Communications in Information and Systems
Volume 24, Number 1, 61–84, 2024

Machine learning predicting COVID-19 in Algeria

Fatima Zohra Younsi, Mohammed Chems Eddine Sahinine, and Ilyes Benarroum

Predicting new and urgent trends in epidemiological data is an important public health problem. This problem has gained increasing attention of the data mining and machine learning research communities. Artificial Intelligence can extract relevant information from an increasingly accessible dataset that would be difficult to navigate in a nonautomated way.

In this paper, our goal is to propose a new spatiotemporal system for predicting COVID-19 cases in the 48 cities of Algeria. This system is mainly based on AI methods, namely: ARIMA, LSTM, SLSTM and Prophet. Real-time data collection was used in our study. The dataset is randomly split into training set and testing set. In our prediction experiment, comparison between observed and predicted values using performance metrics were performed and obtained results were very satisfactory and stable and responses fit completely to each other.

The main purpose of our system is to help public health officials and planners to manage services and organize medical infrastructure as well as evaluate action plans to deal with future course of the COVID-19 epidemic.

1. Introduction

The 2019 novel COronaVIrus Disease (COVID-19 or SARS-CoV-2) is a contagious disease caused by severe acute respiratory syndrome coronavirus 2; the disease broke out in Wuhan, China in December 2019. According to the World Health Organization (WHO), the pandemic has spread to 224 countries and territories around the world with more than 300,000,000 confirmed cases, and causing more than 4 million deaths in 2021. Many government efforts focused to develop and integrate containment planning and in some regions declare emergencies and isolations. With high transmissibility of the virus, many schools and universities have closed down before finishing their year programs. Many persons lost their jobs due to social distancing, self-isolation and travel restrictions. Moreover, the outbreak lead to a reduced workforce across all economic sectors and it is causing an important concern and economic hardship for businesses and consumers.

The main objectives of the field of epidemiology include the study of the frequency of occurrences of diseases in different groups of people as well as the study of the dynamics of different health states. Many scientists and researchers are trying to understand the mechanism and the dynamic of COVID-19 spread in the human population in order to reduce the infection and enable a timely response to the outbreak [1]. In the literature, there are various approaches, methods and models that were carried out to control and predict COVID-19 propagation. Among the developed and implemented models, we find models based on AI methods. As known, AI is a rapidly expanding field of research with a great future ahead of it. Currently, its applications concern all human activities and therefore, it makes possible in particular to improve the quality of healthcare. The main applications of AI are multiplying enormously (automation of tasks, customer relations, logistics, predictive analysis, diagnostics, big data analysis, etc.). In the prediction context, AI driven methods can be useful to predict the parameters, risks, and effects of COVID-19 epidemic. Such predictions can be helpful to control and prevent the spread of such diseases [2].

Current public health information systems often fail to detect emerging infectious diseases or future disease course. Integrating clinical data and relevant AI tools can address critical gaps and accelerate the prediction and understanding of outbreak dynamics.

Various methods and techniques are used in this field; these methods and techniques help decision makers to find the main causes of health problems and disease transmission in populations and it deals with the incidence, distribution and eventual control of diseases and other health-related factors. Indeed, the development of emergency measures, infection prevention and protection plans requires the implementation of tools for the future prediction of the disease [3, 4]. In this context, prediction can be done from the collection of data and the use of AI methods. Indeed, the aim of prediction in epidemiology is to collect and analyze various large datasets (medical, geographical, socio-economic, etc.) and to implement new algorithms and approaches to limit the epidemic spread in order to save as many human lives as possible and to stop its spread in the human population.

In Algeria, the aim of public health information systems used by public health agencies is to collect, manage, store, and transmit data. These systems suffer from several technical drawbacks according to [5, 3]. Among these shortcomings: lack of analytical capabilities, absence of spatiotemporal decision support systems, no integration of artificial intelligence tools or prediction tools in current systems. Therefore, in this study, we are interested in the development of a new prediction system dedicated to the epidemiology of COVID-19 in Algeria. The tool is mainly based on four AI methods, namely: ARIMA, LSTM, SLSTM and PROPHET. The purpose of our system is to help public health managers and planners manage services and organize medical infrastructure as well as evaluate action plans to fight occurrence of a new COVID-19 epidemic wave.

1.1. Our motivation

The COVID-19 infectious disease is now a major concern in public health. Many researchers around the world have been focused on evaluating various actions and strategies to combat the new virus, to understand its spreading mechanism, to find out the real and responsible factors for the epidemic spread and explaining its emergence in the human population. Indeed, decision makers need efficient and effective systems that can predict the COVID-19 evolution in future and in short term.

The time series data of daily new confirmed cases provide information on the force of infection and investment in the prevention resources of epidemics. Using these time series data, it is possible to forecast the number of daily new confirmed cases in the near future. In other words, the number of daily new confirmed cases relates to historical data [6].

1.2. Our contribution

In this article, we investigate the suggested system for predicting COVID-19 future evolution in Algeria and discuss experiment results. Our major contributions are as follows:

1. Develop and discuss a new spatiotemporal system for predicting the number of COVID-19 positive reported cases in Algeria;
2. Develop four different methods of AI, namely: ARIMA, LSTM, SLSTM and Prophet;
3. Validate the used methods by calculating performance metrics;
4. Visualize the prediction results on risk map and making all the plots.

We proceed as follows: In Section 2, we provide some relevant research related to our study. In Section 3 we present and discuss in detail our proposed system. In Section 4, we show the approach adopted by the proposed system followed by the description of the used performance metrics. UML sequence diagram and adopted approach are presented in Section 5. In Section 6, we present datasets that are used for evaluating our system and we discuss experiment results. Finally, we conclude our article in Section 7.

