COVID-19 Patient Count Prediction Using LSTM

Muhammad Iqbal, Feras Al-Obeidat, Fahad Maqbool, Saad Razzaq, Sajid Anwar, Abdallah Tubaishat, Muhammad Shahrose Khan, and Babar Shah

***Abstract*— In December 2019, a pandemic named COVID-19 broke out in Wuhan, China, and in a few weeks, it spread to more than 200 countries worldwide. Every country infected with the disease started taking necessary measures to stop the spread and provide the best possible medical facilities to infected patients and take precautionary measures to control the spread. As the infection spread was exponential, there arose a need to model infection spread patterns to estimate the patient volume computationally. Such patients' estimation is the key to the necessary actions that local governments may take to counter the spread, control hospital load, and resource allocations. This article has used long short-term memory (LSTM) to predict the volume of COVID-19 patients in Pakistan. LSTM is a particular type of recurrent neural network (RNN) used for classification, prediction, and regression tasks. We have trained the RNN model on Covid-19 data (March 2020 to May 2020) of Pakistan and predict the Covid-19 Percentage of Positive Patients for June 2020. Finally, we have calculated the mean absolute percentage error (MAPE) to find the model's prediction effectiveness on different LSTM units, batch size, and epochs. Predicted patients are also compared with a prediction model for the same duration, and results revealed that the predicted patients' count of the proposed model is much closer to the actual patient count.**

***Index Terms*— Covid-19, deep learning, forecasting, long short-term memory (LSTM), pandemics, risk estimation, short term predictio.**

I. Introduction

SINCE December 2019, an outburst of global pandemic Corona Virus disease (COVID-19) [1] started in Wuhan (China) [2] where 2873 deaths were initially reported in China only, and 104 deaths were stated outside the country. It caused the death rate to 3.6% and 1.5%, respectively, till February 2020. Although it originated from China, its quick spread and mortality rate were observed in other areas globally, especially in the US, Italy, U.K., and Spain. By March 1, 2020, around 80 000 patients were reported in China, and 7200 patients were noticed in the rest of the world [3]. In China, the coronavirus's flare-up started, but, by March 15, 2020, the number of infected people surpassed Europe and the USA. The number of mortalities in a few regions presently exceeds the mortality count in China. By the end of April 2020, positive carriers were more than three million worldwide, with more than 210 000 mortalities [4].

COVID-19 is a zoonosis that begins in animals and can be passed to humans following recombination, mutation, and adaption. Although Covid-19 seems similar to influenza virus [5], the transmission speed is a critical contrast between the two infections. Influenza incorporates a shorter middle incubation period and a shorter serial interim than the Covid-19 virus [5]. Here, the incubation period is the time duration between exposure to infection and appearance of symptoms, while serial interim is between progressive cases. The serial interim for COVID-19 is assessed between two to 14 days, whereas, for the influenza virus, it is often around three days. This implies that the dissemination of influenza is shorter than COVID-19. COVID-19 early indications start with any one or combination of more than one symptoms, including temperature, cough, diarrhea, sore throat, high respiratory rate, low oxygen saturation, or shortness of breathing. In later stages, this may lead to acute respiratory distress syndrome. Since December 2019, the virus has been recognized in millions of people around the world and killed thousands of people [6]. According to research and World Health Organizations' findings, COVID-19 overspread overwhelmingly by respiratory beads, but it can also spread from hard surfaces. Both symptomatic and asymptomatic patients can transfer virus at the rate of ranging from 0.5% to 5% [7]. As the infected individuals breathe, hack, or sniffle, they oust small beads of dampness containing the infectious particles. Other individuals in their region may breathe in these particles and get contaminated. The virus-laden beads can too arrive on surfaces that may be touched and get tainted by touching their eyes, nose, or mouth after that. World Health Organization and researchers presently agreed that the virus could stay conveniently on a difficult, nonporous surface, such as steel or plastic, for around three days and unpleasant surface, such as cardboard, for almost a day [8]. The 6-ft-away social distancing rules are implied to keep individuals safe from infected respiratory beads. However, this distance is also not safe enough as sometimes these viral beads can spread out fast [5]. Another reason for the widespread of COVID-19 is that individuals with minor indications and individuals without side effects can spread it around. These asymptomatic cases

Manuscript received July 31, 2020; revised January 20, 2021; accepted January 24, 2021. Date of publication February 19, 2021; date of current version August 2, 2021. *(Corresponding author: Fahad Maqbool.)*

Muhammad Iqbal, Fahad Maqbool, and Saad Razzaq are with the Department of Computer Science and Information Technology, University of Sargodha, Sargodha 40100, Pakistan (e-mail: muhammad.iqbal@uos.edu.pk; fahad.maqbool@uos.edu.pk; saad.razzaq@uos.edu.pk).

Feras Al-Obeidat and Abdallah Tubaishat are with the College of Technological Innovation, Zayed University, Abu Dhabi, United Arab Emirates (e-mail: feras.al-obeidat@zu.ac.ae; abdallah.tubaishat@zu.ac.ae).

Sajid Anwar is with the Center of Excellence in Information Technology, Institute of Management Sciences, Peshawar 25100, Pakistan (e-mail: sajid.anwar@imsciences.edu.pk).

Muhammad Shahrose Khan is with the Department of Computer Science, National University of Computer and Emerging Sciences, Chiniot-Faisalabad Campus, Chiniot 35400, Pakistan (e-mail: shahrose.khan@nu.edu.pk).

