scientific reports

Check for updates

OPEN

Novel cost-effective method for forecasting COVID-19 and hospital occupancy using deep learning

Nabil I. Ajali-Hernández ✉ & Carlos M. Travieso-González

The emergence of the COVID-19 pandemic in 2019 and its rapid global spread put healthcare systems around the world to the test. This crisis created an unprecedented level of stress in hospitals, exacerbating the already complex task of healthcare management. As a result, it led to a tragic increase in mortality rates and highlighted the urgent need for advanced predictive tools to support decision-making. To address these critical challenges, this research aims to develop and implement a predictive system capable of predicting pandemic evolution with accuracy (in terms of Mean Absolute error (*MAE*), Root Mean Square Error (*RMSE*), R^2, and Mean Absolute Percentage Error (*MAPE*)) and low computational and economic cost. It uses a set of interconnected Long Short Term-memory (LSTM) with double bidirectional LSTM (BiLSTM) layers together with a novel preprocessing based on future time windows. This model accurately predicts COVID-19 cases and hospital occupancy over long periods of time using only 40% of the set to train. This results in a long-term prediction where each day we can query the cases for the next three days with very little data. The data utilized in this analysis were obtained from the "Hospital Insular" in Gran Canaria, Spain. These data describe the spread of the coronavirus disease (COVID-19) from its initial emergence in 2020 until March 29, 2022. The results show an improvement in *MAE* (< 161), *RMSE* (< 405), and *MAPE* (> 0.20) compared to other studies with similar conditions. This would be a powerful tool for the healthcare system, providing valuable information to decision-makers, allowing them to anticipate and strategize for possible scenarios, ultimately improving public health outcomes and optimizing the allocation of healthcare and economic resources.

Keywords COVID-19, Covid prediction, Daily covid, Hospital occupancy, LSTM

The global pandemic of COVID-19 currently accumulates more than 770 million cases[1]. Although it seems under control, Coronaviruses are susceptible to genetic mutation, which can cause a high number of cases in a short period of time, putting pressure on hospitals and healthcare systems.

Figure 1 shows the historical number of cases of COVID-19. It can be seen how different peaks of COVID-19 strains have rapidly reached high levels of weekly cases in very short periods. This situation could lead to a new global crisis like the one we experienced from 2020 to 2022.

Artificial Intelligence (AI) and all its branches, such as deep and machine learning algorithms, neural networks, and others, are being intensively explored for new healthcare applications in areas such as diagnostic imaging, lifestyle management, early detection of neurological and neurodegenerative diseases, risk analysis and monitoring, health information management, rehabilitation, and virtual healthcare. The expected benefits in these areas are broad and include optimization of diagnostics, monitoring, and control of disease progression, increased speed in imaging, greater understanding of predictive screening, and reduced costs and inefficiencies in healthcare[2].

In this work, we focus on achieving a new methodology based on combining artificial intelligence algorithms related to time-series and attention layers to predict different daily caseload ranges. In this way, we hope that hospitals around the world will have access to a low-cost technology capable of predicting each day's occupancy for the next three days. This gives them some leeway and reduces pressure and stress on the system.

Signals and Communications Department (DSC), University of Las Palmas de Gran Canaria, Campus Universitario de Tafira, 35017 Las Palmas de Gran Canaria, Spain. ✉email: nabil.ajali101@alu.ulpgc.es

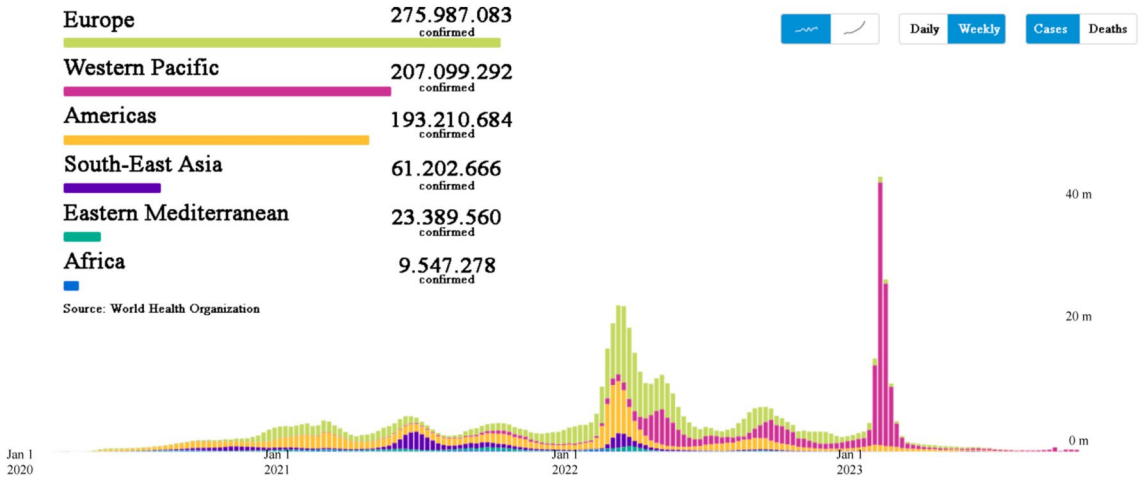

Figure 1. Evolution of the different covid waves from 2020 to the present day[1].

Related works

In the tough battle against the COVID-19 pandemic, the scientific community has joined forces in an unprecedented quest to forecast the trajectory of the pandemic. The arsenal of predictive models, empowered by the skills of AI, has played a critical role in providing insights, anticipating spikes, and guiding critical decisions. This has led the scientific and medical community to a favorable scenario, in a very short time, where AI models are part of every step of the planning and control of everything related to COVID-19 in hospitals. This section shows a selection of some of the most notable studies around the world that have successfully achieved a daily new case forecast or Covid case trend rigorously.

Nayet et al.[3] provide in their work very valuable information about the artificial intelligence models applied in the field of COVID-19 prediction cases. As can be seen in Fig. 2, 52% of the works apply deep learning in their predictions, while the rest of the works focus on machine learning and other mathematical models. In addition, it can be seen in Fig. 3, how in deep learning applications, the most used are Convolutional Neural Networks (CNN) and Generative Adversarial Network (GAN). Being LSTM and BiLSTM a small part of the set.

Pontoh et al.[4] conducted a study from March 2020 to April 2022 to decipher the impact of the pandemic on Jakarta. Using a Neural Network Autoregressive Model (NNAM), their goal was to identify the patterns of positive COVID-19 cases. The results showed a MAE of 302.02, a *RMSE* of 458.51, and an impressive R^2 value of 0.941.

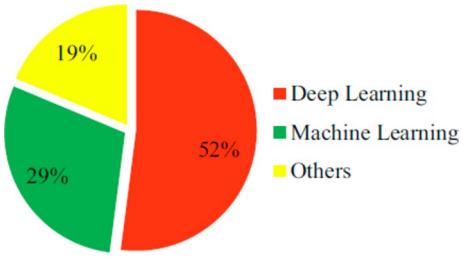

Figure 2. Publication trends in ML, DL, and other techniques for COVID-19 prediction.

