAPE value indicates a more accurate prediction; that is, for the data of Russian Federation, it is better suitable for using the proposed model as a predictive method to COVID-19 data. Second, for all countries, all they have statistical significance with different significance levels. Finally, the top five lowest MAPE is 6.274 (Russian Federation), 15.872 (UK), 18.640 (India), 20.218 (Indonesia) and 33.820 (USA).

For the empirical results of machine learning techniques, Table 5 shows their information for the top three good performance to the most severities of countries for COVID-19, respectively. From Table 5, it is clear that the RF and RT techniques of machine learning models in order have the top two best performance, and they have an outstanding quality than the other listed models.

Table 5
The top three machine learning techniques for the confirmation cases to the most severities of countries

Country	ARIMA	GP	LR	MLP	SMO-R	Bagging	DT	M5-R	DS	M5P	RF	RT	REP-T
Brazil						V					V	V	
China										V	V	V	
France										V		V	V
Germany					V						V	V	
India					V		V				V		
Indonesia										V	V	V	
Italy				V							V	V	
Japan					V						V	V	
Mexico								V			V	V	
Peru						V					V	V	
Republic of Korea						V					V	V	
Russian Federation						V					V	V	
UK				V							V	V	
USA										V	V	V	
Count		0	0	2	3	4	1	1	0	4	13	13	1

Based on the above empirical results, the propose model in the study has a good prediction performance with a lower error rate MAPE with a better parameter setting in the COVID-19 public health dataset, to find out a suitable model under different conditions of country contexts, which also demonstrates the acceptable performance of this prediction model from a variety of evaluation standard with machine learning techniques.

4.2 The Developments and Applications of this Study

In summary, this study has concreted value and impact on the academic and practical applications for further addressing COVID-19 epidemic indicator (confirmation cases) data from perspectives of intelligent forecasting techniques, and it is expected to pace a small step to the benefits of the study to handle the data analysis for prediction model of epidemiological indicators; therefore, it should contribute considerably to the industrial development and academic research of application fields. The overall contribution and importance of this study are subdivided into four main results of directions for a variety of categories:

i. For the total empirical results: Totally, eight core benefits are determined from the proposed model:

- The best parameters of AR(p), MA(q), and ARIMA (p, d, q) models are identified.
- For ARIMA (p, d, q), seven combinations of the best parameters are also defined.
- For AR model, only three countries need lagging periods.
- For MA model, different countries have different lagging periods.
- For ARIMA model, the lowest MAPE 6.274 on Russian Federation is also identified.
- The proposed model is particularly suitable for the data of Russian Federation.
- All countries have statistical significance with different levels.
- The countries of Russian Federation, UK, India, Indonesia, and USA have the top five lowest MAPE orderly. Moreover, it is found that different predictive techniques for the same country's epidemic data have different prediction error rates; thus, different COVID-19 data of different countries needs different forecasting methods. Importantly, it is a good issue to differentiate from the existing works, the study highlights an excellent organization method for three ARIMA-based time-series models and 12 machine learning techniques for addressing a COVID-19 data application. Clearly, this study has a clear contribution of a technical and applicable innovation on benefiting COVID-19 data fields.

ii. For academic value: The study proposes and combines some intelligent forecasting models to provide forecasting needs of data for focusing COVID-19-related epidemics. This study had a combination benefit of time-series forecasting models for epidemic data analyses, which is a relatively new attempt from the limited literature review; thus, this study is somewhat innovative in the application of methodology, and the empirical results are worth a try and expectation for pacing a small academic milestone.

iii. For practical development: This study is intended to use COVID-19-related epidemic indicator (confirmation cases) data as an experimental object, and it is aimed to combine some predictive models with the data analysis of different countries' epidemiological backgrounds, assuming that a number of experimental data sets can be collected in a timely manner in order to increase the prediction and transparency of the epidemiology indicator data of COVID-19 and reduce the impact degree for the waste of national resources caused by the decision-making error in the predictions of COVID-19 related to the epidemic data application.

iv. For public health applications: Some predictive models and the public health application cases for this data analysis of COVID-19-related epidemic indicator can be applied to different industry sectors and different applied fields or similar situation practices in different countries, serving as an excellent reference tool when facing prediction problems at different levels in similar industries or across industries.