Babar Shah is with the College of Information Technology, Zayed University, Abu Dhabi, United Arab Emirates (e-mail: babar.shah@zu.ac.ae).

Digital Object Identifier 10.1109/TCSS.2021.3056769

are the primary cause of their outspread. Corona Virus has spread its roots well deep and could be very dangerous due to its increasing mortality rate. In a few cases, symptom onset might result in death due to intense alveolar damage and respiration failure. The infection can effortlessly spread out at densely populated places like in a stuffed metro car, at a rally, or concert. Social distancing alludes to a measure taken to extend the physical areas between individuals to control the infection's spread by keeping a distance of 6 ft from others when conceivable. Based on the analysis on around 56 000 patients, typical symptoms or indications incorporate 90% fever, 68% dry hack, breath shortness 19%, sickness or spewing 5%, blockage of nasal 5%, and loose bowels 4% [9]. The fever, shortness of breath, and hack are recognized by the Centers for Disease Control (CDC) as common side effects, but others, such as misfortune of scent and taste, have been included in the list, and side effects may take up to 14 days to become evident. According to different reports concluded by health experts, around 81% of affected patients had mild side effects. Of the people infected with considerable side effects, around 14% undergo serious side effects. Critical cases and mortality rate are also related to individuals having an existing medical history, such as a cardiovascular issue, high blood pressure, diabetes, age-related weakness, or low immunity.

Covid-19 outbreak results in increased mortality rate, rise in critical cases, shortage of ventilator, and paramedical services in several areas. World Health Organization has suggested that the lockdown of social, business, and educational activities in most affected areas worldwide counter the situation. Covid-19 has drastic effects on the stock market and economic activities around the globe [10]. The concept of work from home and online education is introduced to overcome the situation. Rapid fall in oil prices and closure of industrial and manufacturing units have started another crisis for underdeveloped and indebted economies [11].

Machine learning and artificial intelligence (AI) algorithms are widely used for forecasting, predictive analysis, and pattern identification [12], [13]. These algorithms can be applied to Covid-19 data to predict and forecast Covid-19 spread out, medical equipment demand, ventilators shortage, or life-saving drug requirement. Covid-19 cases prediction can help out health departments and hospitals to plan for emergencies. In this article, a robust forecasting model has been used to predict COVID-19 cases using long short-term memory (LSTM), a type of RNN. An LSTM model is trained on real available data collected from the National Covid-19 Information Portal [14] of Pakistan to forecast future cases. This article can help to plan better for the upcoming scenario in an epidemic. Hospitals and associated industries can oversee their space, staff, and supplies to provide appropriate and essential care to patients. Areawise prediction can also be exploited to better understand and estimate the geographical specific infection spread patterns. This could also be effective for smart lockdown planning and implementation. Mortality cases can also be predicted from severe patients.

II. Related Work

The Covid-19 outbreak is one of the most severe issues of this century. It disturbed every area of life in no time and affected economic activities and social norms—Covid-19 results in new challenges for business, industry, manufacturing, and almost every social sector. However, Covid-19 made the job of medical practitioners and supporting staff the toughest one. The contagious nature of this disease made it more dangerous. Covid-19 put the whole world in a lockdown state for a long time. However, the lockdown proved fatal to humanity than Covid-19. The World Health Organization and most affected countries coined the term smart lockdown to aid this situation. Only the areas with drastically increasing Covid-19 cases are sealed in smart lockdown. Identification of patterns for such areas is a challenging task. Machine learning is very good at pattern identification and prediction. Researchers applied different techniques to predict Covid-19 patterns in different demographics. Tiwari *et al.* [15] attempted to predict the Covid-19 spread pattern in India using Covid-19 epidemic patterns in China. AI-based techniques can also help the health care professionals in the early detection of disease by correlating the patient symptoms. The authors proposed a forecasting model for predicting affirmed, recouped, and passing cases based on publicly available data (January 22 to April 3) 2020. The forecasting model utilized the time-series prediction method. The impacts of Covid-19 were predicted for the third and fourth week of April 2020. The forecasting model was developed, using WEKA [15]. Marino *et al.* [16] made a similar attempt to predict the Covid-19 spread pattern in Italy, one of the earliest affected countries. Rahman *et al.* [17] predicted Covid pandemic's pattern in Bangladesh. Densely populated areas made this contagious disease more challenging, such as the countries in South Asia. Sharma *et al.* [18] presented a thorough discussion on the challenges of Covid-19 in South Asia. Apart from predicting spread patterns, machine learning can also help detect affected Covid-19 patients. Vaishya *et al.* [19] discussed the importance of AI-based techniques in detecting, monitoring, and tracing Covid-19 cases. Real-time monitoring can help health departments to take advanced measures to counter the spread. Tariq *et al.* [20] proposed a real-time prediction of spread pattern for Covid-19 in Singapore. Deb and Majumdar [21] proposed a time-series method that analyzed the number of Covid-19 episodes and the frequency pattern. They worked on diverse strategies and performed statistical analyses to analyze the flare-up patterns to highlight the current epidemiological stage of a locale so that the distinctive approaches can be recognized to address the Covid-19 widespread in several nations. As per the current circumstance, it is vital to get the disease cases' early spread patterns to arrange and control the successful safety measures [21]. Kucharski *et al.* [22] built a basic SARS-CoV2 transmission model by utilizing diverse data sets to ponder the Covid-19 situation within Wuhan's and surroundings. This model analyzes the conceivable spread of Covid-19 [22]. Recently, several studies have been conducted using exploratory data analysis (EDA) based on the available

