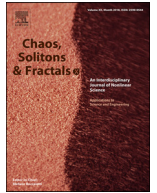**ELSEVIER**

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries

Check for updates

Sarbjit Singh [a,c], Kulwinder Singh Parmar [b,*], Sidhu Jitendra Singh Makkhan [d,e], Jatinder Kaur [b,f], Shruti Peshoria [g], Jatinder Kumar [c]

[a] Guru Nanak Dev University College, Narot Jaimal Singh, Pathankot, Punjab, India
[b] Department of Mathematics, I.K. Gujral Punjab Technical University, Punjab, India
[c] Department of Mathematics, Guru Nanak Dev University, Amritsar, Punjab, India
[d] Department of Mathematics, Sri Guru Angad Dev College, Khadoor Sahib, Tarn Taran, Punjab, India
[e] Department of Mathematics, Lovely Professional University, Punjab, India
[f] Guru Nanak Dev University College, Verka, Amritsar, Punjab, India
[g] Centre for Fire, Explosive and Environment Safety (CFEES), Defence Research and Development Organisation (DRDO), Timarpur, Delhi 110054, India

ARTICLE INFO

ABSTRACT

Discussions about the recently identified deadly coronavirus disease (COVID-19) which originated in Wuhan, China in December 2019 are common around the globe now. This is an infectious and even life-threatening disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It has rapidly spread to other countries from its originating place infecting millions of people globally. To understand future phenomena, strong mathematical models are required with the least prediction errors. In the present study, autoregressive integrated moving average (ARIMA) and least square support vector machine (LS-SVM) models are applied to the data consisting of daily confirmed cases of SARS-CoV-2 in the most affected five countries of the world for modeling and predicting one-month confirmed cases of this disease. To validate these models, the prediction results were tested by comparing it with testing data. The results revealed better accuracy of the LS-SVM model over the ARIMA model and also suggested a rapid rise of SARS-CoV-2 confirmed cases in all the countries under study. This analysis would help governments to take necessary actions in advance associated with the preparation of isolation wards, availability of medicines and medical staff, a decision on lockdown, training of volunteers, and economic plans.

1. Introduction

The world was running very smoothly, and all countries were engaged in improving their gross domestic product (GDP). Suddenly, the novel coronavirus disease (Covid-19) knocked Wuhan, China, in December 2019, which was viewed as the start of an epidemic however it overwhelmed the entire world in a very short span of time and took the form of a pandemic. Covid-19 is highly contagious in nature and is caused by the virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [14,19,34,35,43,48,55]. In the starting phase, it was considered as the spread of viral pneumonia, linked to the exposure to the seafood market [46,51,53]. On January 7, 2020, the virus was isolated, and it confirmed the identity of 2019-nCoV (novel coronavirus) by the Chinese health agencies, which on February 11, 2020, was termed as COVID-19 by the World Health Organization [43]. Till date, COVID-19 has led to the death of 2,86,353 individuals around the globe and affected more than 41,78,156 people as on May 12, 2020, reported by the World Health Organization [43]. Deaths caused by COVID-19 are expected to increase far more than the previous two coronaviruses, namely SARS-CoV and MERS-CoV. [10,24].

The emergency situation caused due to the outbreak of a new infection raises many important questions related to its transmission dynamics, mitigation, and control measures. Researchers are taking the help of mathematical modeling in order to provide answers to such urgent queries [7]. As the medicine is not yet discovered, Governments make efforts to control situations by instructing strategies such as social distancing, quarantine and
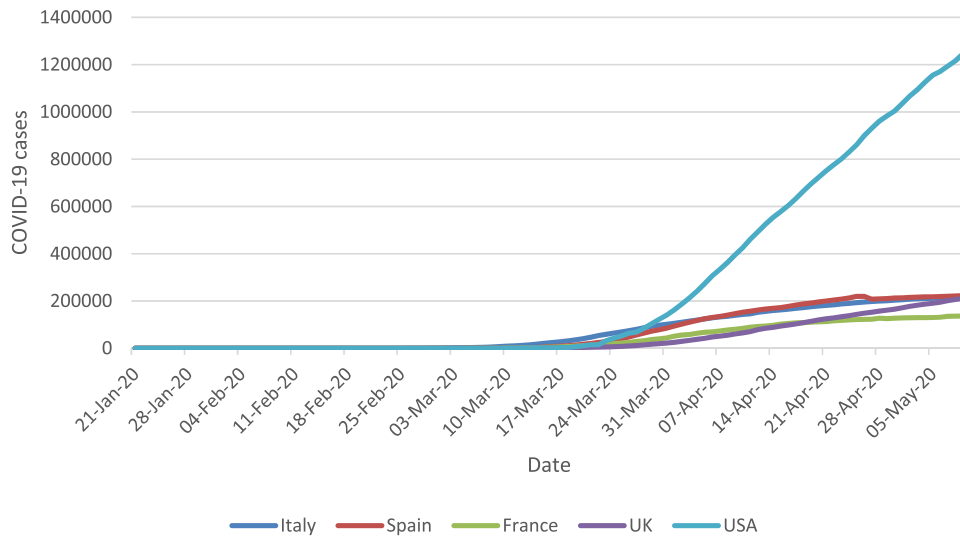
Fig. 1. Time series plot of COVID-19 cases from Five countries.

contact tracing of the infected or suspected people, complete lock-down of the area and impose international/domestic travel restrictions [8,13,32,49]. The stochastic transmission model told about the variation of COVID-19 over a certain period of time, and also discussed the declining trend in Wuhan by restricting international/domestic travel, local transport, etc., [16,32]. Many different models observed the behavior of COVID-19 such as the virus's transmission route being from person to person, and doubling time of 7.4 days [1,3,20,22,46]. To study the stochastic behavior, it demands the consistent data analysis of the COVID-19 data.