2. Related works

Several researchers have conducted prediction methods and models related to the COVID-19 disease [2, 6, 7, 8, 9, 10], etc. In this section, we describe the differences between this article and existing works. For example, [7] used Deep Learning-based models for predicting the number of novel coronavirus (COVID-19) positive reported cases for 32 states and union territories of India. Recurrent neural network (RNN) based long-short term memory (LSTM) variants such as Deep LSTM, Convolutional LSTM and Bi-directional LSTM were applied in the study. Daily and weekly predictions are calculated for all states, and it is found that bi-LSTM gives very accurate results. Similarly, [2] proposed a Bayesian optimization guided shallow LSTM for predicting the country-specific risk of the novel coronavirus (COVID-19). The authors have used the trend data to predict different parameters for the risk classification task. Obtained results show that the proposed pipeline outperforms state-of-the-art methods for data of 180 countries and can be a useful tool for such risk categorization. Data analytics was adopted by developing a nonlinear autoregressive exogenous input (NARX) neural network-based algorithm to predict new COVID-19 cases considering the historical data of COVID-19 cases alongside the external factors that affect the spread of the virus in [11]. The predicted COVID-19 cases help in providing some recommendations for both the government and people of the affected countries. Furthermore, [6] introduced a machine learning model based on the LSTM and XGBoost models to investigate the future trend of COVID-19 in America and evaluate the important features based on the reported COVID-19 cases. The results of test set show that MAPE of the LSTM and XGBoost algorithms reach 2.32% and 7.21%, respectively and the LSTM model has the lower metrics value. The aim of [8]'s study was to test how accurate the ARIMA best-fit model predictions were with the actual values reported after the entire time of the prediction had elapsed. The authors investigated and validated the accuracy of an ARIMA model over a relatively long period of time using Kuwait as a case study. As a result, the accuracy of the prediction provided by the ARIMA proposed model was both appropriate and satisfactory. In the other research, [12] created time series prediction of COVID-19 outcomes is done using ARIMA, LSTM, SLSTM and Prophet models to estimate the future prediction of confirmed, death and recovered cases for the specified time intervals provided in the model. The results of the analysis show that the Stacked LSTM and LSTM models outperformed other studied models with higher accuracy and it proves the reliability for predicting COVID-19 cases. Recently, [13] introduced a hybrid approach to improving medical image analysis for the detection of patients infected with COVID-19. To obtain better results, the authors segmented the CT image dataset to extract information about the lungs. The proposed technique used the opt-aiNet as FS technique in conjunction with DL as well as ML classifiers to detect COVID-19 infected patients. The results showed that the hybrid approach outperformed the simple algorithm while reducing the training time as well. Furthermore, [14] used the time series datasets, collected from all COVID-19-affected countries, for proposing two DL-based prediction models. Two NN-based prediction models, RNN and LSTM, have been evaluated using time series data from three datasets. The proposed LSTM prediction model showed a 98.53% accuracy with regards to the number of confirmed COVID-19 cases and resultant deaths. Finally, the main goals of [10] are reviewing the use of different ML applications in COVID-19 disease for different purposes using various algorithms. Based on the findings of the study, it appears that ML have been utilized mostly for COVID-19 forecasting, detection, identification, and screening from December 2019 to the present. The research examined also the methodology of ML algorithms to COVID-19.

Considering the above-mentioned discussion, in this paper, we introduced a new spatiotemporal COVID-19 Prediction SYStem called COPSYS for predicting new cases of COVID-19 in the 48 cities of Algeria. Furthermore, we develop ARIMA, LSTM, SLSTM and Prophet models and evaluate the performance of the learning models. Machine learning models aid in early detection of COVID-19 outbreak. The used models are evaluated using performance metrics. Thus, we present prediction numerical results and we visualize it on risk map of the country. To the best of our knowledge, this type of study of development new spatiotemporal system for COVID-19 prediction using AI models in Algeria has never been executed.

3. Description of our system

In this section, we describe our proposal in detail. Our goal is to develop a new spatiotemporal COVID-19 Prediction SYStem named COPSYS. The purpose of the latter is to predict the spread of COVID-19 in Algeria and its cities using four methods: ARIMA method, Prophet method, LSTM method and SLSTM method. The suggested system consists of three modules as shown in Figure 1. The modules are: (1) predictive system, (2) visualization system, and (3) accuracy analysis module.

The COPSYS system provides an interface that allows interaction between the decision maker and the system. Through this interface, the decision maker can access and search the epidemiological dataset, choose one of the AI-based prediction methods, or visualize the obtained results. All the methods developed in this study are applied to the dataset to investigate a more pragmatic prediction. The results obtained by the prediction system can be visualized via the visualization system. Moreover, using these results and the dataset the validation accuracy can be achieved.

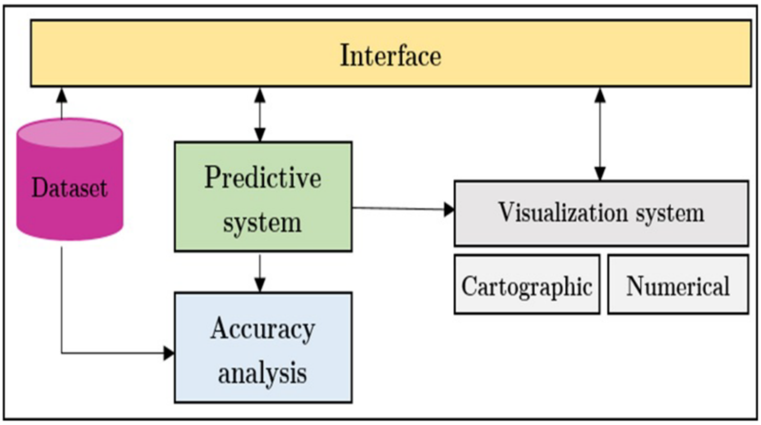


Figure 1: The COPSYS proposed system.

3.1. Predictive system

In what follows, we describe the AI models developed for this study. Our system is designed by using four main components: ARIMA method, Prophet method, LSTM method and SLSTM method.

3.1.1. Long-Short-Term Memory (LSTM) method LSTM is a type of Recurrent Neural Network (RNN). The method is useful for making predictions, classification and processing sequential data. The LSTM configuration the most commonly used in literature was described by Graves and Schmidhuber [15]. The objective of the LSTM developed method is to predict the evolution of COVID-19 in Algeria using Artificial Neural Network (ANN) and deep learning approach. We chose the LSTM method because, in accordance with literature, for prediction tasks, LSTMs are considered to be among the most feasible solutions, and they anticipate the future forecasts dependent on various highlighted features present in the dataset [7]. Besides, LSTM method is well suited for processing and making predictions based on time series data and it has feedback connections. LSTM include a memory cell' that are able to store information for long periods, which is useful when dealing with time series or sequential data. The architecture of LSTM with only input and output gates is shown in Figure 2.