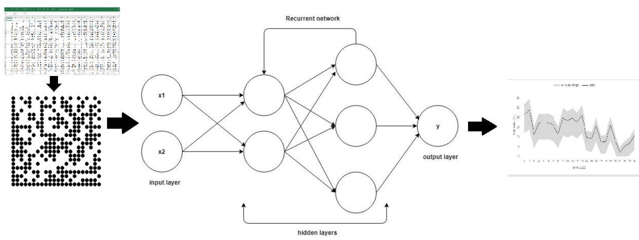

Fig. 1. Architecture of LSTM.

data sets on the epidemiological outbreak of Covid-19, and many researchers are trying to work on the Covid-19 outbreak. The main objectives of these studies focus on the occurrence of confirmed cases, deaths, and recovered cases in Wuhan and throughout the world to understand the incredible threats and upcoming planning for such areas [23]. Tuli *et al.* [24] have used the Robust Weibull model having iterative weighting for fitting generalized Inverse Weibull distribution. They have predicted the growth behavior of Covid-19 and compared the results with the Gaussian model. The above discussions reflect that AI and machine learning-based techniques have been used to identify/detect Covid-19 cases and their associated details. Still, the Covid-19 prediction is in its preliminary phase. With the change in disease symptoms and its associated information, a continuous effort is required to improve machine learning-based forecasting models' performance.

III. Proposed Methodology

A. Data Extraction and Transformation

For predictive analysis, data are extracted from Covid-19 information web portal [14]. The daily updates of new cases, recovered cases, deaths, and critical cases are updated on this web portal and are maintained by the Ministry of Health, Pakistan. This Covid-19 data source [14] is the official resource that contains the provincewise stats of all the above-listed parameters. We have used the data set available till the submission date of this article for model training and testing. A predictive model has been developed using LSTM, as illustrated in Fig. 1, for forecasting cases daily. The LSTM is trained on temporal data that show positive corona cases available from March 11, 2020, to May 31, 2020. For testing the LSTM model, we have used the first 24 days of data of June 2020 that are available until the date before submitting this article. For training, the model utilizes the percentage of daily confirmed cases against a total number of tests performed daily. Once the data are extracted and required preprocessing (on data columns containing dates using date and time functions of Pandas [25]) has been performed, it is divided into two partitions: training data and testing data. It is essential to convert time-series data into a simple structure with input and output components before it fits into the supervised learning model. Here, a time-series generator provided by Keras [26] is used for this conversion. Temporal data batches are created using this time-series generator module.

B. Long Short-Term Memory

LSTM is a special type of artificial neural network that performs well on classification and regression tasks. LSTM is a special type of recurrent neural network (RNN) that uses memory units to serve as short-term memory. LSTM has given outstanding performance on speech recognition [27], handwriting recognition [28], and pattern recognition [29]. It is also a good candidate for pattern recognition. LSTM's prediction performs well on temporal data. An overview of the LSTM model is given in Fig. 1. The LSTM model is defined by passing the hidden nodes and the activation function. Hidden nodes represent the number of neurons used in training the network. The time required to train the model is dependent on the number of nodes. The activation function is used to transform the input into an output. In this article, we have used a rectified linear activation function. Time steps are given as parameters according to the requirements. These are used to define the size of the network's memory. Once the model has been defined, optimizer and loss functions are needed to be initialized. Optimizer calculates the learning rate. We have used the *Adam optimizer* in LSTM model training. The model error is calculated through the loss function, and parameter values are tuned up accordingly. After configuring all parameters, training data are stored in batches and are passed to the compiled model for the training purpose. In this process, the training data are repeated according to a defined number of epochs. The pseudocode for the proposed model is given in Algorithm 1.

Algorithm 1 LSTM Model for COVID-19 Confirm Case Prediction

Input: $COVID-19\ Dataset$
Output: $Graph\ representing\ test\ and\ predicted\ values$
Prerequisites: $Import\ required\ libraries$
Globals: $G : TimeSeriesGenerator()$
$split() : split\ given\ array\ in\ two\ parts$
$N : LSTM\ Network$
$E : Number\ of\ epochs$
$Batch\_size : N$
$D \leftarrow read(data)$
$train_{data},\ test_{data} \leftarrow split(D)$
$K \leftarrow G(train_{data})$
$L \leftarrow G(test_{data})$
Initialize layers,input shape, optimizer and loss function to define N.
for $k = 1\ to\ E$ **do**
  $Train(N, K)$;
end for
$P \leftarrow predictions(N, L)$
$MAPE(test_{data}, P)$
$Graph(test_{data}, P)$

Algorithm 1 demonstrates the flow of execution of the proposed technique. The LSTM network is defined, and the parameters are initialized. The data set is read and stored in variable D. After this, the data are divided into two portions; 80% of the data set is stored in the training set $\text{train}_{\text{data}}$, while the remaining 20% is stored in the testing set $\text{test}_{\text{data}}$. In the next step, training and testing data ($\text{train}_{\text{data}}$ & $\text{test}_{\text{data}}$) are transformed and stored in temporal data batches (K and L),