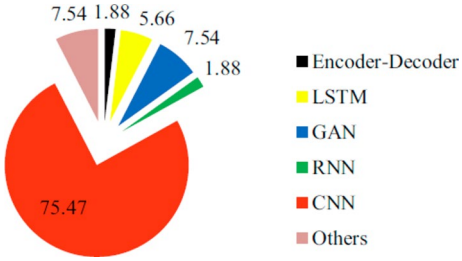

Figure 3. DL techniques used for COVID-19 prediction.

In an extensive study spanning from April 2020 to March 2023, Jin et al.[5] explored the integration of Autoregressive Integrated Moving Average (ARIMA), LSTM, and Back Propagation Neural Network (BPNN) models to forecast COVID-19 trends in Japan and Germany. Their research aimed to provide insights into the pandemic's future trajectory. The results were promising, revealing an *MAE* of 1249.83, an *RMSE* of 2478.28, and an *MSE* of 6,141,895.96. The results show such high values, probably due to the steep slope of the curves and the high gradients resulting from the high number of cases in the different pandemic waves.

On the other hand, Jin et al.[6] proposed a comprehensive analysis in which they aimed to predict the course of the COVID-19 pandemic. For this purpose, they developed an AI model based on ARIMA-LSTM fusion. The duration of this work was from January 2021 to October 2022, covering the territories of China and India. Their efforts yielded promising results, with an *MAE* of 580.35, an *RMSE* of 863.08, an *MSE* of 744.90, and an impressive R^2 value of 0.983.

Moving to the African continent, in the heart of Cameroon, Sandie et al.[7] conducted a large study from March 2020 to May 2021, using a seasonal ARIMA model to uncover trends in COVID-19 infection cases. Their investigation provided valuable information, with an *MAE* of 11,156.77, an RMSE of 18,824.93, and a remarkable *MAPE* of 12.35. During these dates, in the United States, Zhou et al.[8] also conducted an in-depth analysis from April 2020 to April 2021, leveraging an interpretable temporal attention network (ITANet) to make predictions of new confirmed COVID-19 cases. They obtained fairly good results with an *MAE* of 294.46, an *RMSE* of 392.49, and a *MAPE* of 0.1215.

On the other side of the world, in Malaysia, Singh et al.[9] achieved quite acceptable predictions using local hospital data from 22 January to 1 May 2022. They used an ARIMA model that successfully predicted the number of daily cases. Their research revealed a *MAPE* of 16.01 and a p-value of 0.186.

Last but not least, from 15 February 2020 to 22 January 2022, Hssayeni et al.[10], used LSTM for daily forecasting of COVID-19 cases in the USA. Their pioneering work revealed an MAE of 596.66, an R^2 value of 0.82, and a remarkable p-value of 0.0027.

These remarkable discoveries in the field of COVID-19 predictive modeling reflect the unwavering dedication of researchers around the world, who contribute substantially to our collective understanding of this unprecedented global challenge. *MAE* and *RMSE* values tend to show remarkable differences as explained in many studies, due to the steep slope that the contagion curves reached. A detailed summary of the related work section is shown in Table 1. It shows the model used in the study, the dates on which they were carried out, and the relevant results.

Our proposal

Building on previous research, we propose an innovative approach to improve the accuracy and efficiency of COVID-19 prediction models with a small number of training samples.

Our methodology introduces a novel perspective, emphasizing the use of an interconnected LSTM layer with a bidirectional double-layer BiLSTM as the main element of the architecture. This, together with an optimized architecture and a data preprocessing technique not used in previous work leads us to a model that breaks the current limitations shown in the previous section (Figs. 2, 3, and Table 1) and improves on the current state of the art.

This approach consists of segmenting the data into temporal windows with labels of future cases. This facilitates predictive capabilities using daily historical COVID-19 cases as the only input. In addition, a reduction of the training set is achieved, requiring only 40% of the ensemble data to train, with good *MAE,* R^2, and *RMSE* results. This implies that for the nature of the type of data being collected, only an interval of between 6 and 12 months is needed to have a training set ready to deliver good results. The main objective of our proposal is to provide a reliable COVID-19 prediction method with low computational cost, efficient parameters, and the need for simple input. This improvement addresses the limitations observed in previous models, such as ARIMA and BPNN, by exploiting the full potential of the sequential learning capabilities of LSTM and BiLSTM.

In addition, the computational cost associated with our model is significantly lower than some existing approaches. Our method reduces the dependence on a large number of input parameters, making it less data-intensive. Thus, its cost-effectiveness makes it accessible for use in a variety of environments and regions, ensuring its practical applicability.

Study	Dataset	Architecture (layers)	Estimated parameters
Pontoh et al.[4]	1 Mar 2020 to 3 Apr 2022. Jakarta	Neural network autoregressive	Positive cases
Pontoh et al.[4]	1 Mar 2020 to 3 Apr 2022. Jakarta	LSTM	Positive cases
Jin et al.[5]	1 Apr 2020 to 9 Mar 2023. Japan and Germany	ARIMA-LSTM-BPNN	COVID-19 prediction
Jin et al.[6]	1 Jan 2021 to 10 Oct 2022. China and India	ARIMA-LSTM	COVID-19 prediction
Sandie et al.[7]	5 Mar 2020 to 31 May 2021. Cameroon	Seasonal ARIMA model	Infection cases COVID-19
Zhou et al.[8]	1 Apr 2020 to 28 Apr 2021. USA	Interpretable temporal attention network (ITANet)	New confirmed cases
Singh et al.[9]	22 Jan to 1 May 2022. Malaysia	ARIMA	Forecasting daily cases
Hssayeni et al.[10]	15 Feb 2020 to 22 Jan 2022. USA	LSTM	Forecasting daily cases

Table 1. Summary of related works and their conditions.

Material and methods

Dataset

In the field of professional science, our study focuses on the use of a regression model to predict COVID-19 trends using data from the Hospital Insular de Gran Canaria (Spain). This dataset spans from the beginning of 2020 to March 29, 2022, and consists of only two inputs in the simplest case: date and daily new COVID-19 cases. Despite the simplicity of this dataset, our analysis has demonstrated the exceptional ability of the model to accurately predict future COVID-19 trends, identifying temporal patterns, seasonality, and the impact of interventions. This work underscores the value of accessible data and demonstrates how even minimal data inputs can yield profound insights, revolutionizing the landscape of professional research and analysis in the field of science. As mentioned above, the database is owned by the Government of the Canary Islands (Spain) and the data is public[11]. It can be consulted or downloaded from https://opendata.sitcan.es/dataset/capacidad-asistencial-covid-19.