Time series analysis deals with the understanding of the dependence among variables by using various modeling techniques. It provides the ultimate goal of obtaining some information about past observations to predict future values. ARIMA (Autoregressive Integrated Moving Average) model is based upon Box-Jenkins algorithm to estimate the values of *p, d* and *q* which is widely used in many areas of research to forecast the future phenomena [26,29,36–39,41]. ARIMA time series models work based on past data and assume the normality of residuals, so these models are not appropriate to study the non-linear and complex phenomena [4,15]. To overcome the limitations of ARIMA models, kernel-based models are gaining special attention and are being extensively used by most of the researchers. Support vector regression (SVR) and least square support vector machine (LS-SVM) are two widely used kernel-based models which have more tendency to generalize and attain global minima for sample data (Zhou, [27,33,42,47]). LS-SVM is an improved version of SVM which replaces inequality constraints in SVM with equality constraints and uses square errors to replace non-negative errors so that the optimization problem reduces to a linear system of equations. So, kernel-based modeling and forecasting can prove to be a vital tool to deal with such situation by studying its potential of transmission and growth of the virus in the long run [6,25,46,52].

The present study considers modeling and forecasting COVID-19 confirmed cases using autoregressive integrated moving average (ARIMA) and least square support vector machine (LS-SVM) models, which would help in understanding the dynamic nature of the virus transmission. This novel research will help the most affected countries to be well prepared for the medical arrangements, isolation wards, extension of lock-down, recruit new medical staff/ train volunteers, etc. For this, we have considered the daily data obtained from WHO to model the COVID-19 confirmed cases in most affected five countries of the world, namely Italy, Spain, France, the United Kingdom (UK), and the United States of America (USA). The rest of the paper is organized in the following sections: Section 2 contains data and methodology. The applications of the proposed methodology and the results obtained are discussed in Section 3. The final Section 4 contains the conclusions of the study.

2. Data and methodology

For the present study, a dataset consisting of COVID-19 confirmed cases from five most affected countries of the world namely Italy, Spain, France, the United Kingdom, United States of America is used to predict one-month confirmed cases from these countries (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports). The dataset ranges from 21st January 2020 to 09th May 2020, consisting of 110 sample points out of which 88 (*80% of data*) sample points are kept for modeling purposes and the rest for validating the model. A time series plot of COVID-19 confirmed cases from the five countries is shown in Fig. 1. The data modeling techniques ARIMA and LS-SVM are applied in this study to predict one month out of sample data points, which will give an estimate of an increase in the confirmed COVID-19 cases in the future. It will help the governments to take preventive measures to curb the immediate spread of the disease.

2.1. Least square support vector machine

The support vector machine (SVM) developed by Vladimir Vapnik and co-workers at AT&T Bell Laboratories in 1995 is a supervised learning technique based on the principle of Structural Risk Minimization (SRM). Since the SVM model has a tendency to deal with dynamic, nonlinear, and complex time-series data, therefore it is applicable to signal processing, pattern recognition, and non-linear regression [9,17,18]. The dominant methodology of least square support vector machine (LS-SVM) has been derived from SVM, which is used to approximate accurate non-linear relationship between input and output variables and is quite helpful in solving problems related to non-linear classification and regression.

Given an input data vector $X = \{r_k\} \in R^d$ and output data vector $Y = \{s_k\} \in R$ which are combined to form the training dataset $\{(r_1, s_1), (r_2, s_2), ..., (r_N, s_N)\}$ with $s_k \in \{-1, +1\}$ as class labels and $Y(X) = sign[w^T X + b]$ as the linear classifier, where N is the number of samples, b is the bias and w is the weight vector. The separability of training dataset into two classes can be implied as fol-
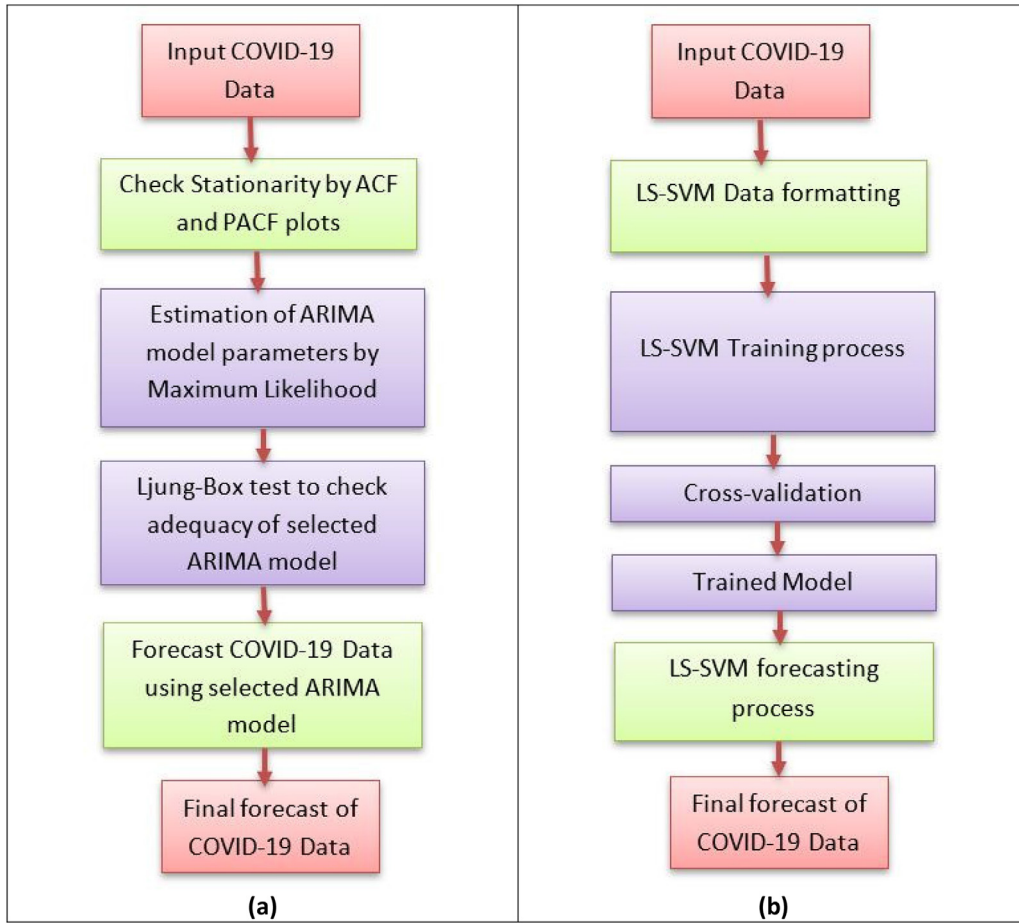
Fig. 2. Flow charts for modeling and forecasting of COVID-19 confirmed cases data using (a) ARIMA model (b) LS-SVM model.