TABLE I
MAPE VALUES FOR VARYING LSTM NODES, EPOCHS, AND BATCH SIZE. EACH CONFIGURATION WAS EXECUTED FIVE TIMES

LSTM UNITS		10				15				20			
Epochs	BATCH SIZE	Mean±StDev	Median	Max	Min	Mean±StDev	Median	Max	Min	Mean±StDev	Median	Max	Min
10	1	0.046±0.015	0.04	0.07	0.03	0.058±0.013	0.06	0.07	0.04	0.082±0.025	0.085	0.105	0.04
	7	0.766±0.167	0.83	0.93	0.49	0.704±0.105	0.7	0.85	0.57	0.832±0.121	0.9	0.935	0.663
	15	0.918±0.045	0.94	0.96	0.86	0.942±0.044	0.96	0.98	0.87	0.934±0.03	0.918	0.98	0.906
20	1	0.082±0.011	0.08	0.1	0.07	0.088±0.026	0.08	0.13	0.06	0.112±0.023	0.121	0.137	0.081
	7	0.398±0.208	0.33	0.67	0.16	0.176±0.079	0.16	0.31	0.11	0.093±0.046	0.076	0.166	0.056
	15	0.862±0.063	0.88	0.92	0.76	0.64±0.177	0.617	0.844	0.408	0.733±0.197	0.847	0.913	0.482
50	1	0.124±0.048	0.1	0.21	0.1	0.178±0.047	0.2	0.22	0.1	0.202±0.027	0.2	0.23	0.16
	7	0.064±0.009	0.07	0.07	0.05	0.094±0.04	0.1	0.15	0.05	0.077±0.025	0.062	0.112	0.057
	15	0.13±0.089	0.11	0.13	0.28	0.061±0.014	0.062	0.081	0.044	0.079±0.044	0.061	0.137	0.042
100	1	0.196±0.045	0.204	0.242	0.121	0.215±0.006	0.212	0.225	0.21	0.213±0.009	0.214	0.223	0.203
	7	0.262±0.381	0.087	0.943	0.078	0.098±0.037	0.084	0.151	0.067	0.166±0.016	0.171	0.177	0.139
	15	0.116±0.102	0.07	0.27	0.034	0.053±0.017	0.055	0.079	0.035	**0.049±0.015**	**0.047**	**0.072**	**0.034**

respectively, using $TimeseriesGenerator$ function. The proposed technique uses the *Adam optimizer* for the optimization process. The optimization process is used to calculate the model error through a loss function for the optimization of the model. The subgroup size of the training sample is represented by $batch\_size$. It is used to train the model during the learning process. Every batch in the network is trained iteratively while considering the updated weights coming from the previous batches. After training the network, trained models N and L are passed to make predictions that are then stored in P. Once a prediction is made, the testing data and predicted values are passed to the MAPE function to measure the predicted values' error. Results have been elaborated on in Section IV.

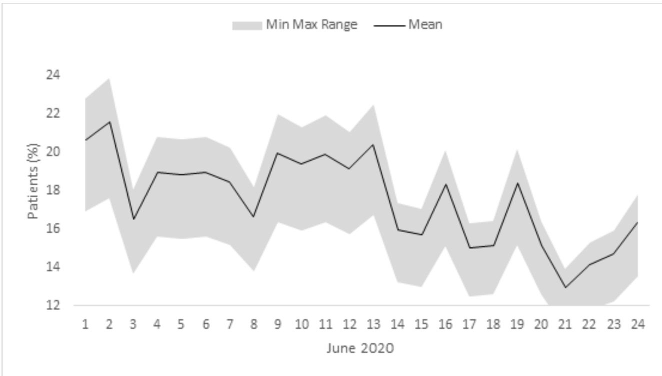

Fig. 2. PPP prediction with ten LSTM units, 100 epochs, and 15 batch size.

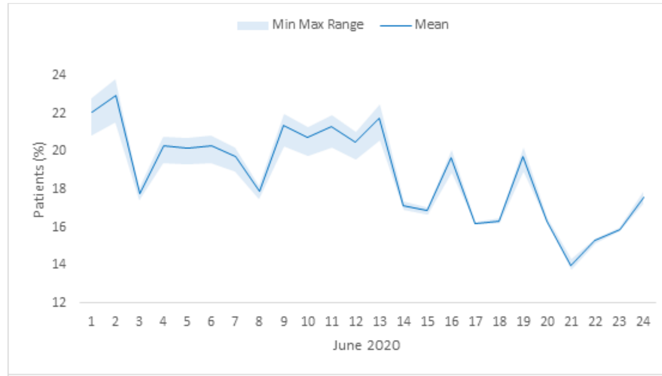

Fig. 3. PPP prediction with 15 LSTM units, 100 epochs, and 15 batch size.

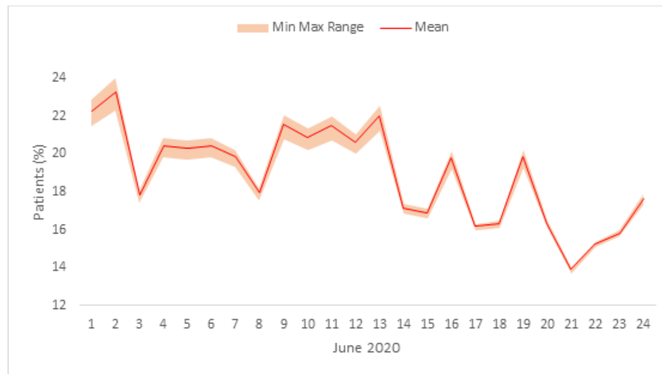

Fig. 4. PPP prediction with 20 LSTM units, 100 epochs, and 15 batch size.