Performance indices

A set of statistical parameters has been used to evaluate the accuracy of the model. The selection of these parameters is based on their widespread use in the literature, which allows us to compare our results with the current state of the art. The most prominent are *RMSE, MAE, MAPE,* and R^2, which measure the precision of the measurements, as well as their dispersion and correlation. Their mathematical expressions are shown in the following equations, where y_i are the observed values, $\widehat{y_i}$ the predicted values and $\overline{y}$ the mean of these values respectively[12–15].

Mean square error (*MSE*):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2 \tag{1}$$

Root mean square error (*RMSE*):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2} \tag{2}$$

Mean average error (*MAE*):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y_i}\right| \tag{3}$$

Mean square error (*MAPE*):

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{\left|y_i - \widehat{y_i}\right|}{y_i} \tag{4}$$

Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\widehat{y_i} - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{5}$$

Data preprocessing

To perform the data preprocessing and labeling, the "new daily cases" variable was separated into one vector and the date variable into another vector. Then, a labeling window with different values was used to assign a label to the "new daily cases" values. This label assigns the values of "Ytrain". These "Ytrain" values depend on the size of the window. Thus, for a window n = 2, the "new daily cases" of date n = 1 and n = 2 would be grouped in the first row of the "Xtrain" vector and their "Ytrain" value would be that of the later date, i.e., n + 1. Next, considering a step = 1, the dates n = 2 and n = 3 would be grouped in the second row, and the value of "Ytrain" would be that of n = 4.

The study was carried out with values from n = 1 to n = 20 to check which window was better suited to the data and which could better handle the high slope presented by the COVID-19 waves. Figure 4 shows a scheme of the starting vectors "date" and "new daily cases" and the labeling process for a window n = 2 and n = 5. The dataset is available from the link provided in the previous section.

Network architecture

To correctly predict the COVID-19 data, an architecture has been designed that is capable of analyzing the time series and capturing the existing gradient differences in the slopes generated by the different waves through the use of deep learning. The different layers used in the whole architecture are described in detail below.

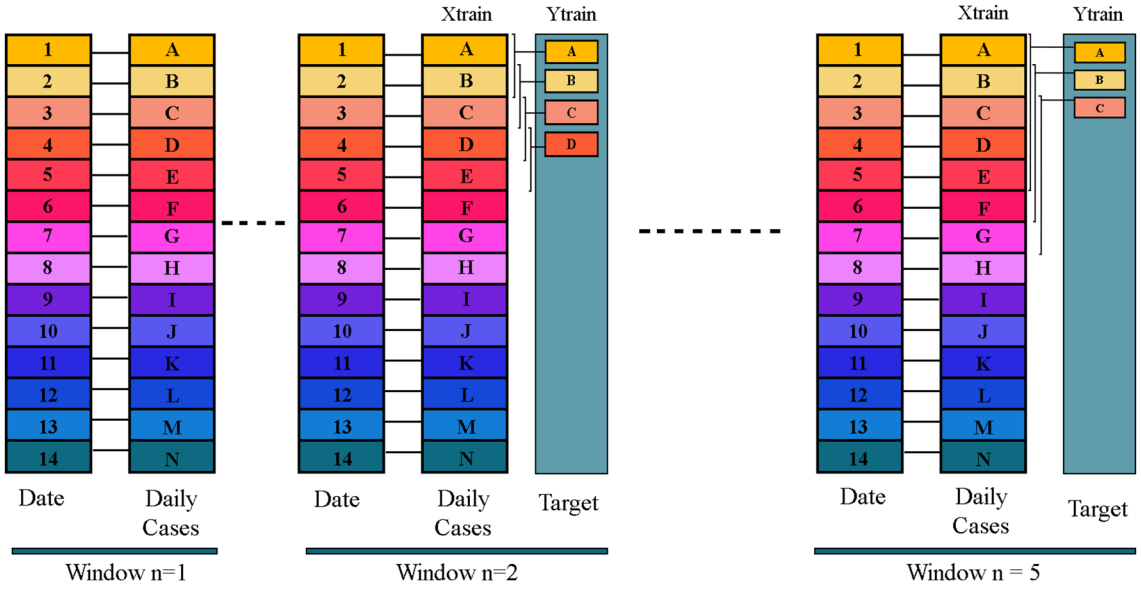

Figure 4. Scheme of the preprocessing and targeting process.

LSTM-BiLSTM

Long-Short Term Memory is a type of recurrent neural network (RNN) that is particularly useful for modeling sequential data. This type of algorithm has been applied to a wide range of tasks, including speech recognition, natural language processing, and time series forecasting[16]. By using memory cells, LSTMs can retain useful data from current or previous stages and use it in the future. Therefore, they use algorithmic gates that are also capable of retaining such information for future use and goal attainment.

They can also be combined to improve the overall network architecture. There are variants with different functions, such as Bidirectional LSTMs (BiLSTMs), Gate Recurrent Units (GRUs), or the new algorithms focused on the attention layer, called "transformers", described by Vaswani et al. in 2017 in their work entitled "Attention is all you need"[17]. In the case of BiLSTMs, the only difference is the relationship between the states, since they are bidirectional and can take into account the data of the previous state as well as the following one.

The LSTM consists of a memory cell and three parts, which can be expressed mathematically as follows[18]:

Input Gate: the layer responsible for updating the state of the network through the sigmoidal function.

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

W_i is the representation of the weight of the input, b_i is the corresponding bias, x_t is the current time step, and h_{t-1} is the output of the previous time step. σ will have a value $\in [0, 1]$, representing full discard or full save of the data, respectively[16,18].

Forget Gate: the layer responsible for deciding whether to save or discard the information. It is the first step of LSTM.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{7}$$

W_f is the weight representation of the input, b_f is the corresponding bias, x_t is the current time step and h_{t-1} is the output of the previous time step.

Output Gate: This is where the information output is determined. This output is based on the filtered version of the cell state. The output value is determined by the sigmoid layer and then multiplied by the cell state[18].

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

$$h_t = o_t \cdot \tan h(C_t) \tag{9}$$

W_o is the weight representation of the input, b_o is the corresponding bias, and x_t is the current time step. h_{t-1} is the output of the LSTM layer at the current time step. Finally, the previous cell state C_{t-1} must be updated. This is computed by forgetting one input gate, as shown in Fig. 5.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot g_t \tag{10}$$

where g_t is the tanh layer.

In addition, the BiLSTM model is composed of two LSTM networks and is capable of reading input evaluations in both directions, forward and backward. The forward LSTM processes information from left to right, while the backward LSTM processes information from right to left[19].