lows:

$$\begin{cases} w^T r_k + b \geq +1, & if \; s_k = +1 \\ w^T r_k + b \leq -1, & if \; s_k = -1 \end{cases}$$

The joint inequality of these two inequalities can be written as:

$$Y \left[w^T X + b \right] \geq 1$$

The methodology of convex optimization theory is behind SVM formulations, which formulate the problem as a constrained optimization problem by constructing Lagrangian using optimality conditions. The problem is finally solved in the dual space of Lagrange multipliers. With resulting classifier

$$s(r) = sign \left[\sum_{k=1}^{N} \alpha_k s_k r_k^T r + b \right]$$

This linear SVM classifier takes into account an additional slack variable while formulating the problem which modifies the set of inequalities into [9]:

$$Y \left[w^T X + b \right] \geq 1 - \xi, \; where \; \xi = \{\xi_1, \xi_2, ..., \xi_N \}$$

The SVM formulation can be used both for linear and non-linear function estimation. As in the classic SVM approach, the least square version of SVM taken into consideration the equality type constraints instead of inequality constraints [40]. This reformulation of solving the LS-SVM problem reduces the lengthy task of solving complex quadratic programs to a set of solutions to linear equations. For LS-SVM classifier, in the primal space, it takes

the form,

$$s(r) = sign \left[w^T r + b \right]$$

where b is a real constant. For non-linear classification, the LS-SVM classifier in the dual space takes the form,

$$s(r) = sign \left[\sum_{k=1}^{N} \alpha_k s_k K(r, r_k) + b \right]$$

where $\alpha_1, \alpha_2, ..., \alpha_N$ are positive real constants, and b is a real constant. In general, $K(r, r_k) = \langle \varphi(r_k), \varphi(r) \rangle$ denotes a kernel function satisfying Mercer's condition, where $\langle \cdot, \cdot \rangle$ is the inner product and $\varphi(r)$ is the non-linear map from original space to higher-dimensional space. The LS-SVM model for functional estimation takes the form

$$s(r) = \sum_{k=1}^{N} \alpha_k s_k K(r, r_k) + b$$

Some popular kernel functions are given by:

a) linear kernel: $K(r, r_k) = r^T r_k$
b) polynomial kernel: $K(r, r_k) = (r^T r_k + 1)^d$, where d denotes the degree of the polynomial kernel.
c) radial basis function (RBF) kernel: $K(r, r_k) = \exp(-\frac{\|r - r_k^2\|}{\sigma^2})$, where σ^2 is a tuning parameter.