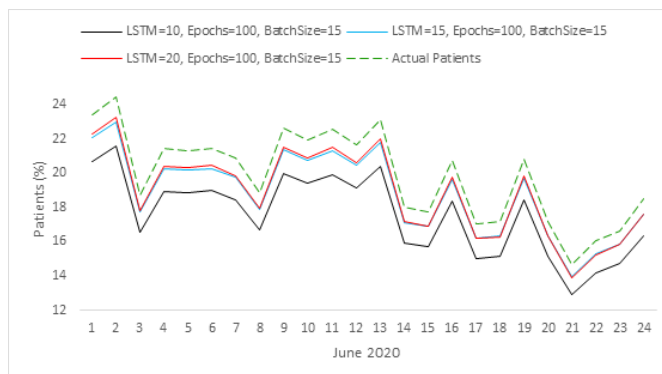

Fig. 5. Comparison of the actual PPP with mean PPP using 100 Epochs, 15 batch size, and varying LSTM units.

IV. Experiments and Results

The LSTM model is trained on the data extracted from Covid-19 data source [14]. Rather than training the LSTM model using daily patient count (PC) and then predicting PC in upcoming days, we have used the Percentage of Positive Patients (PPP) described as

$$\text{PPP} = \frac{\text{No of Daily Positive Tests}}{\text{Total Tests Conducted per Day}}. \tag{1}$$

Using PPP rather than PC is that a limited number of PCR COVID-19 tests are being conducted in Pakistan. Due to the high demand for PCR COVID-19 test kits worldwide, Pakistan can import a limited number of test kits. Hence, direct prediction of PC is not a reasonable consideration. The model was trained with varying LSTM hidden nodes, the number of epochs, and the batch size to find the best performing results. A detailed explanation of these variables has been discussed in the following. Table I represents different experimental values with their corresponding mean absolute percentage error (MAPE). MAPE is used to calculate the prediction accuracy. It indicates how far the model's predictions are off from their corresponding outputs on average. For the evaluation of the forecast performance, the MAPE is a common error indicator benchmark. Based on Lewis' definition [30], if the value of MAPE is greater than 50%, the forecast is considered inaccurate. When its value is less than 50% and greater than 20%, the forecast is reasonable. If the value is less than 20% and more than 10%, it is considered good forecasting. The MAPE value of less than 10% shows outstanding forecasting results. We can see from Table I that the least MAPE value of 0.049 is revealed with LSTM nodes:20, Epochs:100, and Batch Size:15 configurations.

A. LSTM Units Effect on PPP Prediction

Varying LSTM units significantly impact the PPP prediction, but there is no rule for finding optimal units. Fig. 2 shows the PPP prediction across five executions using ten LSTM units. The min–max range is represented through the shaded area representing daily maximum and minimum PPP predicted values across five executions. The solid line represents the average PPP prediction across five executions.

Figs. 2–4 show five executions of each configuration with ten, 15, and 20 LSTM units. It has been shown from the figures that, at 20 LSTM units, the difference between the min–max value decreases, representing that there is not much deviation across multiple executions of the same configuration. It is also clear from Fig. 5 that 20 LSTM units have very close PPP prediction to actual PPP. There is no rule of thumb for finding optimal LSTM units and may vary across multiple domains. In our configurations, 20 LSTM units provide the optimal solution.

B. Batch Size Effect on PPP Prediction

The varying batch size changes min–max difference and standard deviation significantly because a batch size may

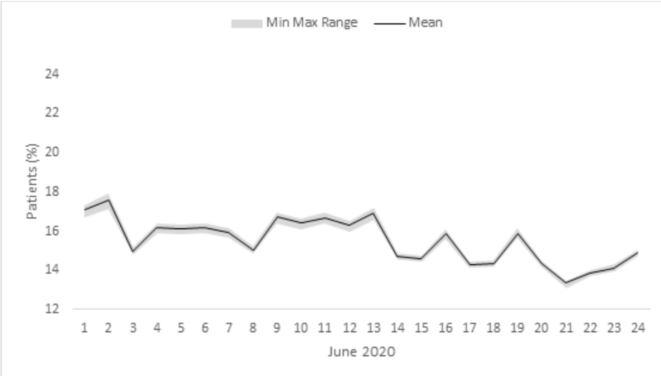

Fig. 6. PPP prediction with 20 LSTM units, 100 epochs, and one batch size.

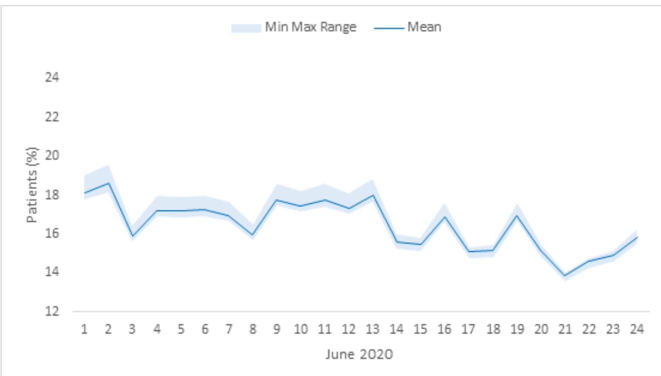

Fig. 7. PPP prediction with 20 LSTM units, 100 epochs, and seven batch size.