5. Conclusions

This study mainly develops a variety of different time-series forecasting models and machine learning techniques for the COVID-19 data actually collected from different countries. This study is mainly focused on the establishment of the enhanced version of the proposed model, followed by an explanation of its data analysis results and discussions. The specific contribution to implement the use of integrated techniques in the study can be divided into several aspects, as follows:

i. Establish a variety of intelligent time-series forecasting models and machine learning techniques to execute predictions of COVID-19-related outbreak data as a helpful reference base for predicting future outbreak indicator, and it is possible to derivatively forecast other industry applications or industry needs to other countries around the world in the future.
ii. Provide an objective model assessment method to validate the corresponding performance differences, thereby enhancing the performance of intelligent forecasting models.
iii. Provide an objective evaluation standard for assessing the performance of predictions for COVID-19-related outbreak indicator data from different countries.
iv. Provide an intelligent forecasting model that applies COVID-19-related outbreak indicator data as a tool for mining this forecasting knowledge for different interest considerations for the interested parties.

Until now, the study has completed the empirical experiments of models in using the practical COVID-19 data evaluation experiments, and the study results for data analysis have showed that the proposed model has an acceptable performance. Although the study has a good empirical result, it has still a room to further improve the quality of the proposed model. Thus, there are two core ways to further well do it in the future research. First, it is expected that more different models, such as Generalized Method of Moments (GMM) model [41], and more other machine learning algorithms [42,43] or a typical model of systematic review and meta-analysis [44,45] is more smoothly carried out, and better research results can be achieved. Second, it is also hoped and then searched to analyse from multi-perspectives of different time-series forecasting models, such as deep multivariate time-series forecasting [46] and fuzzy time-series forecasting model [47] for further addressing different time interval datasets.

Acknowledgment
The authors would wish to sincerely thank the National Science and Technology Council of Taiwan on the grant number NSTC 111-2221-E-167-036-MY2 for the financial support of this research.

References
[1] MacDonald, Robert, Hannah King, Emma Murphy, and Wendy Gill. "The COVID-19 pandemic and youth in recent, historical perspective: more pressure, more precarity." Journal of Youth Studies 27, no. 5 (2024): 723-740. https://doi.org/10.1080/13676261.2022.2163884
[2] Miller, Amalia R., Carmit Segal, and Melissa K. Spencer. "Effects of the COVID-19 pandemic on domestic violence in Los Angeles." Economica 91, no. 361 (2024): 163-187. https://doi.org/10.1111/ecca.12493
[3] Lun, Low Han, Adam Daniel Effendi, Dzeti Farhah Mohshim, Nabilla Afzan Abdul Aziz, Noreen Izza Arshad, and Mitra Mohd Addi. "Readiness and Effectiveness of Synchronous Online Teaching and Learning in Higher Education during COVID-19 Pandemic." Journal of Advanced Research in Computing and Applications 20, no. 1 (2020): 1-11.
[4] Haussner, William, Antonio P. DeRosa, Danielle Haussner, Jacqueline Tran, Jane Torres-Lavoro, Jonathan Kamler, and Kaushal Shah. "COVID-19 associated myocarditis: a systematic review." The American journal of emergency medicine 51 (2022): 150-155. https://doi.org/10.1016/j.ajem.2021.10.001
[5] Jackson, William M., Jerri C. Price, Lisa Eisler, Lena S. Sun, and Jennifer J. Lee. "COVID-19 in pediatric patients: a systematic review." Journal of Neurosurgical Anesthesiology 34, no. 1 (2022): 141-147. https://doi.org/10.1097/ANA.0000000000000803
[6] Alkrimi, Jameela Ali, Raja Salih Mohammed, Ameer Hamdi Hakeem Al-Ameedee, Safaa Hakeem Alkahfaji, Ban Alwash, and Rawan Al-Rubaye. "Measures of Effectiveness for E learning of University Students During the Covid-19 Pandemic Using the Statistical Model." International Journal of Scientific Research in Network Security and Communication 11, no. 3 (2023): 1-7.
[7] Ardolino, Marco, Andrea Bacchetti, Alexandre Dolgui, Guglielmo Franchini, Dmitry Ivanov, and Anand Nair. "The impacts of digital technologies on coping with the COVID-19 pandemic in the manufacturing industry: a systematic literature review." International Journal of Production Research 62, no. 5 (2024): 1953-1976. https://doi.org/10.1080/00207543.2022.2127960