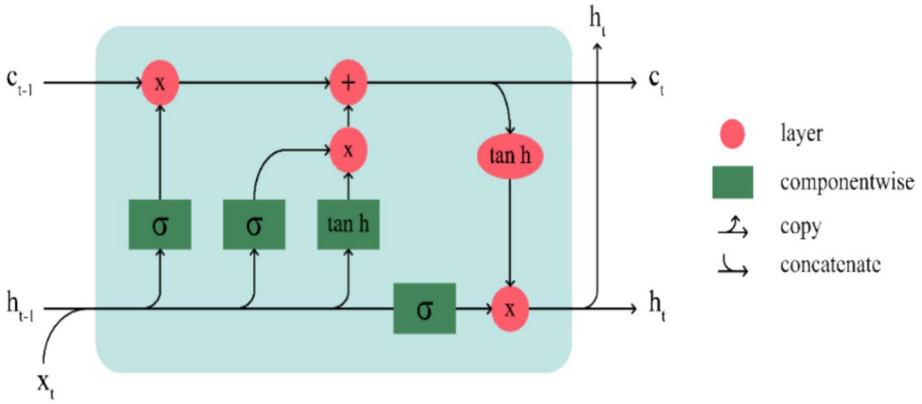

Figure 5. Representation of the components of a LSTM cell.

Dense layer

A dense or fully connected layer, also known as a fully connected feedforward neural network, is a type of artificial neural network in which each neuron in one layer is connected to each neuron in the next layer. The basic formula for a fully connected neural network with a hidden layer and an output layer (y_{fc}) can be represented as follows[20]:

$$y_{fc} = f\left(\sum_{i=1}^{n}(W_i * x_i) + b\right) \tag{11}$$

where x_i is the input vector to the network and W_i are the weight matrices for the connections between layers. b is the bias and f, is the activation function applied to the output of each layer (sigmoid, ReLU, tanh).

It is important to note that this formula is for a single hidden layer neural network, but in practice, fully connected neural networks usually have multiple hidden layers, in which case the formula would be more complex and would include additional weight matrices and bias vectors for each additional layer.

Dropout

Dropout is a regularization technique used in deep learning to avoid overfitting. It works by randomly "dropping" (i.e., setting to zero) a certain number of neurons during each training iteration. The mechanism of a dropout layer is quite simple: it is applied to the output of the previous layer and consists of multiplying the input vector by a mask. This mask is a binary mask that is randomly generated for each training iteration, it has the same shape as the input and each element is either 0 or 1. The probability that each element of the mask is 1 is called the dropout rate. The dropout rate is a hyperparameter that is usually set between 0.2 and 0.5, depending on the specific application and the complexity of the model. Typically, a low dropout rate is used for the input layer and a higher dropout rate is used for the hidden layers. During the testing phase, it is common to use a dropout rate of 0, which means that all neurons are active. This is because dropout is only applied during the training phase and is not used during the testing phase[21].

Hyperparameters

The network was trained and tested using Python's TensorFlow. The adaptive moment estimation or Adam method, which is a widely used optimization algorithm for neural network training, is used. Adam combines techniques from the RMSprop and Momentum optimizers to efficiently and effectively adjust the weights of a neural network during training. See Eqs. (12)–(14) below[22,23]:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \tag{12}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \tag{13}$$

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{v_t}+\epsilon} m_t \tag{14}$$

where m_t is the first updated moment (mean), v_t is the second updated moment (variance), β_1 and β_2 are the moment decay parameters, g_t is the gradient at the current step, α is the learning rate, "epsilon" (ϵ) is a small numerical constant to avoid division by zero, and θ_t is the current value of the parameter being updated. This is the parameter that the algorithm optimizes.

The values of these hyperparameters were $1{\cdot}10^{-6}$ for "Epsilon" in the training options and $1{\cdot}10^{-4}$ for the learning rate. The batch size was set to 5 and 15, with epochs equal to 1000 and a "shuffle" in each epoch. The value of β_1 was set to 0.99, and β_2 to 0.999. Training and testing are performed using the holdout method for regressions with a training percentage test of 40–60.

Complete network architecture

The architecture developed in this work consists of the sequential input of the previously defined temporal windows. This sequence passes through 3 levels. In the first one, there is an LSTM layer with 128 units in its hidden layer and with sequence return enabled. Then, in levels 2 and 3, there are 2 BiLSTM layers with sequence return enabled and 128 units each. Finally, at the output of this last level, a dense fully connected layer with 128 connections is implemented. Then, to reduce the randomness of the weights, a Dropout layer with a value of 0.4 and a Flatten layer are added to flatten the output sequence into a vector[24]. Finally, a dense layer with one neuron and linear activation is added to obtain the output. In Fig. 6, a scheme of the entire architecture is shown.

Results and discussion

Different strategies have been used in the design of this study. Then, the results are shown and we proceed to compare the different models used with their architectures. Also, a discussion about the chosen model is made and the best results obtained in the prediction of COVID-19 for this model are shown.

Temporal windows

To determine the number of windows to be used in the preprocessing, an analysis is performed under similar conditions, varying only the number of data to be grouped. A double LSTM with a 75–25 training-test set is used and the prediction error results are compared to obtain the optimal window. Table 2 shows the obtained results, where n is the size of the window.

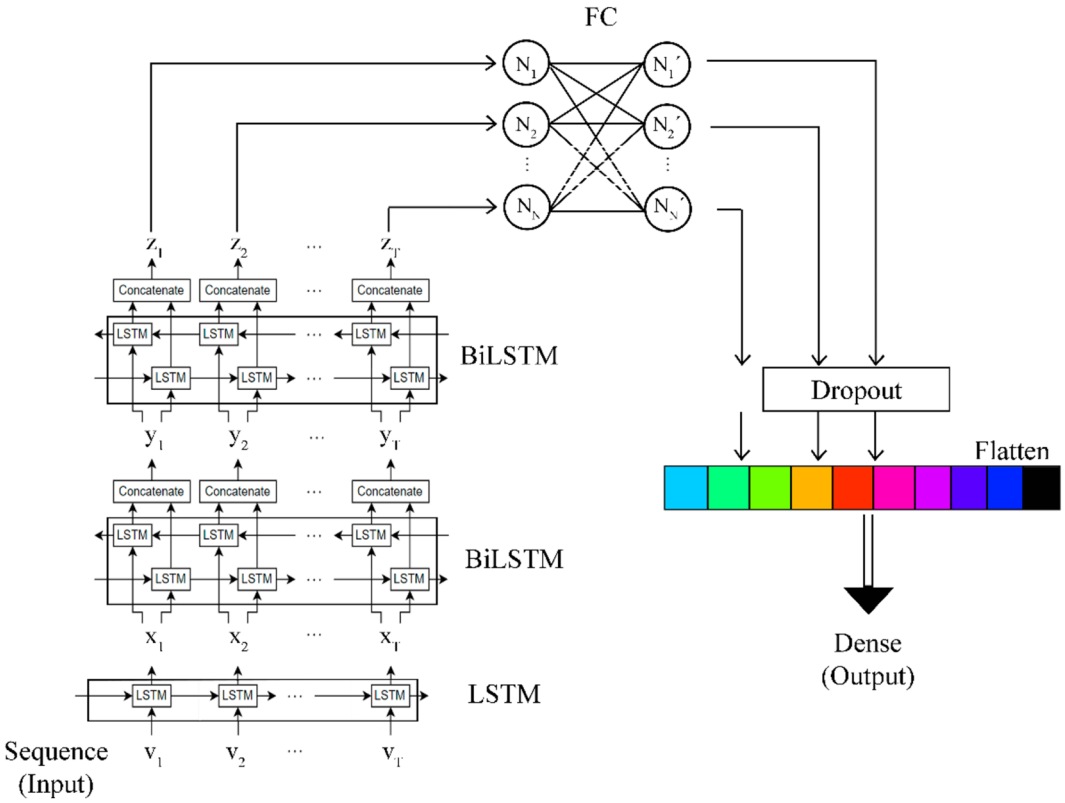

Figure 6. Scheme of the implemented network.