The structure of high dimensional feature space $\varphi(r)$ is dependent on kernel function parameters which affect the accuracy of the final solution. The working of the LS-SVM model has been briefly described in the flowchart given in Fig. 2(b).
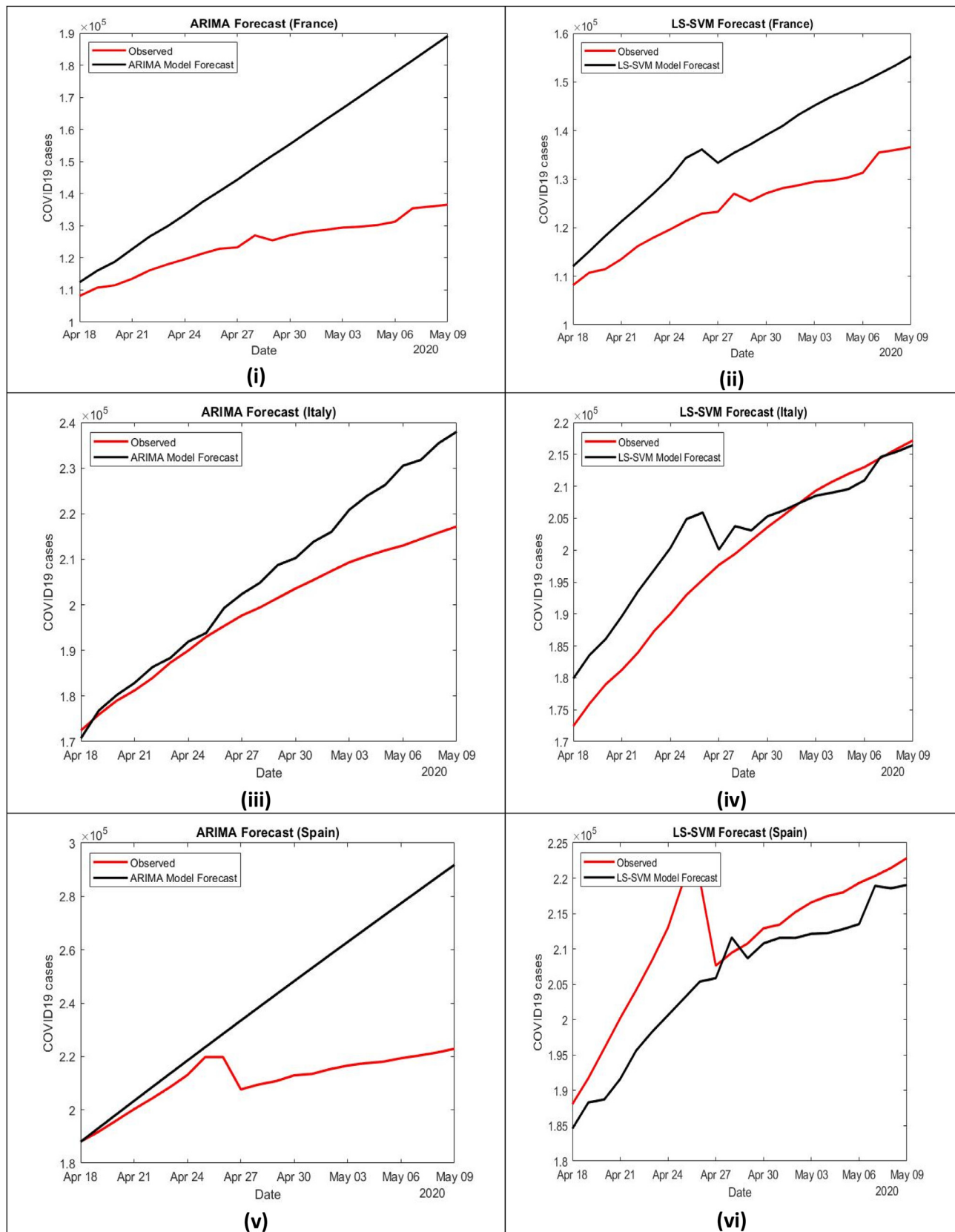
Fig. 3. Comparison of ARIMA and LS-SVM models forecast of COVID-19 confirmed cases for the testing phase of the countries (i & ii) France (iii & iv) Italy (v & vi) Spain (vii & viii) UK and (ix & x) USA.
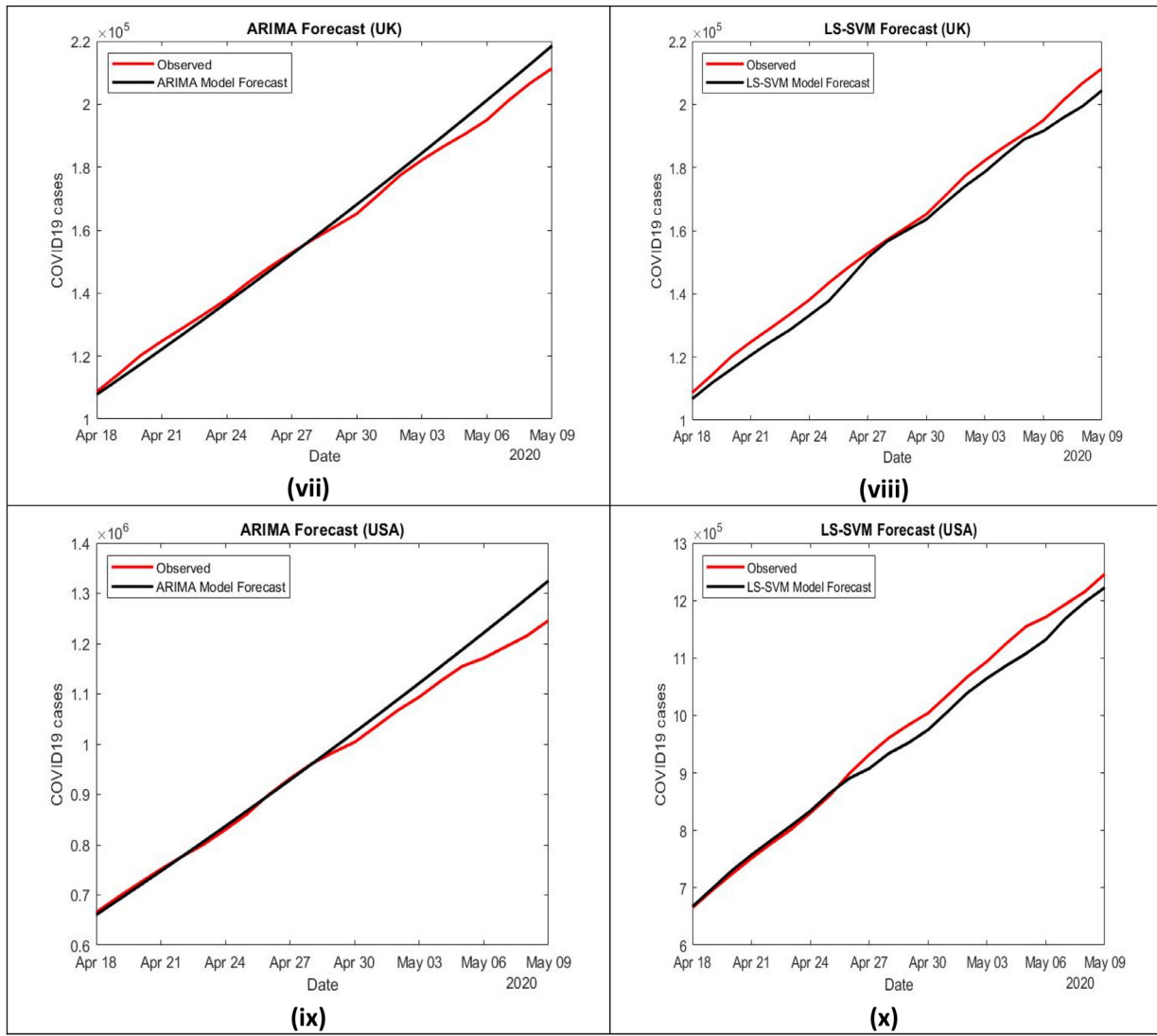
Fig. 3. Continued

2.2. Autoregressive integrated moving average (ARIMA) model

Among the various time series models used in the field of mathematical forecasting, Box Jenkins methodology based ARIMA (*p, d, q*) model is the most popular and widely used time series model due to its simplicity in application and capability of handling non-stationary data. It is an extension of the ARMA (*p, q*) model which is formed by the combination of the Auto-Regressive (AR (*p*)) and Moving Average (MA (*q*)) time series models along with the differencing parameter *d* which is used to convert a non-stationary time series into a stationary series. Differencing of the first order is done by finding the difference between the current and the previous data values, the further second-order difference is obtained by finding the difference among the first-order differences and this process continues, it is also termed as finite difference technique [2,29–31,37].