[8] Woldometers. “Report coronavirus cases of Worldometer website.” https://www.worldometers.info/coronavirus
[9] Aga, Beza Zeleke, Temesgen Duressa Keno, Debela Etefa Terfasa, and Hailay Weldegiorgis Berhe. "Pneumonia and COVID-19 co-infection modeling with optimal control analysis." Frontiers in Applied Mathematics and Statistics 9 (2024): 1286914. https://doi.org/10.3389/fams.2023.1286914
[10] Fonseca Lima, Eduardo Jorge, Luiza Campos Corrêa de Araújo, Karine Ferreira Agra, Ana Julia Xavier Mendoza, Julia Pierre de Brito Siebra, and Carmina Silva Dos Santos. "Analysis of Childhood Pneumonia: A Comparison Between the Pre-and During the COVID-19 Pandemic in a Reference Hospital in Brazil." Pediatric Health, Medicine and Therapeutics (2024): 103-110. https://doi.org/10.2147/PHMT.S451735
[11] Sakuramoto, Kazuhito, Daiki Wada, Shuhei Maruyama, Takashi Muroya, Fukuki Saito, Yasushi Nakamori, and Yasuyuki Kuwagata. "Evaluation of characteristics and prognosis of COVID-19 patients requiring invasive mechanical ventilation during dominance of nonvariant, alpha, delta, and omicron variants in tertiary hospitals of Japan." BMC Infectious Diseases 24, no. 1 (2024): 223. https://doi.org/10.1186/s12879-024-09131-4
[12] Scott, Hayley, Aleena Zahra, Rafael Fernandes, Bettina C. Fries, Henry C. Thode Jr, and Adam J. Singer. "Bacterial infections and death among patients with Covid-19 versus non Covid-19 patients with pneumonia." The American journal of emergency medicine 51 (2022): 1-5. https://doi.org/10.1016/j.ajem.2021.09.040
[13] Chhabra, Anureet, Sunil K. Singh, Akash Sharma, Sudhakar Kumar, Brij B. Gupta, Varsha Arya, and Kwok Tai Chui. "Sustainable and intelligent time-series models for epidemic disease forecasting and analysis." Sustainable Technology and Entrepreneurship 3, no. 2 (2024): 100064. https://doi.org/10.1016/j.stae.2023.100064
[14] Huang, Siyuan, Yepeng Liu, Fan Zhang, Yue Li, Jinjiang Li, and Caiming Zhang. "Crosswavenet: A dual-channel network with deep cross-decomposition for long-term time series forecasting." Expert Systems with Applications 238 (2024): 121642. https://doi.org/10.1016/j.eswa.2023.121642
[15] Wu, Yuhan, Xiyu Meng, Junru Zhang, Yang He, Joseph A. Romo, Yabo Dong, and Dongming Lu. "Effective LSTMs with seasonal-trend decomposition and adaptive learning and niching-based backtracking search algorithm for time series forecasting." Expert Systems with Applications 236 (2024): 121202. https://doi.org/10.1016/j.eswa.2023.121202
[16] Liu, Dandan, and Hanlin Wang. "Time series analysis model for forecasting unsteady electric load in buildings." Energy and Built Environment 5, no. 6 (2024): 900-910. https://doi.org/10.1016/j.enbenv.2023.07.003