Parameters	n = 3	n = 5	n = 8	n = 10	n = 15
MSE	762,778.11	862,210.50	**481,293.11**	608,270.42	1,287,370.01
RMSE	873.20	928.55	**693.75**	779.36	1134.15
MAE	570.57	653.51	**334.34**	530.25	819.42
R^2	0.7202	0.6720	**0.7967**	0.7646	0.6147
MAPE	0.3392	0.3848	**0.2834**	0.3135	0.4752

Table 2. Comparison of the results of a classification by varying only the input window. Significant values are given in bold.

There is a clear tendency to obtain better results with an 8-sample window, i.e., a weekly period. This is because infections tend to vary from week to week, not from day to day, as demonstrated by Docherty et al.[25].

Thus, the results of a week's new cases are labeled together, as shown in Fig. 4. This gives the system a high degree of stability and allows fluctuations in the gradient of the curves to be captured.

LSTM layers

Once the optimal number in the window was obtained, the use of the proposed network consisting only of LSTMs was compared by varying the number of layers of the network. Table 3 shows the results of performing the classification using 1, 2, 3, and 4 LSTM layers in series. The 75–25 ratio is used because it is a common ratio in this type of work, and it provides an indicative point from which to begin the search for the optimum point[3].

As can be seen in Table 3, as the number of LSTM layers increases, a lower error is obtained, the best case being 3 LSTM layers. However, after 3 layers there is no improvement, and the computational cost increases. Therefore, it was decided to add three LSTM layers and then continue with the training/testing ratio.

Other layers

The decision to use combinations of different LSTM layers instead of adding CNN or other layers is explained in "Related works", Figs. 2 and 3, where it can be seen that CNN layers are the most widely used. To use a different type of approach and to differentiate ourselves from the majority of works, we decided to follow the aforementioned path where we focus on recurrent layers. Since the percentage of "other" layers has a rather insignificant value (7.54%), simulations were made with other types of architectures, see Table 4. The results show that these layers did not significantly improve the model and are therefore not included in our work. It would be interesting to go further in the future with another work to study the influence of adding or modifying the layers of our architecture together with other types of layers such as transformers or other models.

Training and test set

A series of experiments are performed with the 3 LSTM layers, varying the train/test set to find the best ratio in the predictions. The results are shown in Table 5.

Parameters	LSTM layer (75–25)	LSTM 2 layers (75–25)	LSTM 3 layers (75–25)	LSTM 4 layers (75–25)
MSE	665,540.42	481,293.11	**410,887.56**	431,885.88
RMSE	815.14	693.75	**641.00**	338.81
MAE	556.00	334.34	**330.82**	330.82
R^2	0.7642	0.7967	**0.8264**	0.8176
MAPE	0.3516	0.2834	**0.2497**	0.2589

Table 3. LSTM's statistics in daily case prediction using 1, 2, 3, and 4 layers. Significant values are given in bold.

Architecture	LSTM and attention layer	GRU	Multi-layer perceptron (MLP)	LSTM-attention-MLP
MSE	695,431.12	> 100,000	1,587,740.24	737,851
RMSE	833.93	> 100,000	1260.00	858.15
MAE	590.33	> 100,000	863.15	568.93
R^2	0.7521	–	0.6541	0.7156
MAPE	0.3819	–	0.6045	0.4382

Table 4. Comparison of results using different types of architectures, n = 8.

Parameters	Training/test (%)					
	75–25	60–40	50–50	40–60	30–70	20–80
MSE	410,887.56	322,300.93	245,984.84	**219,317.31**	**193,612.13**	447,370.26
RMSE	641.00	567.71	495.96	**468.31**	**440.01**	668.85
MAE	330.82	269.86	211.31	**205.32**	**176.86**	254.00
R^2	0.8264	0.8112	0.8313	**0.8256**	**0.8259**	0.5496
MAPE	0.2497	0.2549	0.2415	**0.2498**	**0.2606**	0.3635

Table 5. Comparison of different training/test sets for a 3-layer LSTM architecture. Significant values are given in bold.

On the other hand, it is observed that after varying the train-test range from a percentage of 80–20% train-test to 20–80% train-test, two relevant results are obtained. Both for the 40–60% train-test range and for the 30–70% train-test range, the values are similar. While better *MAE* and *RMSE* values are obtained for the second set, a lower *MAPE* value is obtained for the first set. This means that the second system is better suited to abrupt changes in the curves than the first, and therefore can better capture the peaks in the COVID-19 waves, while on average the first case makes a smaller error in the prediction, being able to obtain better results in daily cases. Moreover, the value of R^2, a parameter that is particularly important in regressions, is identical in both cases and equal to 0.825[26]. Therefore, it was decided to choose a train-test ratio of 40–60%, since it is more common to have a more compensated set when adding other layers or changing the architecture[27].

LSTM vs BILSTM layers

After obtaining the appropriate number of layers and the optimal train-test set, we proceed to a final step. Then, taking into account the nature of the data and its time dependence, the idea of using BiLSTM algorithms is established. The result is compared after adding or replacing different LSTM layers with BiLSTM. All possibilities of adding or replacing 1–3 LSTM and 1–3 BiLSTM are obtained. The results are shown in Table 6.

As can be seen in Table 6, the best performance is achieved by the LSTM with double BiLSTM. It achieves a lower value of MAE (161.80), RMSE (404.53), MAPE (0.19), and a better regression (0.87). Figure 7 shows a comparison of the 3 best performances of each set; the 1-LSTM-1BiLSTM, the 2-LSTM-2BiLSTM, and the 1-LSTM-2-BiLSTM. The different metrics of these 3 sets are compared and the worst case is also used. These 4 scenarios and their RMSE, MAE, r-squared and MAPE values are shown in Fig. 7.

The graphs are followed by Table 7 comparing our work with that mentioned in "Related work".