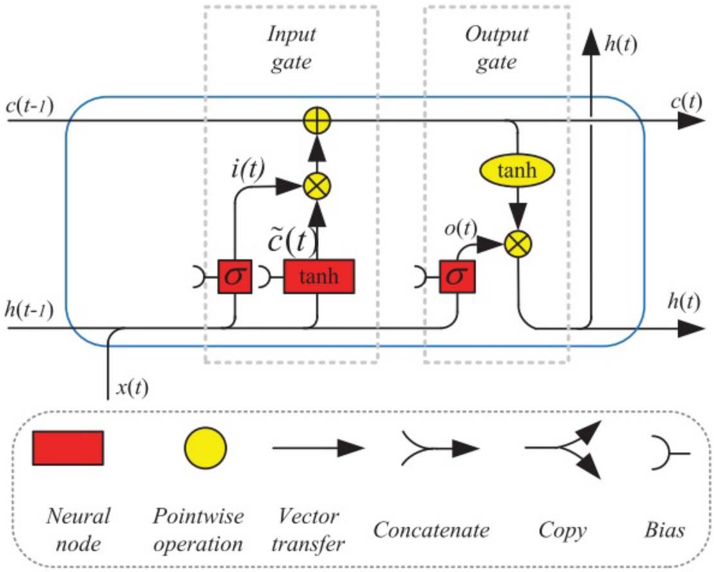


Figure 2: LSTM network architecture [16].

3.1.2. Stacked LSTM (SLSTM) or Deep LSTM method The stacked LSTM is an extension of standard LSTM described in the previous subsection (Section 3.1.1). The methods were proposed by [17]; it consists of multiple hidden layers with multiple memory cells. The SLSTM model uses multiple LSTM layers that are stacked before the forwarding to a dropout layer and output layer at the final output. In a stacked LSTM, the first LSTM layer produces sequence vectors used as the input of the subsequent LSTM layer. Moreover, the LSTM layer receives feedback from its previous time step, thus allowing for the capturing of data patterns. The dropout layer also excludes 10% of the neurons to avoid overfitting [18]. A block of three recurrent layers is illustrated in Figure 3 according to [16].

3.1.3. Auto-Regressive Integrated Moving Average model (ARIMA) method ARIMA models were popularized by [19]. The method is a time-series forecasting approach that is used in predicting the future value of a variable from its own past values. It uses auto-regression and moving average, and incorporates a differencing order to remove trend and/or

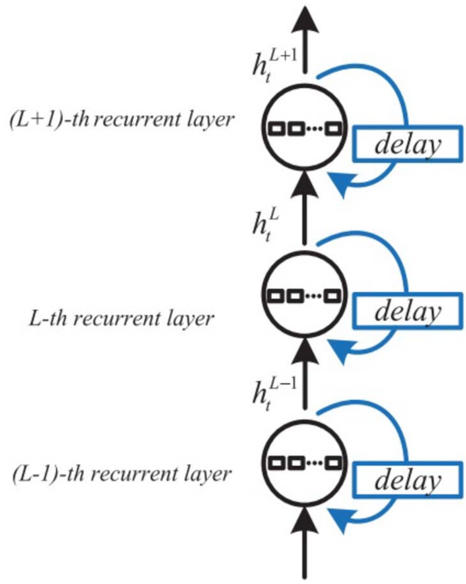


Figure 3: Stacked LSTM network architecture.

seasonality [8]. The aim of this model is to predict the evolution of COVID-19 in Algeria using classic time series. We will use classical methods such as ARIMA because these outperform machine learning and deep learning methods for one-step forecasting on univariate datasets and multi-step forecasting on univariate datasets. At present, machine learning and deep learning methods are not yet delivering on their promise of univariate time series predictions and there is a lot of work to be done. Figure 4 illustrates the ARIMA model.

3.1.4. Prophet method The Prophet method is designed for automatic forecasting of univariate time series data. It is an open-source library developed by Facebook's Core Data Science team in 2017. The method is based on an additive model where non-linear trends are adjusted for annual, weekly, and daily seasonality. Prophet is robust to missing data and trend shifts, and it handles outliers well. Prophet is an additive regression model with four components that are combined in the following equation [21]:

$$y(t) = g(t) + s(t) + h(t) + t \tag{1}$$

where, $g(t)$: piecewise linear or logistic growth curve for modeling non-periodic changes in time series. $s(t)$: periodic changes (e.g., weekly/yearly

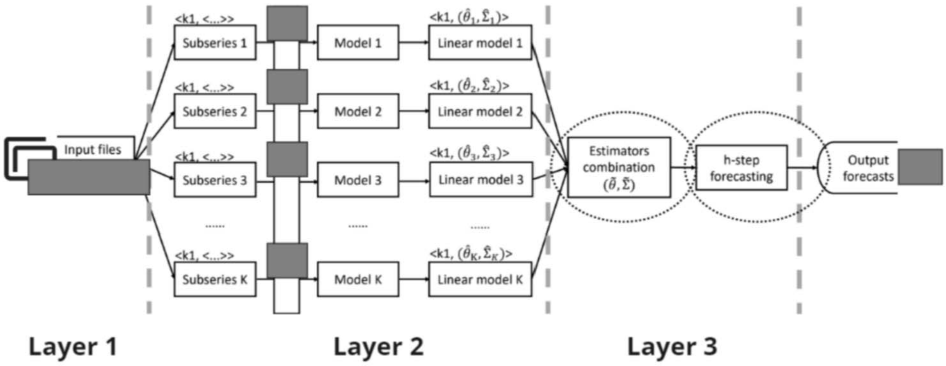


Figure 4: The ARIMA model [20].

seasonality). $h(t)$: holiday effects (provided by the user) with irregular schedules. t: the error term accounts for any unusual changes that the model does not take into account.

Prophet uses time as a regressor trying to fit several linear and nonlinear time functions as components [21]. The following diagram illustrates the prediction process used in Prophet model according to [22]:

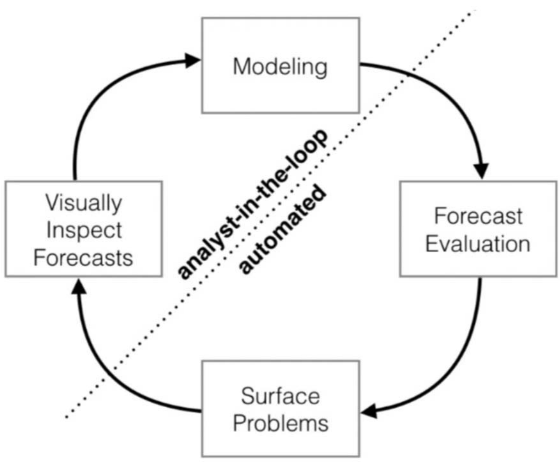


Figure 5: Prediction process using Prophet model.

3.2. Accuracy analysis: performance metrics

In this section, we evaluate the performance of the learning models in terms of R-squared (R^2) score, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square error (RMSE), and Mean Absolute Percentage Error (MAPE) where, the y_j notation is considered the real values, and the ($\hat{y}_i$) notation, the predicted values.