[17] Petropoulos, Fotios, and Ivan Svetunkov. "A simple combination of univariate models." International journal of forecasting 36, no. 1 (2020): 110-115. https://doi.org/10.1016/j.ijforecast.2019.01.006
[18] Islam, M. A., Hang Seng Che, M. Hasanuzzaman, and N. A. Rahim. "Energy demand forecasting." In Energy for sustainable development, pp. 105-123. Academic Press, 2020. https://doi.org/10.1016/B978-0-12-814645-3.00005-5
[19] Khan, Rehan Ullah, Waleed Albattah, Suliman Aladhadh, and Shabana Habib. "Learning Patterns from COVID-19 Instances." Computer Systems Science & Engineering 40, no. 2 (2022). https://doi.org/10.32604/csse.2022.019757
[20] Zheng, Jianqin, Haoran Zhang, Yuanhao Dai, Bohong Wang, Taicheng Zheng, Qi Liao, Yongtu Liang, Fengwei Zhang, and Xuan Song. "Time series prediction for output of multi-region solar power plants." Applied Energy 257 (2020): 114001. https://doi.org/10.1016/j.apenergy.2019.114001
[21] Anderson, Roy M., Hans Heesterbeek, Don Klinkenberg, and T. Déirdre Hollingsworth. "How will country-based mitigation measures influence the course of the COVID-19 epidemic?." The lancet 395, no. 10228 (2020): 931-934. https://doi.org/10.1016/S0140-6736(20)30567-5
[22] Hadwan, Mohammad, Basheer M. Al-Maqaleh, Fuad N. Al-Badani, Rehan Ullah Khan, and Mohammed A. Al-Hagery. "A Hybrid Neural Network and Box-Jenkins Models for Time Series Forecasting." Computers, Materials & Continua 70, no. 3 (2022). https://doi.org/10.32604/cmc.2022.017824
[23] Obisesan, Omodara E. "Machine Learning Models for Prediction of Meteorological Variables for Weather Forecasting." Int J Environ Clim Change 14, no. 1 (2024): 234-52. https://doi.org/10.9734/ijecc/2024/v14i13829
[24] Abbasimehr, Hossein, Ali Noshad, and Reza Paki. "A novel featurization methodology using JaGen algorithm for time series forecasting with deep learning techniques." Expert Systems with Applications 235 (2024): 121279. https://doi.org/10.1016/j.eswa.2023.121279
[25] Das, Kaushik, Roushan Kumar, and Anurup Krishna. "Analyzing electric vehicle battery health performance using supervised machine learning." Renewable and Sustainable Energy Reviews 189 (2024): 113967. https://doi.org/10.1016/j.rser.2023.113967
[26] Saeed, Nadia, and Nismat Javed. "Lessons from the COVID-19 pandemic: Perspectives of medical students." Pakistan journal of medical sciences 37, no. 5 (2021): 1402. https://doi.org/10.12669/pjms.37.5.4177
[27] Fauci, Anthony S., H. Clifford Lane, and Robert R. Redfield. "Covid-19—navigating the uncharted." New England Journal of Medicine 382, no. 13 (2020): 1268-1269. https://doi.org/10.1056/NEJMe2002387