It can be seen that this study covers a large data set and has the lowest *MAE* value compared to other studies related to the estimation of daily new cases of COVID-19. This low *MAE* indicates that, on average, the model predictions have a small magnitude of error compared to the actual cases. Each prediction tends to be closer to the actual value, suggesting that the model effectively captures trends and patterns in daily new cases of COVID-19.

On the other hand, this study has a lower *RMSE* than the other papers, except the paper by Zhou et al.[8]. Their performance is good, but they have a worse *MAE* and a much more limited database. Low *RMSE* values mean that the differences between the model predictions and the actual values are small and consistent. The *RMSE*, by penalizing larger errors, indicates that even cases where the model is wrong tend to be fairly close to reality.

In addition, the *MAPE* and R^2 values of this work are quite acceptable considering the extent of the data and the shape of the curve, since they cover 2 years and 2 months of data, i.e., the 3 waves of COVID-19. Therefore, this work is hardly comparable with the other studies that have used *MAPE* as a parameter[7–9], since they have data with a duration of only 1 year and are therefore less affected by the variations of the curves. Papers with an R^2 higher than this study have higher values for *RMSE* and *MAE* too and although they perform well in cases where there is a rapid growth of new daily cases, they do not respond as favorably as this study. Therefore, taking into account that the objective is to determine the new cases of COVID-19 to avoid collapse and manage the healthcare system, the set of results of this study makes it a model that from very little data can give a very accurate prediction and has a high adaptability in case of large changes in trend.

Finally, in Table 8 the model prediction parameters are included for 3 days ahead, i.e., the next day (N + 1) and the two following days (N + 2 and N + 3).

Parameters	3 LSTM + 1 BiLSTM	3 LSTM + 2 BiLSTM	3 LSTM + 3 BiLSTM
MSE	**164,677.98**	227,299.50	221,403.32
RMSE	**405.81**	476.76	470.54
MAE	**174.16**	198.51	199.91
R^2	**0.8691**	0.8193	0.8240
MAPE	**0.2165**	0.2194	0.2439
Parameters	**2 LSTM + 1 BiLSTM**	**2 LSTM + 2 BiLSTM**	**2 LSTM + 3 BiLSTM**
MSE	306,529.07	**167,790.98**	212,088.06
RMSE	553.65	**409.62**	460.53
MAE	218.83	**180.41**	191.667
R^2	0.7563	**0.8667**	0.8314
MAPE	0.2351	**0.2193**	0.2177
Parameters	**1 LSTM + 1 BiLSTM**	**1 LSTM + 2 BiLSTM**	**1 LSTM + 3 BiLSTM**
MSE	194,284.38	**163,647.00**	175,088.22
RMSE	440.78	**404.53**	418.44
MAE	191.60	**161.80**	164.95
R^2	0.8456	**0.8712**	0.8608
MAPE	0.2349	**0.1930**	0.2077

Table 6. Comparison of use of BiLSTM layers vs. LSTM layers. Significant values are given in bold.

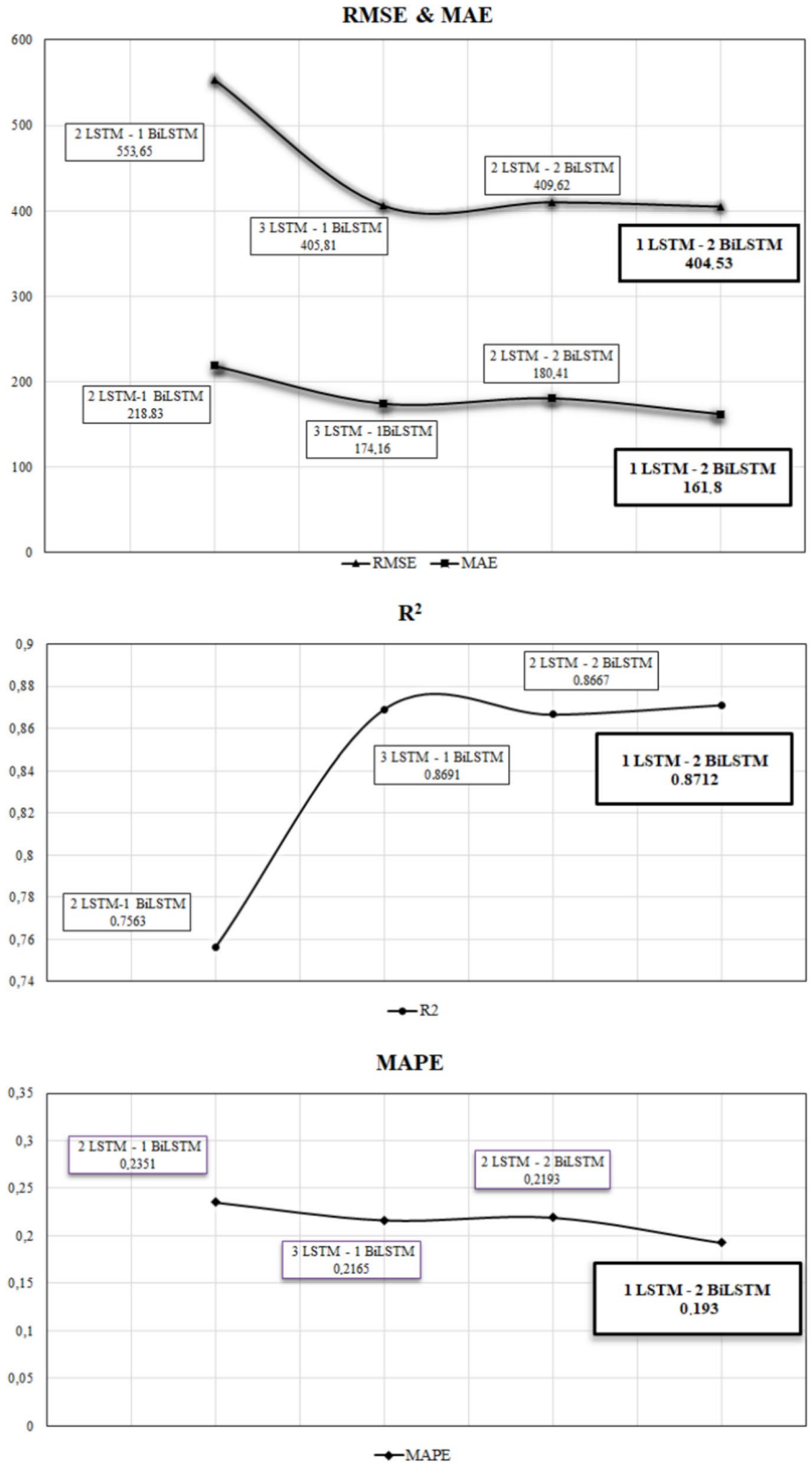

Figure 7. Model performance curves using *RMSE*, *MAE*, R^2 and *MAPE* in different conditions.