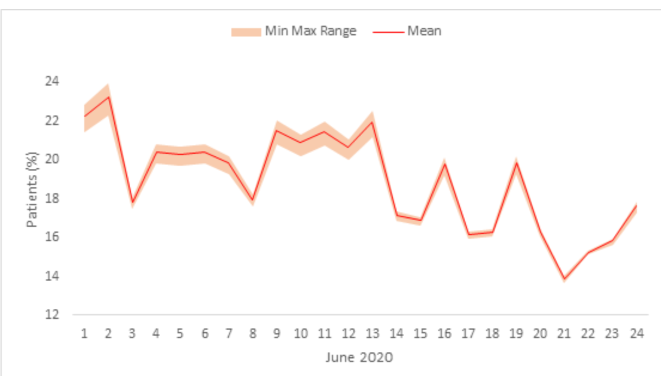

Fig. 8. PPP prediction with 20 LSTM units, 100 epochs, and 15 batch size.

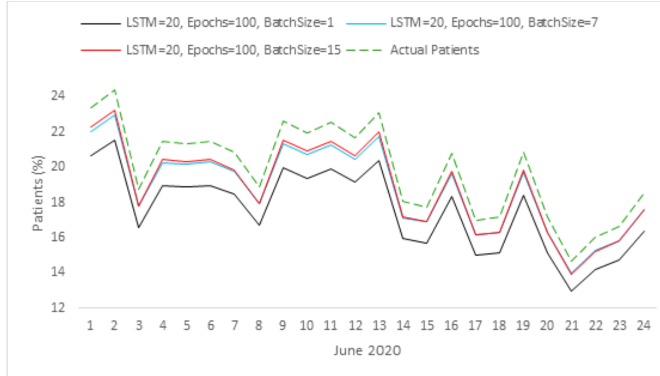

Fig. 9. Comparison of the actual PPP with mean PPP using 20 LSTM units, 100 epochs, and varying batch sizes.

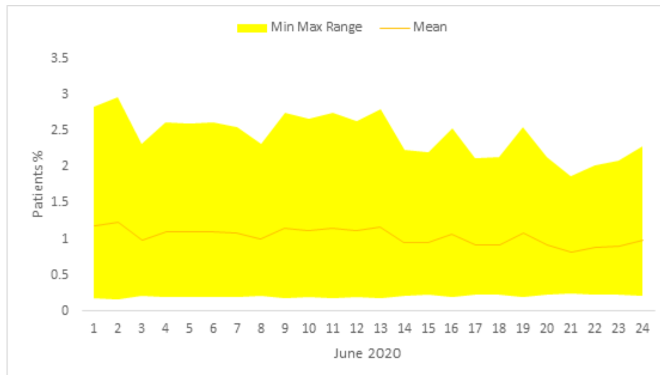

Fig. 10. PPP prediction with 15 LSTM units, 10 epochs, and 15 batch size.

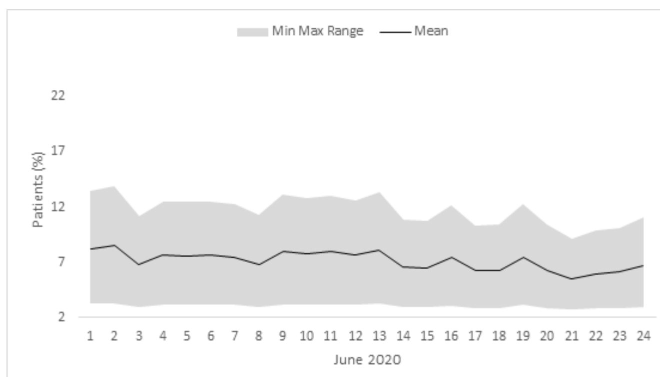

Fig. 11. PPP prediction with 15 LSTM units, 20 epochs, and 15 batch size.

represent the number of records after which a prediction is made. Fig. 6 shows the PPP prediction across five executions using a batch size of 1. The min–max range is represented through the shaded area representing daily maximum and minimum PPP predicted values. The solid line represents the average PPP prediction.

Figs. 7 and 8 show five executions with a batch size of 7 and 15, respectively. It is evident from the figures that, as the batch size varies, the difference between min–max values changes, representing a deviation across multiple executions of the same configuration. It is also clear from Fig. 9 that the batch size of 15 has a very close PPP prediction to actual PPP. This shows that predicting PPP after 15 days of data is a more suitable choice.

C. Epochs Effect on PPP Prediction

Varying epochs also affect the PPP prediction significantly. Fig. 10 shows the PPP prediction across five executions using 10 epochs, 15 LSTM units, and 15 batch size. The min–max range is represented through the shaded area representing daily maximum and minimum PPP predicted values. The solid line represents the average PPP prediction.

Figs. 10–13 show the executions with 15 LSTM units, 15 batch size, and with 10, 20, 50, and 100 epochs, respectively. It is evident from the figures that, at 100 epochs, the difference between min–max values decreases, representing that there is not much deviation across multiple executions of the same configuration. It is also clear from Fig. 14 that 100 epochs

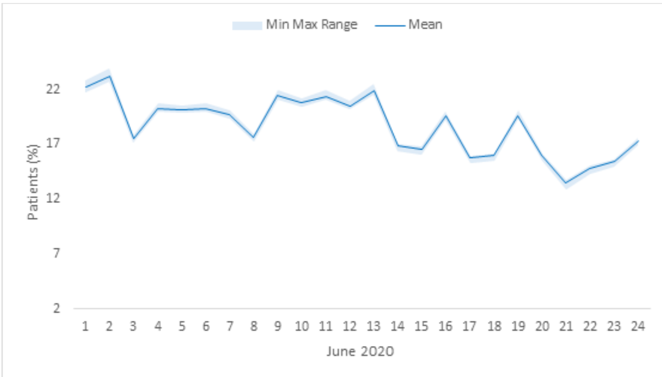

Fig. 12. PPP prediction with 15 LSTM units, 50 epochs, and 15 batch size.