3.2.1. R-squared ($\mathbf{R^2}$) score It is an important indicator used in statistics to measure the quality of a regression model. Mathematically, it is the proportion of the variance of a dependent variable that is explained by one or more independent variables in the regression model. It is expressed either between 0 and 1, or as a percentage. R-squared indicates how well a regression model fits a data set.

The formula for calculating R-squared is [23]:

$$R^2 = \left(\sum_{i=1}^{n} (yi - \widehat{y})(yi - \widehat{y}) / \sum_{i=1}^{n} (yi - \bar{y})^2 \sum_{i=1}^{n} (\hat{y}i - \bar{y})^2 \right)^2 \tag{2}$$

where $\bar{y}$ is the mean of the y values.

3.2.2. Mean Absolute Error (MAE) It is the mean of the absolute errors. The absolute error is the absolute value of the difference between the predicted value and the actual value. The MAE value indicates how big, on average, is the error that can be expected from the forecast [24].

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^{n} |yj - \widehat{yj}| \tag{3}$$

3.2.3. Mean Squared Error (MSE) Mean Squared Error (MSE) is included in the performance evaluation of the experiments since they provide quadratic loss functions that measure the forecasting uncertainty while focusing on the impact of large errors. The values of MSE could express the sum of the variance and square value of bias, further contributing to the performance analysis of a model [25].

$$\text{MSE} = \frac{1}{n} \sum_{j=1}^{n} (yj - \hat{y}j)^2 \tag{4}$$

3.2.4. Root Mean Squared Error (RMSE) Another relevant indicator is the RMSE. The later provides an indication with respect to the dispersion or variability of the quality of the prediction. The RMSE can be related to the variance of the model.

Additionally, the values of RMSE increase with the variance of the frequency distribution of error magnitudes, resulting in larger values when large error values are present [25]. The formula for calculating RMSE is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(yj - \hat{y}j)^2} \tag{5}$$

3.2.5. Mean Absolute Percentage Error (MAPE) MAPE quantifies accuracy as a percentage that can be calculated as a cumulative absolute percent error for each time frame, as the actual values minus the predicted values divided over the actual values [6]. That is, it depicts the mean error in percentage terms. The MAPE is described as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{j=1}^{n}\frac{|yj - \widehat{y}|}{|yj|} \tag{6}$$

3.3. Visualisation system

Our system aims to help public health officials carry out an epidemiological surveillance process by trying to predict the evolution of the spread of epidemics in order to control them and therefore be able to control an epidemiological situation in the future.

Indeed, we have developed a prediction system based on Artificial Intelligence methods namely: ARIMA, LSTM, SLSTM and Prophet to predict the future evolution of COVID-19 in Algeria. To ensure proper analysis by public health personnel, we have presented the prediction results in the form of a prediction map.

4. Approach adopted by the COPSYS system

After describing the proposed COPSYS system, description of the COPSYS approach is necessary. The approach adopted by COPSYS is given in the flowchart of Figure 6.

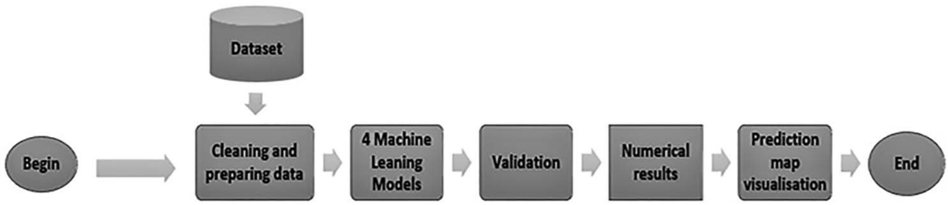


Figure 6: The approach adopted by COPSYS.

5. The sequence diagram UML Modelling COPSYS

Sequence diagram, called also event diagram, represents the flow of communications in the system. The diagram describes the interaction between two lifelines as a series of events ordered in time, as if these lifelines were present at the same time. UML is designed for object-oriented programming, and these communications between classes are recognized. Figure 7 illustrates the sequence diagram UML Modelling COPSYS.

6. Case study and experiments

The experiments aim to validate the performance of our COPSYS system. We present a case study illustrating the use of COPSYS. First, we describe the dataset used in this study as well as the study area. Then, we present and discuss the numerical results obtained from COPSYS. Next, we visualize the evolution of epidemic. Finally, we evaluate the performance of the learning models.

6.1. Dataset description

The dataset used in our experiments is composed of COVID-19 positive reported cases in Algeria between May 1st, 2020 and April 24th, 2021.

During this period, the country had recorded over 120,000 confirmed cases and 4000 deaths. We used this period because updated data are not currently available for all wilayas (cities) and territories. The dataset was obtained from a CSV file provided by Algeria Coronavirus Tracker API[1] for COVID-19 data in Algeria and it is the input of our suggested system. The CSV file includes confirmed COVID-19 cases per day during the period indicated. The collected data concern 48 cities of Algeria; the total number

[1] https://api.corona-dz.live/province/all

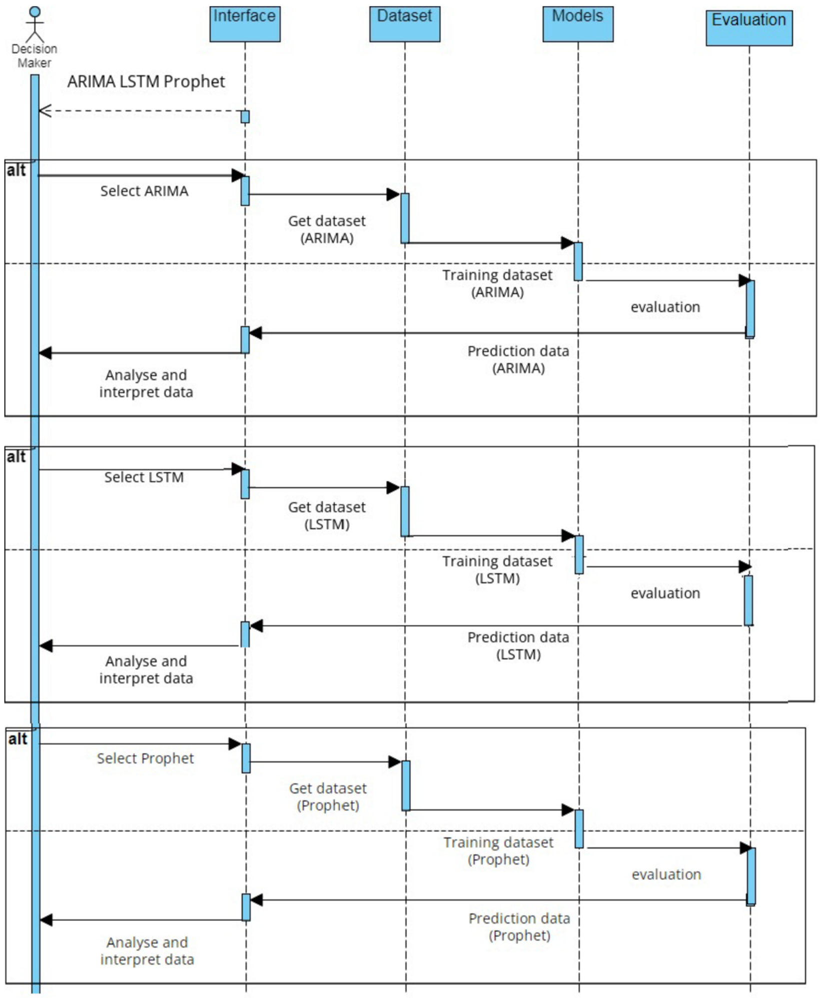
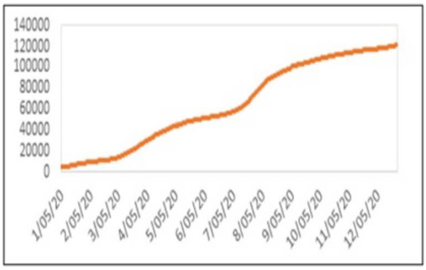