Mathematically it is given as,

$$x_t = u_t + \sum_{i=1}^{p} \varphi_i x_{t-i} - \varepsilon_t - \sum_{j=1}^{q} \vartheta_j \varepsilon_{t-j}$$

where x_t and x_{t-i} refers to the predicted and the observed values of the time series. p and q refer to the order of the AR and MA

model. ε_t represents the error terms, φ_i and ϑ_j are the model parameters [54].

ARIMA modeling of a time-series data makes use of three steps main steps viz, model identification, parameter estimation, diagnostic checking. For identification, if needed, suitable differencing of the time series is done in order to obtain normality and stationarity. Then the values of the parameters p and q are to be identified, which are examined with the help of autocorrelation (ACF) and partial autocorrelation (PACF) functions plots [50]. The autocorrelation function measures the amount of relationship between the actual data values with its lagged values at time t while the partial autocorrelation measures the amount of correlation between a present value with the lagged value after controlling and removing all the linear dependence between the low order lags values. Finally, the adequacy of the model is verified with the help of the Ljung-Box test so that the further modeling of data series is required. In the case of the inadequacy of the selected model, all three steps of the model building process are repeated until an adequate model is obtained. Thus, appropriately selected ARIMA models are used for predicting the future values of time series data [5,11]. ARIMA model building and forecasting procedure have been depicted in the flow diagram given in Fig. 2(a).
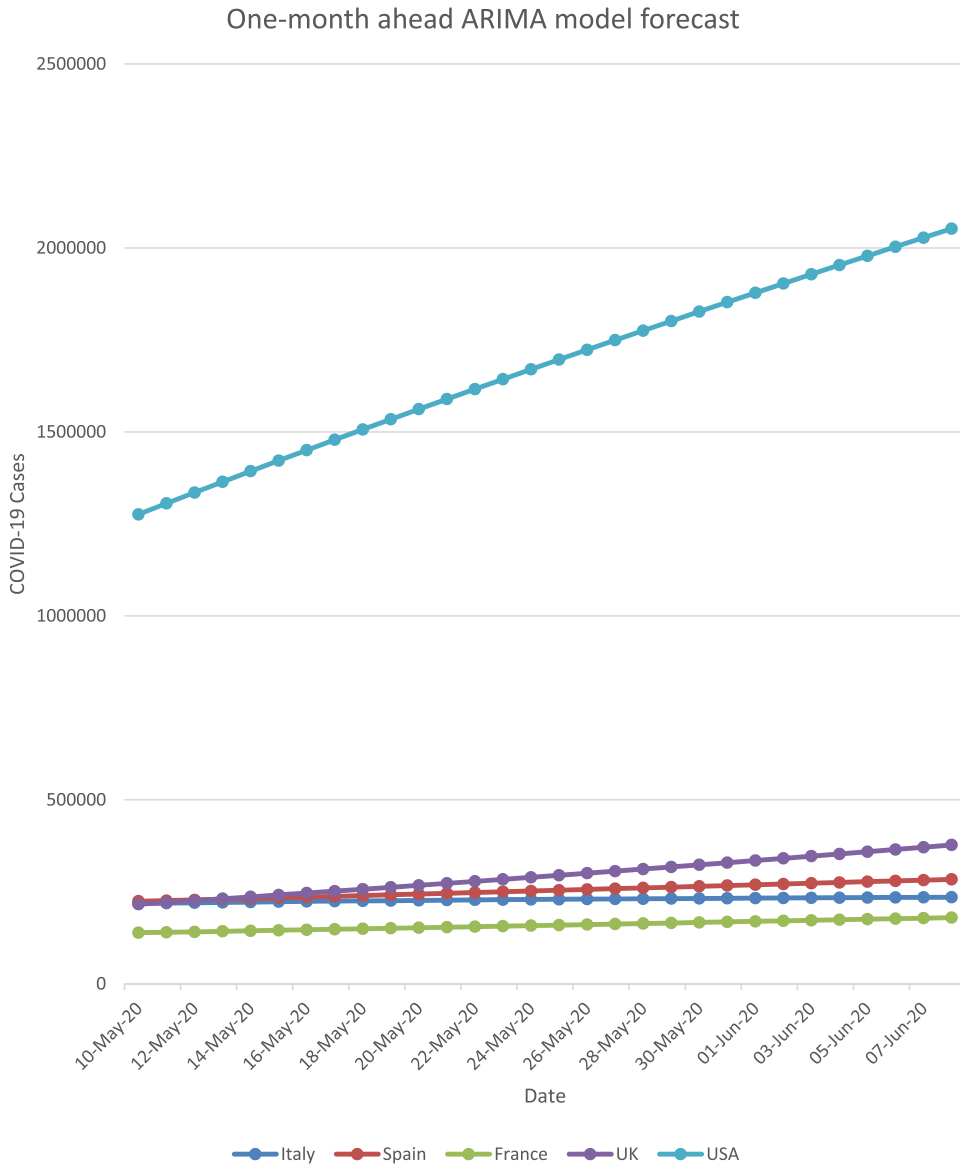
One-month ahead ARIMA model forecast

Fig. 4. One month ahead of COVID-19 affected cases in different countries by the ARIMA model.

3. Application and results

In the present study, COVID-19 confirmed cases from five countries, Italy, Spain, France, the United Kingdom, and the United States of America, were used as input for modeling and forecasting. Mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and coefficient of determination (R^2) were used as evaluation criteria for validating model outputs obtained by ARIMA and LSSVM models [23]. The MAE, MSE, and RMSE are given by

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |f_{i,o} - f_{i,e}|$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (f_{i,o} - f_{i,e})^2$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (f_{i,o} - f_{i,e})^2}$$

where N is the number of data points, $f_{i,o}$ and $f_{i,e}$ are the observed values and model's estimated values respectively of COVID-19 affected cases.