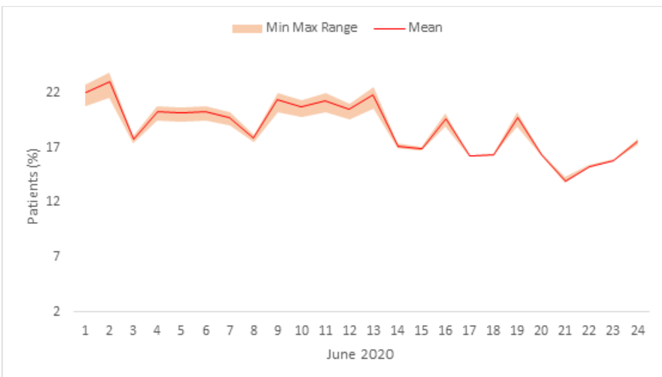

Fig. 13. PPP prediction with 15 LSTM units, 100 epochs, and 15 batch size.

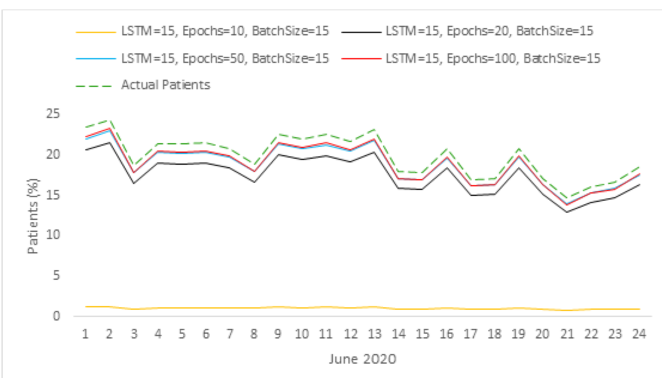

Fig. 14. Comparison of the actual PPP with mean PPP using 15 LSTM units, 15 batch size, and varying epochs.

have very close PPP prediction to actual PPP due to significant training iterations.

D. PC Prediction

In this section, we finally conclude the total number of predicted positive COVID-19 cases between June 1, 2020, and June 24, 2020, using the following equation:

$$\text{Predicted PC}_{\text{StartDate}}^{\text{EndDate}} = \sum_{i=\text{StartDate}}^{\text{EndDate}} (\text{Predicted PPP}_i * \text{Predicted Tests}_i) \quad (2)$$

Predicted PPP has been discussed in detail in Sections IV-A–IV-C. To predict PCR COVID-19 tests that would be performed daily, we have again used LSTM, and results have been shown in Table II. We can see from COVID-19 data set that the actual PC between the said dates is **120 510**, and the predicted PC using (2) is **111 860**. This

TABLE II

TOTAL TESTS PER DAY PREDICTION USING 20 LSTM UNITS, 15 BATCH SIZE, AND 100 EPOCHS. EXPERIMENT WAS EXECUTED FIVE TIMES

Date	Actual	Predicted Tests			
Year (2020)	Tests	Avg ± StDev	Median	Max	Min
June 1	16146	16146 ± 510	16140	16926	15359
June 2	16948	16948 ± 536	16942	17767	16121
June 3	19676	19676 ± 622	19670	20628	18717
June 4	22257	22257 ± 704	22249	23333	21172
June 5	21645	21645 ± 684	21638	22692	20590
June 6	22538	22538 ± 712	22530	23627	21439
June 7	22099	22099 ± 699	22091	23167	21022
June 8	24021	24021 ± 759	24013	25182	22850
June 9	23220	23220 ± 734	23212	24342	22088
June 10	25926	25926 ± 820	25917	27180	24662
June 11	27654	27654 ± 874	27644	28991	26306
June 12	29123	29123 ± 921	29113	30531	27704
June 13	28827	28827 ± 911	28817	30220	27421
June 14	28377	28377 ± 897	28367	29749	26994
June 15	24406	24406 ± 771	24398	25586	23216
June 16	27432	27432 ± 867	27423	28759	26095
June 17	30733	30733 ± 971	30722	32219	29235
June 18	28122	28122 ± 889	28113	29482	26751
June 19	30909	30909 ± 977	30899	32404	29403
June 20	28152	28152 ± 890	28143	29514	26780
June 21	29777	29777 ± 941	29767	31217	28325
June 22	24000	24000 ± 759	23992	25161	22830
June 23	22811	22811 ± 721	22803	23914	21699
June 24	21304	21304 ± 673	21296	22334	20265

aggregation predicted PPP and predicted tests were calculated using 20 LSTM units, 15 Batch Size, and 100 epochs, as shown in Fig. 8, and Table II, respectively. To find the effectiveness of our proposed work, we compared our predicted PC with the predicted PC of Covid-19 Scenario Analysis tool developed by Imperial College London [31]. The comparison revealed that the LSTM model provides a closet percentage of PC compared to actual PC. Imperial College London model's estimated PC for the abovementioned time duration is approximately **14 million**.

V. Conclusion

In this article, we have used the LSTM model to predict the COVID-19 PC in Pakistan. Due to the limited availability of PCR COVID-19 test kits, we cannot directly predict PC. Instead, we have calculated PC by predicting both PPP and total tests per day separately. Results revealed that our predicted PC is much closer to the actual PC between June 1, 2020, and June 24, 2020. The proposed model can also predict Covid-19 cases, areawise, that help smart lockdown decisions and areawise sampling. In the future, we have plans to train the model on death cases and will try to forecast the mortality ratio and its correlations with the critical cases.