[28] Lauer, Stephen A., Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, and Justin Lessler. "The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application." Annals of internal medicine 172, no. 9 (2020): 577-582. https://doi.org/10.7326/M20-0504
[29] Jiang, Fang, Liehua Deng, Liangqing Zhang, Yin Cai, Chi Wai Cheung, and Zhengyuan Xia. "Review of the clinical characteristics of coronavirus disease 2019 (COVID-19)." Journal of general internal medicine 35 (2020): 1545-1549. https://doi.org/10.1007/s11606-020-05762-w
[30] Xu, Zhe, Lei Shi, Yijin Wang, Jiyuan Zhang, Lei Huang, Chao Zhang, Shuhong Liu *et al.,* "Pathological findings of COVID-19 associated with acute respiratory distress syndrome." The Lancet respiratory medicine 8, no. 4 (2020): 420-422. https://doi.org/10.1016/S2213-2600(20)30076-X
[31] Zheng, Ying-Ying, Yi-Tong Ma, Jin-Ying Zhang, and Xiang Xie. "COVID-19 and the cardiovascular system." Nature reviews cardiology 17, no. 5 (2020): 259-260. https://doi.org/10.1038/s41569-020-0360-5
[32] Heymann, David L., and Nahoko Shindo. "COVID-19: what is next for public health?." The lancet 395, no. 10224 (2020): 542-545. https://doi.org/10.1016/S0140-6736(20)30374-3
[33] Rothan, Hussin A., and Siddappa N. Byrareddy. "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak." Journal of autoimmunity 109 (2020): 102433. https://doi.org/10.1016/j.jaut.2020.102433
[34] Box, George EP, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. Time series analysis: forecasting and control. John Wiley & Sons, 2015.
[35] Ee, Jonathan Yong Chung, Jin Yuan Chan, and Gan Lik Kang. "Carbon reduction analysis of Malaysian green port operation." Progress in Energy and Environment (2021): 1-7.
[36] Yan, Chao, Yankun Zhang, Weiyi Zhong, Can Zhang, and Baogui Xin. "A truncated SVD-based ARIMA model for multiple QoS prediction in mobile edge computing." Tsinghua Science and Technology 27, no. 2 (2021): 315-324. https://doi.org/10.26599/TST.2021.9010040
[37] Alsharif, Mohammed H., Mohammad K. Younes, and Jeong Kim. "Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea." Symmetry 11, no. 2 (2019): 240. https://doi.org/10.3390/sym11020240
[38] Ospina, Raydonal, João AM Gondim, Víctor Leiva, and Cecilia Castro. "An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil." Mathematics 11, no. 14 (2023): 3069. https://doi.org/10.3390/math11143069
[39] Mohamed, Mohamed Ali, Ibrahim Mahmoud El-Henawy, and Ahmad Salah. "Price Prediction of Seasonal Items Using Machine Learning and Statistical Methods." Computers, Materials & Continua 70, no. 2 (2022). https://doi.org/10.32604/cmc.2022.020782
[40] Dhahad, Hayder Abed, Ahmed Mudheher Hasan, Miqdam Tariq Chaichan, and Hussein A. Kazem. "Prognostic of diesel engine emissions and performance based on an intelligent technique for nanoparticle additives." Energy 238 (2022): 121855. https://doi.org/10.1016/j.energy.2021.121855
[41] Zheng, Qikang, Fariya Sharmeen, Chengcheng Xu, and Pan Liu. "Assessing regional road traffic safety in Sweden through dynamic panel data analysis: Influence of the planned innovative policies and the unplanned COVID-19 pandemic." Transportation Research Part A: Policy and Practice 179 (2024): 103918. https://doi.org/10.1016/j.tra.2023.103918
[42] Husnain, Ali, Hafiz Khawar Hussain, Hafiz Muhammad Shahroz, Muhammad Ali, Ahmed Gill, and Saad Rasool. "Exploring ai and machine learning applications in tackling covid-19 challenges." Revista Espanola de Documentacion Cientifica 18, no. 02 (2024): 19-40.
[43] Iwendi, Celestine, C. G. Y. Huescas, Chinmay Chakraborty, and Senthilkumar Mohan. "COVID-19 health analysis and prediction using machine learning algorithms for Mexico and Brazil patients." Journal of Experimental & Theoretical Artificial Intelligence 36, no. 3 (2024): 315-335.
[44] Gioia, Francesca, Laura N. Walti, Ani Orchanian-Cheff, and Shahid Husain. "Risk factors for COVID-19-associated pulmonary aspergillosis: a systematic review and meta-analysis." The Lancet Respiratory Medicine 12, no. 3 (2024): 207-216. https://doi.org/10.1016/S2213-2600(23)00408-3
[45] Özgüç, Safiye, Emine Kaplan Serin, and Derya Tanriverdi. "Death anxiety associated with coronavirus (COVID-19) disease: A systematic review and meta-analysis." OMEGA-Journal of Death and Dying 88, no. 3 (2024): 823-856. https://doi.org/10.1177/00302228211050503
[46] Ye, Junchen, Weimiao Li, Zhixin Zhang, Tongyu Zhu, Leilei Sun, and Bowen Du. "MvTS-library: An open library for deep multivariate time series forecasting." Knowledge-Based Systems 283 (2024): 111170. https://doi.org/10.1016/j.knosys.2023.111170
[47] Shi, Xinjie, Jianzhou Wang, and Bochen Zhang. "A fuzzy time series forecasting model with both accuracy and interpretability is used to forecast wind power." Applied Energy 353 (2024): 122015. https://doi.org/10.1016/j.apenergy.2023.122015