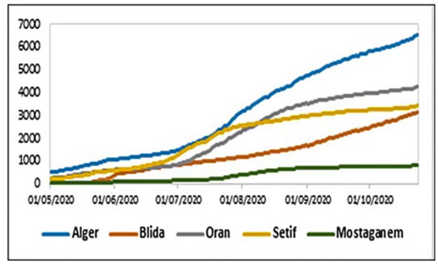
Figure 7: The sequence diagram.

of confirmed cases in the file were 123,672. In this study, 70% of the dataset was randomly selected for training, and test predictions were performed on the remaining 30% of the test data.

Figure 8a and 8b illustrate daily-confirmed COVID-19 cases dataset in Algeria and daily-confirmed COVID-19 cases dataset in top five affected areas used in this study, respectively.



(a) Daily-confirmed COVID-19 cases dataset in Algeria



(b) Daily-confirmed COVID-19 cases dataset in top five affected areas used in this study

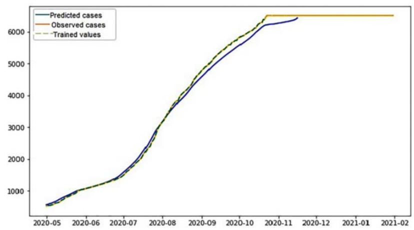
Figure 8: Number of COVID-19 positive cases.

6.2. Results and discussion

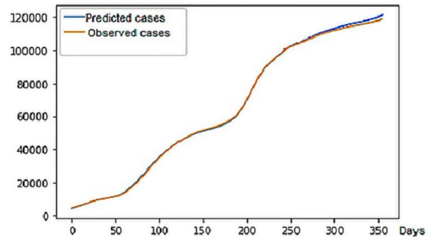
In this section, we present the results obtained from our system COPSYS. We illustrate the comparison graphs between the actual data and the predicted data for the wilayas with the high-confirmed COVID-19 cases. Finally, we present the COVID-19 prediction map.

It should be noted that the results obtained by the SLSTM method are the same as the results obtained by LSTM. This is due to the small size of the dataset used in this study. Indeed, in the following sections, we will not discuss the results of SLSTM.

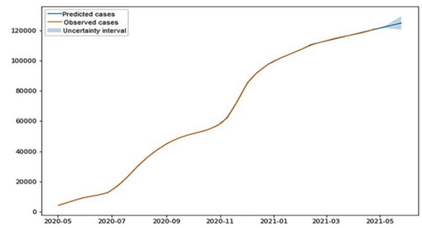
6.2.1. Numerical results Figure 9 illustrates the two curves (confirmed and predicted COVID-19 cases) up to 24th April 2021 using the three methods (Figure 9a) ARIMA, (Figure 9b) LSTM and (Figure 9c) Prophet in Algeria where the x-axis represents the date and the y-axis is the number of cases in Algeria. We notice that the difference between the two curves is small. As shown in the figure, the blue area in the Prophet Model curve (Figure 9c) indicates the uncertainty intervals, which shows the real observation within the range. Moreover, the prediction results for the next 60 days show a lower trend evolution COVID-19 incidence in all wilayas of Algeria. This was the result of many Algerian government efforts focused to develop and integrate containment planning and isolations, and in some regions declare emergencies. During this time, with high transmissibility of the virus, many schools and universities have closed down before finishing their year programs and many persons stopped working due to social distancing, self-isolation and travel restrictions (borders closure: sea, land and air).



(a) ARIMA method



(b) LSTM method

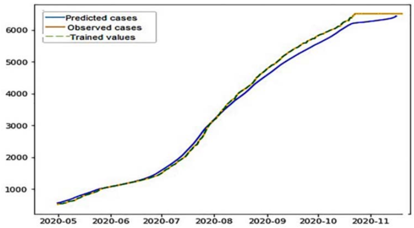


(c) Prophet method

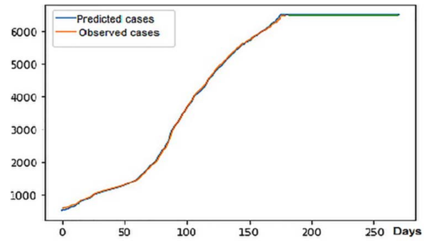
Figure 9: Prediction using (a) ARIMA, (b) LSTM and (c) Prophet Models for a count of confirmed COVID-19 cases (for the next 90 days) in Algeria.

In what follows, we display the three top affected wilayas, namely: Algiers, Blida and Oran in which the number of confirmed COVID-19 cases reaches above 1000 at the beginning of May 2020. Here, it should be noted that we have carried out learning and validation using available real databases for each wilaya and we provided the prediction of the next months. The data from 1st May 2020 till 23th October 2020 has been considered from which 70% of the data is used for training and 30% is used for test.