References

[1] F. Wu *et al.*, "A new coronavirus associated with human respiratory disease in China," *Nature*, vol. 579, no. 7798, pp. 265–269, 2020.

[2] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, pp. 497–506, May 2020.

[3] *Coronavirus Disease 2019 (COVID-19) Situation Report*, World Health Organization, Geneva, Switzerland, 2020.

[4] E. Dong, H. Du, and L. Gardner, "An interactive Web-based dashboard to track COVID-19 in real time," *Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, May 2020.

[5] *Coronavirus Disease 2019 (COVID-19) Situation Report*, World Health Organization, Geneva, Switzerland, 2020.

[6] *World Health Organization*. Accessed: May 31, 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[7] K. J. Clerkin *et al.*, "COVID-19 and cardiovascular disease," *Circulation*, vol. 141, no. 20, pp. 1648–1655, 2020.

[8] B. Resnik. *World Health Organization*. Accessed: May 31, 2020. [Online]. Available: www.vox.com/science-and-health/2020/4/2/21197617/coronavirus-pandemic-covid-19-death-rate-transmission-risk-factors-lockdowns-social-distancing

[9] *Report of the Who-China Joint Mission on Coronavirus Disease 2019 (COVID-19)*, World Health Organization, Geneva, Switzerland, 2020.

[10] P. K. Ozili and T. Arun, "Spillover of COVID-19: Impact on the global economy," Tech. Rep., Mar. 2020. [Online]. Available: https://ssrn.com/abstract=3562570

[11] M. Gupta *et al.*, "COVID-19 and economy," *Dermatol. Therapy* vol. 33, no. 4, 2020, Art. no. e13329, doi: 10.1111/dth.13329.

[12] F. Al-Obeidat, Á. Rocha, M. S. Khan, F. Maqbool, and S. Razzaq, "Parallel tensor factorization for relational learning," *Neural Comput. Appl.*, vol. 4, pp. 1–10, Jan. 2021.

[13] S. Razzaq, F. Maqbool, and A. Hussain, "Modified cat swarm optimization for clustering," in *Proc. Int. Conf. Brain Inspired Cognit. Syst.* Cham, Switzerland: Springer, 2016, pp. 161–170.

[14] Health Advisory Platform. *COVID-19 Pakistan*. Accessed: Jun. 24, 2020. [Online]. Available: https://covid.gov.pk/

[15] S. Tiwari, S. Kumar, and K. Guleria, "Outbreak trends of coronavirus disease–2019 in India: A prediction," *Disaster Med. Public Health Preparedness*, vol. 14, pp. 1–6, Oct. 2020.

[16] M. Gatto *et al.*, "Spread and dynamics of the COVID-19 epidemic in italy: Effects of emergency containment measures," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 19, pp. 10484–10491, May 2020.

[17] M. R. Rahman, A. H. M. H. Islam, and M. N. Islam, "Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID-19) pandemic in bangladesh towards vulnerability zoning and management approaches," *Model. Earth Syst. Environ.*, vol. 7, pp. 1–29, Sep. 2020.

[18] G. D. Sharma, G. Talan, M. Srivastava, A. Yadav, and R. Chopra, "A qualitative enquiry into strategic and operational responses to COVID-19 challenges in South Asia," *J. Public Affairs*, Jun. 2020.

[19] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial intelligence (AI) applications for COVID-19 pandemic," *Diabetes Metabolic Syndrome, Clin. Res. Rev.*, vol. 14, no. 4, pp. 337–339, Jul. 2020.

[20] A. Tariq *et al.*, "Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020," *BMC Med.*, vol. 18, no. 1, pp. 1–14, Dec. 2020.

[21] S. Deb and M. Majumdar, "A time series method to analyze incidence pattern and estimate reproduction number of COVID-19," 2020, *arXiv:2003.10655*. [Online]. Available: http://arxiv.org/abs/2003.10655

[22] A. J. Kucharski *et al.*, "Early dynamics of transmission and control of COVID-19: A mathematical modelling study," *Lancet Infectious Diseases*, vol. 20, no. 5, pp. 553–558, 2020.

[23] S. K. Dey, M. M. Rahman, U. R. Siddiqi, and A. Howlader, "Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach," *J. Med. Virology*, vol. 92, no. 6, pp. 632–638, Jun. 2020.

[24] S. Tuli, S. Tuli, R. Tuli, and S. S. Gill, "Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing," *Internet Things*, vol. 11, Sep. 2020, Art. no. 100222.

[25] Pandas team. *Pandas*. Accessed: Jun. 24, 2020. [Online]. Available: https://pandas.pydata.org/

[26] Keras development team. *Keras*. Accessed: Jun. 24, 2020. [Online]. Available: https://keras.io/

[27] A. Graves, N. Jaitly, and A.-R. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*, Dec. 2013, pp. 273–278.

[28] T. M. Breuel, A. Ul-Hasan, M. A. Al-Azawi, and F. Shafait, "High-performance OCR for printed english and fraktur using LSTM networks," in *Proc. 12th Int. Conf. Document Anal. Recognit.*, Aug. 2013, pp. 683–687.

[29] F. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016.

[30] C. D. Lewis, *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. Oxford, U.K.: Butterworth-Heinemann, 1982.

[31] *COVID-19 Scenario Analysis Tool*. Accessed: Jun. 24, 2020. [Online]. Available: https://covidsim.org/v2.20200706/?place=Pakistan