The N + 1, N + 2, and N + 3 values predict future daily Covid cases. As can be seen, the errors increase as we move away from the prediction horizon. This is logical since each time we increase the value of N, the error

Study	Architecture (layers)	Metrics
Pontoh et al.[4]	Neural network autoregressive	*MAE* = 302.02; *RMSE* = 458.51; R^2 = 0.941
Pontoh et al.[4]	LSTM	*MAE* = 462.43; *RMSE* = 964.63; R^2 = 0.9334
Jin et al.[5]	ARIMA-LSTM-BPNN	*MAE* = 1249.83; *RMSE* = 2478.28; *MSE* = 6,141,895.96
Jin et al.[6]	ARIMA-LSTM	*MAE* = 580.35; *RMSE* = 863.08; *MSE* = 744.90; R^2 = 0.983
Sandie et al.[7]	Seasonal ARIMA model	*MAE* = 11,156.77; *RMSE* = 18,824.93; *MAPE* = 12.35
Zhou et al.[8]	Interpretable temporal attention network (ITANet)	*MAE* = 294.46; *RMSE* = 392.49; *MAPE* = 0.1215
Singh et al.[9]	ARIMA	*MAPE* = 16.01; *p-value* = 0.186
Hssayeni et al.[10]	LSTM	*MAE* = 596.66; R^2 = 0.82, *p-value* = 0.0027
This study	1-LSTM-2-BiLSTM	*MAE* = 161.80; *RMSE* = 404.53; *MAPE* = 0.1930; R^2 = 0.8712

Table 7. Comparison of COVID-19 forecasting metrics from previous studies vs the proposed study.

Parameters	N + 1	N + 2	N + 3
MSE	**163,647.00**	198,511.32	246,272.67
RMSE	**404.53**	445.54	496.26
MAE	**161.80**	194.24	214.32
R^2	**0.8712**	0.8422	0.8149
MAPE	**0.1930**	0.2355	0.2509

Table 8. Metrics for predicting daily COVID-19 cases for days N + 1, N + 2, and N + 3. Significant values are given in bold.

and uncertainty also increase. Specifically, the MSE and RMSE values increase by 18% and 9% respectively between the prediction of day N + 1 and the prediction of day N + 2. This relationship is somewhat respected when the prediction range is increased, since from prediction N + 2 to prediction N + 3, the MSE and RMSE error values increase by 19% and 10%, respectively. The increase from N + 1 to N + 3 is 50% and 23% for these values, respectively.

In the case of MAE, something similar happens, but the values from N + 1 to N + 2 increase by 17%, and then from N + 2 to N + 3 the value increases more slowly, by only 9%. The total difference in MAE from N + 1 to N + 3 is 32%.

In the case of R^2, the value loses correlation linearly, as it simply seems to lose a 3% correlation with each day it is increased.

Finally, the MAPE value is more affected by the change from N + 1 to N + 2 than from N + 2 to N + 3, the first increase in MAPE being 18% and the second only 6%.

All these values show that as the horizon of prediction gets further away, there are still competitive forecast values, but they lose consistency. Table 9 summarizes these data.

In addition, Fig. 8 is included, which shows the good performance of the model in its predictions with a 90% confidence interval.

Conclusions

In the context of daily COVID case prediction in a hospital, time and accuracy are key factors to avoid system saturation and optimally manage patient volume and occupancy.

In this study, a deep learning model based on the union of LSTM and BiLSTM layers was developed to predict the daily number of cases from a large database. This database, with a duration of more than two years of daily cases in a hospital in Spain, allows predictions with only 2 parameters as inputs. Being of low computational cost and easy access.

Parameters	N + 1 to N + 2 (%)	N + 2 to N + 3 (%)	N + 1 to N + 3 (%)
MSE	18	19	50
RMSE	9	10	23
MAE	17	9	32
R^2	− 3	− 3	− 6
MAPE	18	6	30

Table 9. Metric changes between the statistical parameters of N + 1, N + 2 and N + 3.

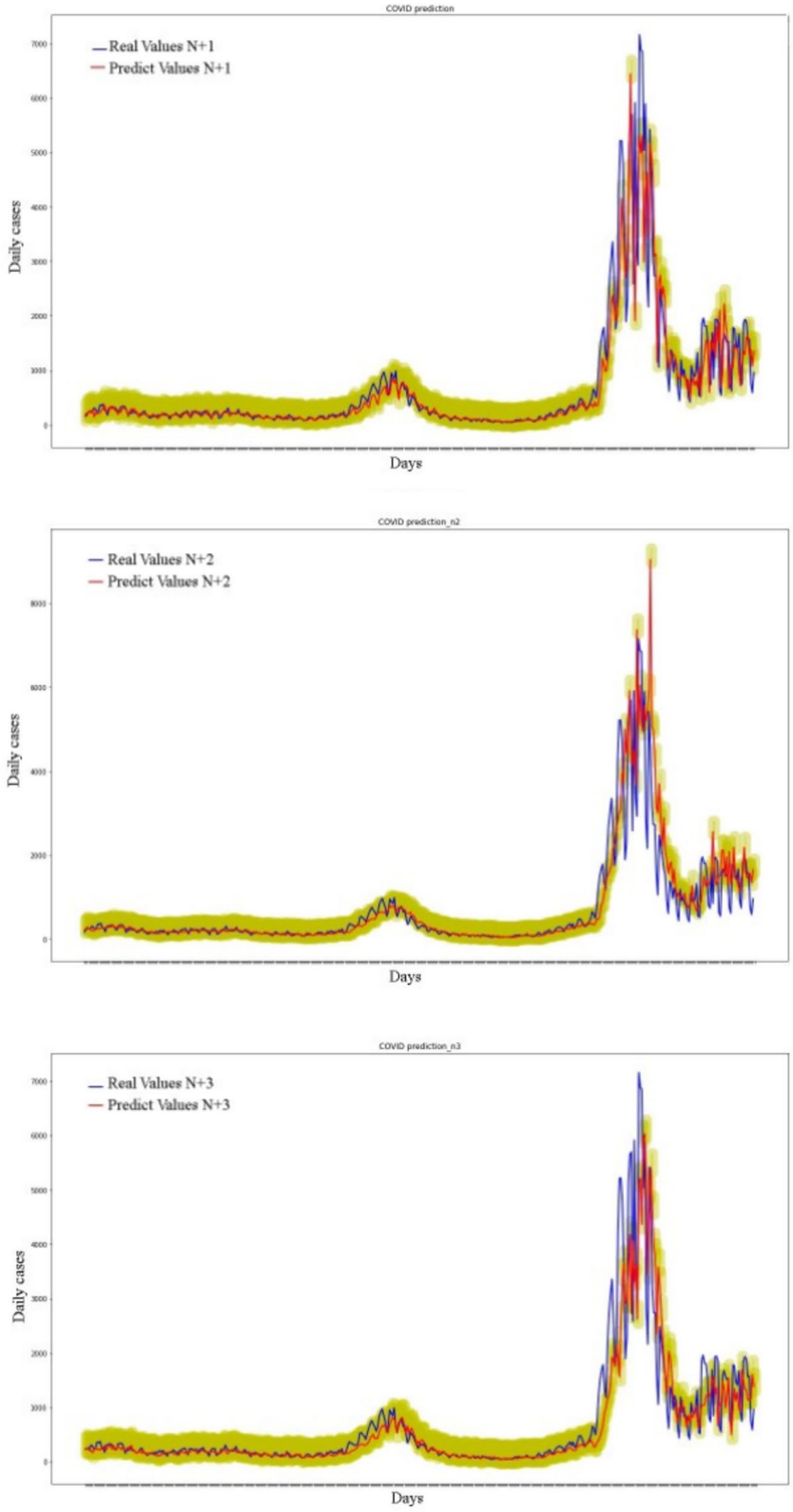

Figure 8. Evolution of N + 1, N + 2, and N + 3 predictions of daily COVID-19 cases with a 90% confidence interval.