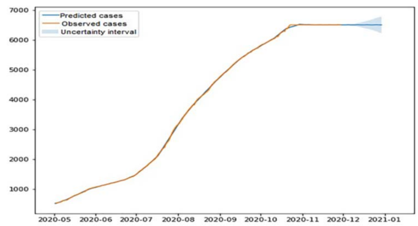
Figure 10 presents the two curves (confirmed and predicted COVID-19 cases) using the three methods (Figure 10a) ARIMA, (Figure 10b) LSTM and (Figure 10c) Prophet in Algiers. The x-axis represents the date and the y-axis is the number of cases in the wilaya of Algiers. The latter is the capital and largest city in Algeria. The Algiers population now totals over 3 million residents. During the COVID-19 epidemic, Algiers has recorded a significant number of new cases. According to the Figure 10, we notice that the spread of COVID-19 will continue to increase in the next days. Thus, the blue area in the Prophet Model curve (Figure 10c) indicates uncertainty (95% confidence interval) that shows the real observation within the range. The used models provided the prediction of the next three months. Data from 1st May 2020 to 1st October 2022 were used for training; the remaining 22 days were used for test. The suggested system provides prediction for two months using LSTM and prophet models and the next 24 days using ARIMA model.



(a) ARIMA method



(b) LSTM method

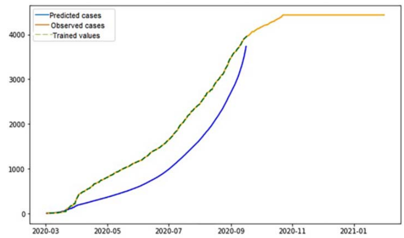


(c) Prophet method

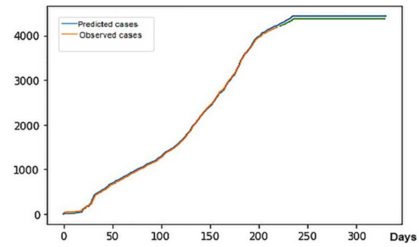
Figure 10: Prediction using (a) ARIMA, (b) LSTM and (c) Prophet Models for a count of confirmed COVID-19 cases (for the next 90 days) in the wilaya of Algiers.

The first positive case was declared in the wilaya of Blida in March 2020. According to the Scientific Monitoring Committee of the Coronavirus pandemic evolution, this city has reported more than 130 deaths by this virus and more than 4,400 new cases in the first weeks of the pandemic. The city was placed under quarantine for over a month. Figure 11 illustrates the two curves (confirmed and predicted COVID-19 cases) using the three methods (Figure 11a) ARIMA, (Figure 11b) LSTM and (Figure 11c) Prophet for the wilaya of Blida where the x-axis represents the date and the y-axis represents the number of cases in Blida. The used ARIMA model provided the prediction of next days. Data from 2nd March 2020 to 24th August 2020 were used for training and the remaining data were used for test. The suggested system provide prediction for two months using LSTM and prophet models and the next 24 days using ARIMA model.

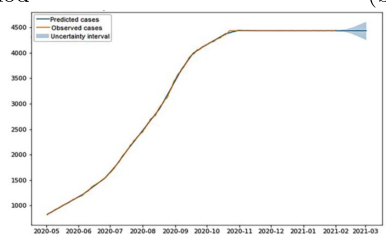
The wilaya of Oran is the second city of Algeria; the Oran population now totals over 2 million inhabitants. This city has recorded a significant number of COVID-19 confirmed cases. Indeed, the study of predicting the future course of COVID-19 seems important. We show in Figure 12 the comparison between the two curves confirmed and predicted COVID-19 cases obtained by the three methods (a) ARIMA, (b) LSTM and (c) Prophet for



(a) ARIMA method



(b) LSTM method



(c) Prophet method

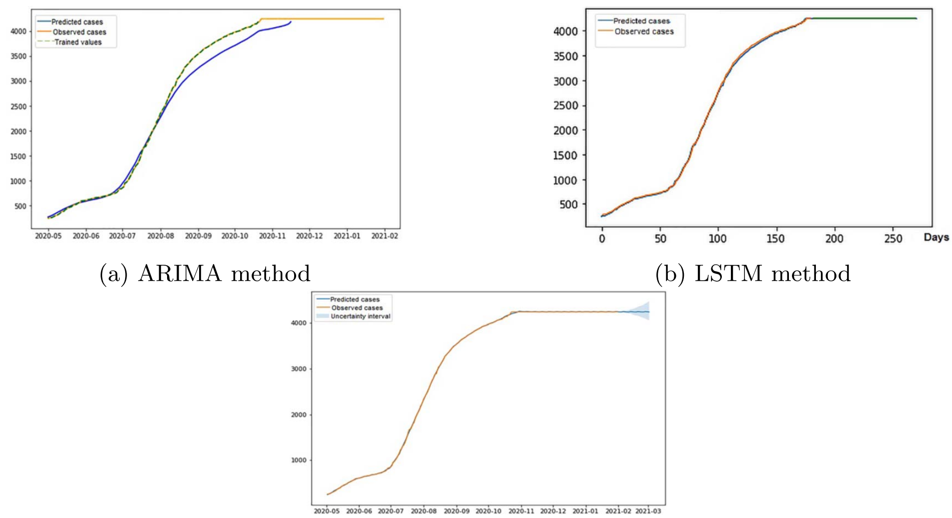
Figure 11: Prediction using (a) ARIMA, (b) LSTM and (c) Prophet Models for a count of confirmed COVID-19 cases (for the next 90 days) in the wilaya of Blida.

the wilaya of Oran. For the graph, the x-axis and y-axis represent the date and the number of cases in Oran, respectively.

Figures 12a, 12b, 12c illustrate the two months' prediction of confirmed cases with confidence intervals (CI) of 95% by December 2020. As shown by the figure, the number of COVID-19 new cases will continue to increase slowly in the next days. This explains that, the situation was not very serious and worrisome as the number of hospital admissions was very low compared with other cities in the world, which have reported sharp increases in COVID-19 incidence. Hence, this gives a chance to eliminate and control the disease in the future.

6.2.2. Statistical analysis As result of the evaluation of the test set, the calculated metrics are presented in Tables 1, 2 and 3. Here, we confirm that each predictive model contain errors.

Tables 1, 2 and 3 present the performances of each model in terms of average MAPE, RMSE, MAE, R-squared score, Correlation Coefficient (CC) and MSE for COVID-19 data obtained for Algeria and the top five cities affected by COVID-19 in the country, namely: Algiers, Blida, Oran, Setif and Mostaganem respectively.



(a) ARIMA method

(b) LSTM method

(c) Prophet method

Figure 12: Prediction using (a) ARIMA, (b) LSTM and (c) Prophet Models for a count of confirmed COVID-19 cases (for the next 90 days) in the wilaya of Oran.