The input data of COVID-19 confirmed cases of five countries is first tested for stationarity by using autocorrelation function (ACF) and partial autocorrelation (PACF) plots. In the case of non-stationarity, appropriate differencing transformation is applied to make data stationary. ARIMA model parameters p and q are estimated by regenerating ACF and PACF plots of stationary data obtained by the differencing process. Maximum likelihood is another commonly used method for parameter estimation. After estimating the order of parameters of an ARIMA model, the overall adequacy of the selected model is finally checked by the Ljung-Box test so that no further modeling is required. After fitting input data with an appropriate ARIMA model, the predictions are made and compared with the testing data set to validate the model.

LSSVM methodology discussed in Section 2.1 is used to train, validate, and predict one-month confirmed cases of COVID-19 using input data of confirmed cases in most affected five countries of the world. In carrying out LS-SVM methodology, the dataset of 110 COVID-19 confirmed cases from five countries divided into two
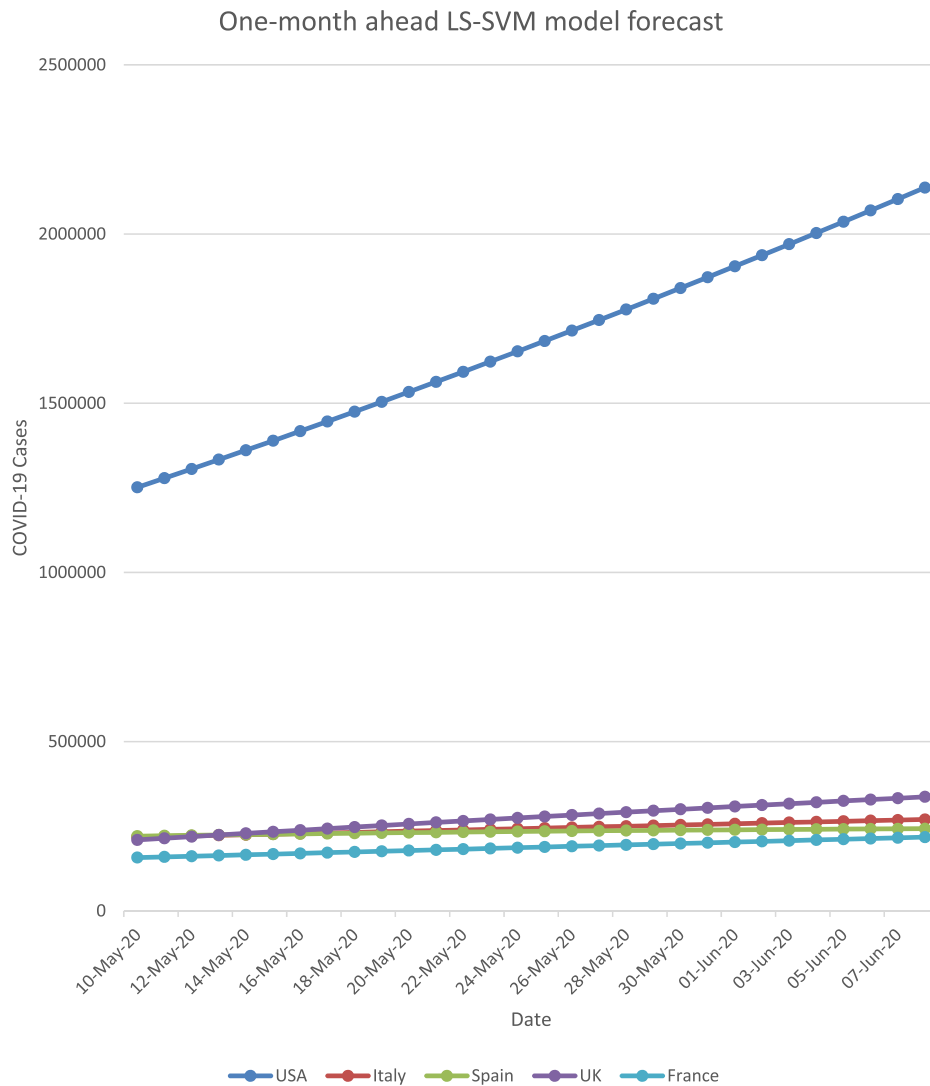
One-month ahead LS-SVM model forecast

Fig. 5. One month ahead, COVID-19 affected cases in different countries by LS-SVM.

Table 1
Prediction performance of an ARIMA model.

Country	MAE (x10³)	MSE(x10⁸)	RMSE(x10³)	R^2
Italy	7.773	1.029	10.147	0.9817
Spain	29.322	14.246	37.744	0.7203
France	26.128	9.170	30.283	0.9627
UK	2.750	0.114	3.381	0.9990
USA	21.339	10.222	31.972	0.9963

sub-sets: (a) the training dataset containing 88 sample points (80% of the data) required to train model and (b) the testing dataset containing 22 sample points (20% of the data) to assess the performance of a model. There is a difference between LS-SVM performance in testing and training. Generally, LS-SVM has better performance in the training phase than in the testing phase due to overtraining. LS-SVM can avoid overtraining so that generalizations can be better made [12,21,28].

The performance of the prediction of ARIMA and LS-SVM models is given in Tables 1 and 2 separately. The results indicate a drop in errors (MAE, MSE, and RMSE) and an increase in accuracy of prediction (coefficient of determination) using the LS-SVM model. For the countries Italy, Spain, and France, LS-SVM reduces errors approximately by 60% as compared to the errors for the UK and

the USA. The RMSE values using the LS-SVM model are found to be reduced by *80%* and *64%* as compared to the ARIMA model for the countries Spain and France respectively. For Italy, the decrease in RMSE value using the LS-SVM model is nearly *50%* as compared to the ARIMA model. The observed and LS-SVM model forecast values are found to be closer as compared to those obtained from the ARIMA model. A comparison of observed and forecast values of a testing dataset of COVID-19 confirmed cases from five countries by ARIMA and LS-SVM models shown in Fig. 3. One-month ahead out of sample forecast of COVID-19 confirmed cases from most affected five countries of the world generated by ARIMA, and LSSVM models are shown in Figs. 4 and 5. These figures show a sharp rise in COVID-19 confirmed cases from most affected five countries of the world.