The results show that the model is reliable and accurate in its predictions, since it achieves lower error values than similar works, with an MAE of 161 and an RMSE of 404. It also achieves good values of R^2 (0.8712) and MAPE (0.193), which are justified by the size of the database, since it includes the 3 waves of COVID-19, unlike many previous studies. In addition, the model is proposed to be able to make predictions for the next 3 days with acceptable margins of error.

All of this can have several positive effects, including accurate forecasting that can help the hospital better plan and allocate resources such as beds, medical staff, and supplies to meet expected caseloads. With these accurate forecasts, administrators and healthcare professionals can make more informed decisions about treatment strategies, preventive measures, and vaccine distribution.

In addition, the model's performance allowed for rapid response to changes. Because the model performs well, it is more likely to detect changes in case trends quickly, allowing for an early and efficient response to potential increases in COVID-19 incidence.

Future lines of research include increasing the complexity of the model by adding additional layers, such as transformations, and including variables such as new discharges and health parameters.

Data availability

The link to the datasets analyzed during the current study can be found in Section "Dataset", and is also available from the authors upon request.

Received: 5 April 2024; Accepted: 2 August 2024
Published online: 29 October 2024

References

1. World Health Organization [Internet]. 2023. WHO COVID-19 Dashboard. (World Health Organization, Geneva, 2020). Available from: https://covid19.who.int/.
2. Racine, E., Boehlen, W., & Sample, M. Healthcare uses of artificial intelligence: Challenges and opportunities for growth. In *Healthcare management forum*, pp. 272–275 (SAGE Publications Sage CA, Los Angeles, CA, 2019).
3. Nayak, J. *et al.* Intelligent system for COVID-19 prognosis: A state-of-the-art survey. *Appl. Intell.* **51**, 2908–2938 (2021).
4. Pontoh, R. S. *et al.* Jakarta pandemic to endemic transition: Forecasting COVID-19 using NNAR and LSTM. *Appl. Sci.* **12**(12), 5771 (2022).
5. Jin, Y. C. *et al.* Prediction of COVID-19 data using improved ARIMA-LSTM hybrid forecast models. *IEEE Access.* **1**, 1 (2023).
6. Jin, Y. *et al.* Prediction of COVID-19 data using an ARIMA-LSTM hybrid forecast model. *Mathematics.* **10**(21), 4001 (2022).
7. Sandie, A. B. *et al.* Observed versus estimated actual trend of COVID-19 case numbers in Cameroon: A data-driven modeling. *Infect. Dis. Model.* **8**(1), 228–239 (2023).
8. Zhou, B., Yang, G., Shi, Z. & Ma, S. Interpretable temporal attention network for COVID-19 forecasting. *Appl. Soft Comput.* **120**, 108691 (2022).
9. Singh, S. *et al.* Forecasting daily confirmed COVID-19 cases in Malaysia using ARIMA models. *J. Infect. Dev. Countries.* **14**(9), 971–976 (2020).
10. Hssayeni, M. D. *et al.* The forecast of COVID-19 spread risk at the county level. *J. Big Data.* **8**, 1–16 (2021).
11. Gobierno de Canarias. (s. f.). *Capacidad asistencial COVID-19 - SITCAN Open Data*2024. 2024 May; Available from: https://opendata.sitcan.es/dataset/capacidad-asistencial-covid-19.
12. Willmott, C. J. & Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **30**(1), 79–82 (2005).
13. Hyndman, R. J. & Koehler, A. B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **22**(4), 679–688 (2006).
14. Draper, N. R., & Smith, H. Applied regression analysis. Vol. 326 (John Wiley & Sons, 1998).
15. Steel, R. G. D., & Torrie, J. H. Principles and procedures of statistics: with special reference to the biological sciences (1960).
16. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997).
17. Vaswani, A. *et al.* Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, 1 (2017).
18. Van Houdt, G., Mosquera, C. & Nápoles, G. A review on the long short-term memory model. *Artif. Intell. Rev.* **53**, 5929–5955 (2020).
19. Hameed, Z. & Garcia-Zapirain, B. Sentiment classification using a single-layered BiLSTM model. *IEEE Access.* **8**, 73992–74001 (2020).
20. Heaton, J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618. *Genet. Program. Evol. Mach.* **19**(1–2), 305–307 (2018).
21. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014).
22. Kingma, D. P., & Ba, J. Adam: A method for stochastic optimization. arXiv preprint. arXiv:14126980 (2014).
23. Jesús. ¿Qué es un optimizador y para qué se usa en deep learning? DataSmarts Español [Internet]. 2020 Jul; Available from: https://datasmarts.net/es/que-es-un-optimizador-y-para-que-se-usa-en-deep-learning/.
24. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature.* **521**(7553), 436–444 (2015).
25. Docherty, A. B. *et al.* Features of 20 133 UK patients in hospital with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol: Prospective observational cohort study. *BMJ.* **369**, 1 (2020).
26. Chicco, D., Warrens, M. J. & Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **7**, e623 (2021).
27. Poldrack, R. A., Huckins, G. & Varoquaux, G. Establishment of best practices for evidence for prediction: A review. *JAMA Psychiatry.* **77**(5), 534–540 (2020).

Author contributions

Nabil I. Ajali-Hernández wrote the main manuscript text, prepared all figures, wrote and executed the experiments. Carlos M., Travieso-González reviewed the main manuscrit text, procured the database, supervised the project and contributed to the conceptualisation of the main text. All authors reviewed the manuscript.

Funding

This work was funded by "Agencia Canaria de Investigación, Innovación y Sociedad de la Información de la Consejería de Economía Conocimiento y Empleo y por el Fondo Social Europeo (FSE) Programa Operativo

Integrado de Canarias 2014-2020, Eje 3 Tema Prioritario 74 (85%)" from "Gobierno de Canarias" in Spain, under the reference "TESIS2020010118".

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to N.I.A.-H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.