Table 1: Performance evaluation using ARIMA model

	ARIMA model					
	Algeria	Algiers	Blida	Oran	Setif	Mostaganem
MAPE	1.15	3.69	10.99	5.55	6.37	9.15
RMSE	439.69	146.80	192.15	453.95	170.78	39.02
MAE	158.142	120.17	157.53	173.04	138.96	32.56
MSE	172.096	21549.01	36921.22	206075.13	29164.69	1522.21
CC	0.99994	0.9996	0.9997	0.998	0.9981	0.9975
$\mathbf{R}^2$ score	0.9998340	0.999196	0.9994	0.9978	0.9962	0.9951

Table 2: Performance evaluation using LSTM model

	LSTM model					
	Algeria	Algiers	Blida	Oran	Setif	Mostaganem
MAPE	17.21	1.36	5.04	1.28	1.10	1.68
RMSE	11197.41	39.63	29.94	27.6390	22.22	7.1825
MAE	1988.60	34.85	24.55	21.35	19.22	5.15
MSE	125382052.32	1570.92	896.19	763.91	51.59	51.59
CC	0.999723	0.99990	0.999909	0.999839	0.999868	0.99983
R^2 score	0.999446	0.999854	0.999765	0.99964	0.999768	0.999659

Table 3: Performance evaluation using Prophet model

	Prophet model					
	Algeria	Algiers	Blida	Oran	Setif	Mostaganem
MAPE	1.47	0.40	4.44	1.13	1.18	1.13
RMSE	462.61	16.70	22.24	25.11	22.48	3.89
MAE	214.36	10.34	15.46	16.54	13.97	2.47
MSE	15769.03	278.90	494.71	630.64	505.47	15.17
CC	0.99995	0.99997	0.99994	0.99992	0.99985	0.99992
R^2 score	0.99989	0.99991	0.99985	0.99980	0.99976	0.99983

In terms of the calculated performance metrics, in the case of Algeria, we observe that ARIMA Model achieved the best prediction performance with the lowest MAPE, RMSE, MAE, MSE values and good value for correlation coefficient and $\mathbf{R}^2$. However, for the five wilayas, the prediction performed by Prophet Model shows high accuracy in terms of average MAPE, RMSE, MAE, $\mathbf{R}^2$ score and MSE. From the obtained results, it is clear that Prophet Model has a minimal error in percentage values for confirmed COVID-19 cases than the other models for the five wilayas. Moreover, we notice that the other performance metrics have higher values because a less amount of data is available for COVID-19. This indicates that the Prophet model can well predict the future course of COVID-19 disease most accurately among the three models. It could be related to its ability to fit on a small amount of dataset. In addition, the obtained results for the performance metrics were stable and it did not vary after various executions.

6.2.3. Prediction map It is clearly shown in this research that our aim is to develop a spatiotemporal tool for predicting COVID-19 disease. As known, Geographic Information System (GIS) is a critical tool in tracking, analysis and helping to combat epidemics contagion. Analysis tools are necessary to treat different data and build relevant, reliable and accurate information for monitoring, prediction and prevention [26]. Indeed, we create maps predicting for the country showing which wilayas will be worst affected by a future surge in COVID-19 cases using the three models. Figure 13 illustrates the COVID-19 risk map for Algeria (showing the epidemic prediction for each wilaya) on 15th November 2020. We divided the country into five different categories based on the number of confirmed COVID-19 cases to classify the COVID-19 future incidence from the lowest level to the highest level. From the prediction map, it was inferred that the very high risk of COVID-19 epidemic cases was recorded in Algiers, Blida and Oran. Areas of high and average incidence were recorded in some wilayas, for example:

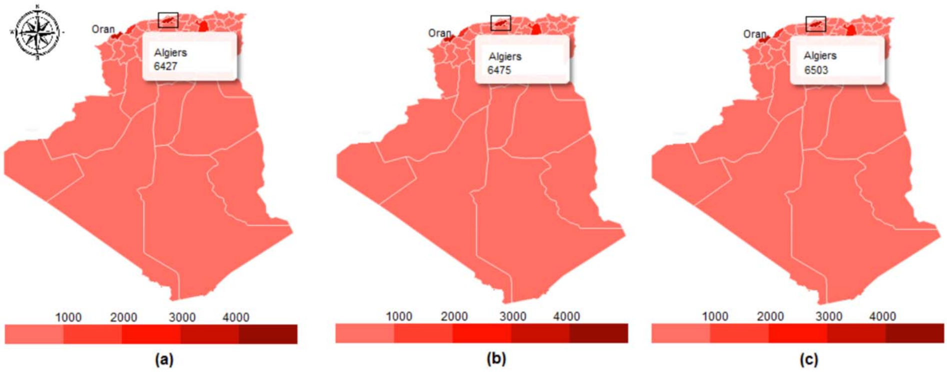


Figure 13: Prediction risk map of COVID-19 in Algeria wilayas.

Setif, Bouira, Khenchela, etc. The remaining wilayas were designated as low and very low incidence areas, respectively.

During the epidemic period, a total confinement was applied in the country to stop the epidemic spread. Over time and given the improved epidemic situation (decrease in the number of new cases) other measures were introduced. Partial confinement at home (from 8 p.m. until 5 a.m. the next day) was applied from Tuesday, October 1, 2020, for a period of two months for 32 wilayas, including Algiers and Oran. Closure during this period of the following activities (applicable in the 32 wilayas affected by partial home confinement): sports halls, places of pleasure, recreational and leisure areas, beaches, youth centers and cultural centers, sales markets as well that the prohibition, throughout the national territory, of any type of gathering of people and family reunification. Moreover, to control the epidemic situation in the country, a weekly coordination meeting on COVID-19 was held between the WHO-Algeria office and the WHO regional office (AFRO) every Monday.

7. Conclusion

The COVID-19 is a viral respiratory infection which broke out in December 2019 in China. It is one of the viruses that spread rapidly with high mortality risks among people; it is very contagious and causes different symptoms. This leads to a need to maintain and to strengthen the use of technologies and artificial intelligence in the epidemiology field. In this context, several studies and research have been carried out in order to control and monitor the disease spread and predict its future course.

Indeed, the aim of this work is to introduce a new spatiotemporal prediction system based on machine learning algorithms. We investigated four prediction models: LSTM, SLSTM, ARIMA and Prophet Models, to improve COVID-19 predicting accuracy. In the current study, we used a real dataset obtained from Algeria Coronavirus Tracker API[2] for COVID-19 data in Algeria. The latter was randomly divided into training (70%) and testing (30%) sets. To analyse the performance of the proposed COPSYS system, we examined our prediction models using performance metrics (RMSE, MAE, MSE, $\mathbf{R}^2$, CC and MAPE). The results showed that the Prophet model can well forecast the future course of COVID-19 disease most accurately among the other models. Moreover, we created prediction risk maps to help decision makers in their process control and to facilitate better understanding of the epidemic evaluation in the time and the space.