4. Conclusions

The present study deals with the development of ARIMA and the LS-SVM models to predict one-month confirmed cases of COVID-19 from most affected five countries of the world namely Italy, Spain, France, the United Kingdom, and the United States of America. ARIMA and the LS-SVM models have different tendencies to handle COVID-19 data from the five countries. The input of COVID-19 confirmed cases from five countries were given to

Table 2
Prediction performance of the LS-SVM model.

Country	Training				Testing			
	MAE $(x10^3)$	MSE $(x10^8)$	RMSE $(x10^3)$	R^2	MAE $(x10^3)$	MSE $(x10^8)$	RMSE $(x10^3)$	R^2
Italy	7.484	0.849	9.214	0.9788	4.616	0.337	6.114	0.9299
Spain	2.436	0.073	2.719	0.9988	5.790	0.515	7.176	0.8067
France	0.770	0.012	1.103	0.9992	12.180	1.680	12.965	0.9782
UK	0.705	0.006	0.782	0.9991	3.520	0.157	3.964	0.9966
USA	3.082	0.131	3.628	0.9997	18.405	5.137	22.667	0.9969

ARIMA and LS-SVM models according to their systematic methodologies and one-month ahead predictions were obtained after validating these models. The results showed more accuracy of the LS-SVM model as compared to the ARIMA model. The accuracy of the prediction of the LS-SVM model in forecasting COVID-19 confirmed cases for Spain was *80%* more than the ARIMA model whereas the accuracy of the LS-SVM model over the ARIMA model was approximately *60%* for the countries Italy and France. The values of the coefficient of determination (R^2) are improved by the LS-SVM model. Moreover, one-month ahead predictions of confirmed COVID-19 cases from five countries using LSSVM models showed a sharp increase in COVID-19 confirmed cases as compared to the ARIMA model. Besides various safety measures of lock-down, quarantine, mass testing of symptoms, both ARIMA and LS-SVM models indicate a sharp rise in COVID-19 confirmed cases across the globe. The values of forecasting errors MAE, MSE, and RMSE in the validation phase point towards more accuracy of the LS-SVM model over the ARIMA model. The prediction results will help all the above-said countries to estimate the future cases which will help to prepare the number of isolation wards, availability of medicine, decision on lock-down, train volunteers, further economic plans, etc. The study depicts that governments have to take important strict decisions to control the spread and cure the problem.

Author credit statement

All authors contributed in the manuscript

Sarbjit Singh did the data analysis part using ARIMA and LSSVM.

Dr. Kulwinder Singh Parmar (corresponding), writes the methodology and results.

Sidhu Jitendra Singh Makkhan compiles the research data and helps to write the introduction.

Jatinder Kaur helps in writing the abstract, introduction sections.

Dr. Shruti Peshoria provides the main idea and helps to improve the entire manuscript.

Prof. (Dr.) Jatinder Kumar helps in writing the conclusion section and references setting etc.

Data statement

Data is freely available at the WHO website https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [44,45].

Declaration of Competing Interest

Authors have no conflict of interest.

Acknowledgement

References

[1] Abdirizak F, Lewis R, Chowell G. "Evaluating the potential impact of targeted vaccination strategies against severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) outbreaks in the healthcare setting. Theor Biol Med Model 2019;16(1):16.

[2] Aladaga CH, Egrioglub E, Kadilar C. "Forecasting non-linear time series with a hybrid methodology. Appl Math Lett 2009;22(9):1467–70.

[3] Binti Hamzah FA, Lau C, Nazri H, Ligot DV, Lee G, Tan CL, et al. CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. [Submitted].. Bull World Health Organ 2020. doi:10.2471/BLT.20.255695.

[4] Box GEP, Jenkins GM. Time series analysis, forecasting and control. San Francisco, CA: Holden Day; 1976.

[5] Brockwell PJ, Davis RA. Introduction to time series and forecasting. Springer Texts in Statistics; 2002.

[6] Che J, Wang J. "Short-term load forecasting using a kernel-based support vector regression combination model. Appl Energy 2014;132:602–9.

[7] Chen TM, Rui J, Wang QP, Zhao ZY, Cui JA, Yin L. "A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. Infect Dis Poverty 2020;9(1):1–8.

[8] Choi SC, Ki M. "Estimating the reproductive number and the outbreak size of Novel Coronavirus disease (COVID-19) using mathematical model in Republic of Korea. Epidemiol Health 2020:e2020011.

[9] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.

[10] Dye C, Gay N. Modeling the SARS epidemic. Science 2003;300(5627):1884–5.

[11] Faruk DO. "A hybrid neural network and ARIMA model for water quality time series prediction. Eng Appl Artif Intell 2010;23(4):586–94.

[12] Gong Q, Lu W, Gong W, Wang X. Short-term load forecasting of LSSVM based on improved PSO algorithm. In: Chinese conference on pattern recognition. Berlin, Heidelberg: Springer; 2014. p. 63–71.

[13] Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Munday JD, Kucharski AJ, Edmunds WJ, Sun F, Flasche S. "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Global Health 2020;8:e488–96.

[14] Huang C, Wang Y, Li X, Rem L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet North Am Ed 2020. doi:10.1016/S0140-6736(20)30183-5.

[15] Kantz H, Schreider T. Non-linear time series analysis. Cambridge University Press; 1997.

[16] Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Eggo RM, Sun F, Jit M, Munday JD, Davies N. "Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020. doi:10.1016/S1473-3099(20)30144-4.

[17] Kisi O, Parmar KS. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. J Hydrol 2016;534:104–12.

[18] Kisi O, Parmar KS, Soni K. Modeling of air pollutants using least square support vector regression, multivariate adaptive regression spline, and M5 model tree models. Air Qual Atmos Health 2017;10:873–83. doi:10.1007/s11869-017-0477-9.

[19] Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. Int J Antimicrob Agents 2020;55(3):105924 ISSN 0924-8579. doi:10.1016/j.ijantimicag.2020.105924.