There are several limitations to our proposed models. First, the SLSTM gives the same outputs as LSTM. This is due to the small size of the used dataset. Deep learning models combined by stacking technique are suitable for small-size datasets. Therefore, we can discuss the new model construction. Moreover, It is necessary to predict the numbers of new cases of COVID-19, as well as recoveries and deaths. In the future, we aim to extend COPSYS system for predicting the three patient health situations during the disease. Further improvements and extensions can be made based on this study. The model will be improved towards identifying the leading risk factors that put persons at increased risk. The model will be based on machine learning algorithms using clinical features, demographic and socio-economic data. Our novel COPSYS system will enhance the benefit of monitoring and prediction for ongoing secondary outbreaks of COVID-19 or similar future outbreaks of other emergent infectious diseases.

References

[1] K. H. Almotairi, A. M. Hussein, L. Abualigah, S. K. M. Abujayyab,E. M. Mahmoud, B. O. Ghanem and A. H. Gandomi, Impact of artificial intelligence on COVID-19 pandemic: a survey of image processing, tracking of disease, prediction of outcomes, and computational medicine. *Big Data and Cognitive Computing* **7** (2023), 11.

[2] R. Pal, A. A. Sekh, S. Kar, and D. K. Prasad, Neural network based country wise risk prediction of COVID-19. *Applied Sciences* **18** (2020), 6448.

[2] https://api.corona-dz.live/province/all

[3] F. Z. Younsi and D. Hamdadou, Prediction system-based community partition for tuberculosis outbreak spread. *International Journal of Information Technologies and Systems Approach* **15** (2022), 1–20.

[4] A. Heidari, N. Jafari Navimipour, M. Unal and S. Toumaj, Machine learning applications for COVID-19 outbreak management. *Neural Computing and Applications* **34** (2022), 15313.

[5] I. H. Bennacef and R. Benamirouche, La performance du système d'information sanitaire comme préalable à la prise de décisions de santé. *Revue Académique des Études Sociales et Humaines* **14** (2022), 15.

[6] J. Luo, Z. Zhang, Y. Fu and F. Rao, Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms. *Results in Physics* **27** (2021), 104462.

[7] P. Arora, H. Kumar and B. K. Panigrahi, Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, Solitons & Fractals* **139** (2020), 110017. MR4114870

[8] H. Alabdulrazzaq, M. N. Alenezi, Y. Rawajfih, B. A. Alghannam, A. A. Al-Hassan and F. S. Al-Anzi, On the accuracy of ARIMA based prediction of COVID-19 spread. *Results in Physics* **27** (2021), 104509.

[9] V. Poleneni, J. K. Rao and S. A. Hidayathulla, COVID-19 prediction using ARIMA model. In: *11th International Conference on Cloud Computing, Data Science & Engineering*, India, pp. 860–865 (2021).

[10] S. Tiwari, P. Chanak and S. K. Singh, A review of the machine learning algorithms for COVID-19 case analysis. *IEEE Transactions on Artificial Intelligence* (2022).

[11] A. E. Eltoukhy, I. A. Shaban, F. T. Chan and M. A. Abdel-Aal, Data analytics for predicting COVID-19 cases in top affected countries: observations and recommendations. *International Journal of Environmental Research and Public Health* **19** (2021), 7080.

[12] J. Devaraj, R. M. Elavarasan, R. Pugazhendhi, G. M. Shafiullah, S. Ganesan, A. K. Jeysree and E. Hossain, Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant? *Results in Physics* **21** (2021), 103817.

[13] S. Kanwal, F. Khan, S. Alamri, K. Dashtipur and M. Gogate, COVID-opt-aiNet: A clinical decision support system for COVID-19 detection. *International Journal of Imaging Systems and Technology* (2022).

[14] M. O. Alassafi, M. Jarrah and R. Alotaibi, Time series predicting of COVID-19 based on deep learning. *Neural Networks* **18** (2005), 602–610.

[15] A. Graves and J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neurocomputing* **468** (2022), 335–344.

[16] Y. Yu, X. Si, C. Hu and J. Zhang, A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation* **31** (2019), 1235. MR3988464

[17] A. Graves, A. R. Mohamed and G. Hinton, Speech recognition with deep recurrent neural networks. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6645–6649 (2013).

[18] C. C. Wei, Development of stacked long short-term memory neural networks with numerical solutions for wind velocity predictions. *Advances in Meteorology* (2020).

[19] G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, *Time Series Analysis: Forecasting and Control*. John Wiley & Sons (2015). MR3379415

[20] M. Nkongolo, Using ARIMA to predict the growth in the subscriber data usage. *Eng* (2023), 92–120.

[21] N. N. Sánchez-Pozo, S. Trilles-Oliver, A. Solé-Ribalta, L. L. Lorente-Leyva, D. Mayorca-Torres and D. H. Peluffo-Ordóñez, Algorithms air quality estimation: A comparative study of stochastic and heuristic predictive models. In: *International Conference on Hybrid Artificial Intelligence Systems*, pp. 293–304 (2021).

[22] G. Rafferty, Forecasting time series data with facebook prophet: Build, improve, and optimize time series forecasting models using the advanced forecasting tool. *Packt Publishing Ltd.* (2021).

[23] O. Renaud and M. P. Victoria-Feser, A robust coefficient of determination for regression. *Journal of Statistical Planning and Inference* **140** (2010), 1852–1862. MR2606723

[24] N. N. Sánchez-Pozo, S. Trilles-Oliver, A. Solé-Ribalta, L. L. Lorente-Leyva, D. Mayorca-Torres and D. H. Peluffo-Ordóñez, Algorithms air quality estimation: A comparative study of stochastic and heuristic predictive models. In: *International Conference on Hybrid Artificial Intelligence Systems*, pp. 293–304. Springer, Cham (2021).

[25] J. Hale, Which evaluation metric should you use in machine learning regression problems? *Medium* (2021).

[26] F. Z. Younsi, A. Bounnekar, D. Hamdadou and O. Boussaid, Integration of multiple regression model in an epidemiological decision support system. *International Journal of Information Technology & Decision Making* **18** (2019), 1755–1783.

Fatima Zohra Younsi
LIO Laboratory
University of Oran 1
Oran
Algeria
Department of Computer Science
University of Mostaganem
Mostaganem
Algeria
E-mail address: fatima.younsi@univ-mosta.dz

Mohammed Chems Eddine Sahinine
Department of Computer Science
University of Mostaganem
Mostaganem
Algeria
E-mail address: sahinine.math@gmail.com

Ilyes Benarroum
Department of Computer Science
University of Mostaganem
Mostaganem
Algeria
E-mail address: iliesbenarroum@gmail.com

Received March 22, 2023