[20] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY, Xing X. "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med 2020;382:1199–207.

[21] Li Y. Short-term load forecasting based on LS-SVM optimized by BCC algorithm. In: 2009 15th International conference on intelligent system applications to power systems. IEEE; 2009. p. 1–5.

[22] Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, Wang MH, Cai Y, Wang W, Yang L, He D. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. Int J Infect Dis 2020;93:211–16.

[23] Lohani AK, Goel NK, Bhatia KKS. "Improving real time flood forecasting using Fuzzy Inference system. J Hydrol 2014;509:25–41.

[24] Ma, Z. (2020). "Spatiotemporal fluctuation scaling law and metapopulation modeling of the novel coronavirus (COVID-19) and SARS outbreaks". arXiv:2003.03714.

[25] Ma, Z.E., Zhou, Y.C., Wang, W.D. (2004). "Mathematical modeling and research of infectious disease dynamics."

[26] Melard G, Pasteels JM. "Automatic ARIMA modeling including interventions, using time series expert software. Int J Forecast 2000;16(4):497–508.

[27] Misra D, Oommen T, Agarwal A, Mishra SK, Thompson AM. "Application and analysis of support vector machine-based simulation for runoff and sediment yield. Biosystems Eng 2009;103:527–35.

[28] Mustaffa Z, Yusof Y, Kamaruddin SS. "Gasoline price forecasting: an application of LSSVM with improved ABC. Procedia-Soc Behav Sci 2014;129:601–9.

[29] Parmar KS, Bhardwaj R. River water prediction modeling using neural networks, fuzzy and wavelet coupled model. Water Resour Manage 2015;29(1):17–33.

[30] Parmar KS, Bhardwaj R. "Water quality management using statistical analysis and time-series prediction model. Appl Water Sci 2014;4:425. doi:10.1007/s13201-014-0159-9.

[31] Parmar KS, Bhardwaj R. Statistical, time series, and fractal analysis of full stretch of river Yamuna (India) for water quality management. Environ Sci Pollut Res 2015;22(1):397–414.

[32] Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, Clifford S, Pearson CA, Munday JD, Abbott S. "The effect of control trategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health 2020. doi:10.1016/S2468-2667(20)30073-6.

[33] Sapankevych NI, Sankar R. "Time series prediction using support vector machines: a survey. IEEE Comput Intell Mag 2009;4(2):24–38.

[34] Shen M, Peng Z, Xiao Y. "Modeling the epidemic trend of the 2019 novel coronavirus outbreak in China. bioRxiv 2020.

[35] Singh S, Parmar KS, Kumar J, Makkhan SJS. Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19. Chaos Solitons Fractals 2020:109866 (In Press). doi:10.1016/j.chaos.2020.109866.

[36] Soni K, Parmar KS, Kapoor S, Kumar N. Statistical variability comparison in MODIS and AERONET derived aerosol optical depth over Indo-Gangetic Plains using time series. Sci Total Environ 2016;553:258–65.

[37] Soni K, Parmar KS, Kapoor S. "Time series model prediction and trend variability of aerosol optical depth over coal mines in India. Environ Sci Pollut Res 2015;22(5):3652–71.

[38] Soni K, Kapoor S, Parmar KS, Kaskaoutis DG. "Statistical analysis of aerosols over the Gangetic-Himalayan region using ARIMA model based on long-term MODIS observations. Atmos Res 2014;149:174–92.

[39] Soni K, Parmar KS, Agrawal S. "Modeling of air pollution in residential and industrial sites by integrating statistical and Daubechies wavelet (level 5) analysis. Model Earth Syst Environ 2017;3:1187–98.

[40] Suykens JAK, Vandewalle J. "Least squares support vector machine classifiers. Neural Process Lett 1999;9(3):293–300.

[41] Valenzuela O, Rojas I, Rojas F, Pomares H, Herrera LJ, Guillen A, et al. "Hybridization of intelligent techniques and ARIMA models for time series prediction. Fuzzy Sets Syst 2008;159(7):821–45.

[42] Van Gestel T, Suykens J, Baestaens D-E, Lambrechts A, Lanckriet G, Vandaele B, et al. "Financial time series prediction using least squares support vector machines within the evidence framework. IEEE Trans Neural Netw 2001;12(4):809–21.

[43] World Health Organization (WHO) Coronavirus disease 2019 (COVID-19) situation report – 70. WHO; 2020.

[44] World Health Organization. Coronavirus. World Health Organization, cited January 19, 2020. Available: https://www.who.int/health-topics/coronavirus.

[45] World Health Organization (WHO). Coronavirus 2020.

[46] Wu JT, Leung K, Leung GM. "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet North Am Ed 2020;395(10225):689–97. doi:10.1016/S0140-6736(20)30260-9.

[47] Xu R, Bian G. "Discussion about non-linear time series prediction using least squares support vector machine. Commun Theor Phys 2005;43:1056–60.

[48] Yang P, Wang X. "COVID-19: a new challenge for human beings. Cell Mol Immunol 2020. doi:10.1038/s41423-020-0407-x.

[49] Yuan Z, Xiao Y, Dai Z, Huang J, Chen Y. "A simple model to assess Wuhan lockdown effect and region efforts during COVID-19 epidemic in China Mainland. [Submitted].. Bull World Health Organ 2020 E-pub: 02 March. doi:10.2471/BLT.20.254045.

[50] Zhang GP. "Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 2003;50:159–75.

[51] Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. Int J Infect Dis 2020. doi:10.1016/j.ijid.2020.01.050.

[52] Zhao S, Musa SS, Lin Q, Ran J, Yang G, Wang W, et al. Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of January 2020: a data-driven modelling analysis of the early outbreak. J Clin Med 2020. doi:10.3390/jcm9020388.

[53] Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. "A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020. doi:10.1038/s41586-020-2012-7.

[54] Zhu B, Wei Y. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. Omega 2013;41(3):517–24.

[55] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020. doi:10.1056/NEJMoa2001017.