Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

Infectious Disease Patient Count Prediction Using Machine Learning Algorithm

Mrs. K.Karpagavalli Assistant Professor in Department of CSE, Annamacharya Institute Of Technology And Sciences(Autonomous), Rajampet, Andhra Pradesh, India-516126

T.Lakshmiprasanna[2], E.Sai Vandana[3], N.Manohar Reddy[4], K.Renuka[5] , P.Naveen Kumar[6]
Department of Computer Science And Engineering, Annamacharya Institute Of Technology And Sciences(Autonomous), Rajampet, Andhra Pradesh, India-516126

ABSTRACT: A pandemic COVID-19, which started in Wuhan, China, in December 2019, and within Several weeks it had expanded to around 200 nations. Each nation that had the sickness began to take its toll the required steps to preventing the spread, give sick individuals the greatest medical treatment possible, and implement safeguards to restrict the growth. It became necessary to simulate the progress of infection in order to predict the number of patients, using calculation because the infection spread exponentially. The essential ingredient is steps perhaps regional governments might implement to stop the span, manage hospital burden, and allocate resources is their evaluation of these people. In this study, long short-term memory was used to anticipate the number of COVID-19 patients in Pakistan. For classification, prediction, and regression, recurrent neural networks of a specific type known as LSTM are used. We trained the RNN model with Pakistani Covid-19 data from March 2020 to May 2020 in order to anticipate the Covid-19 Percentage of Positive Patients for June 2020. Finally, we computed the mean absolute percentage error to measure the model's prediction performance on various LSTM units, batch sizes, and epochs. (MAPE). When anticipated patients are compared to a prediction model for the same time, the results reveal that the proposed model's predicted patient count is substantially closer to the actual patient count.
Keywords— COVID-19, robust learning, prognosticate, lstm, infection, risk assessment, prognosis.

1.INTRODUCTION

Since December 2019 in Wuhan, China, the global pandemic Corona infection illness (COVID-19) [1] began. Initially, only 2873 deaths in China were reported, while 104 deaths were reported outside of China. Up until February 2020, it increased mortality rate by 3.6% and 1.5%. Although it started in China, it quickly spread over the world and had a high fatality rate, particularly in the US, Italy, the UK, and Spain. Despite it started in China, it quickly lay off over the world and had a high fatality rate, particularly in the US, Italy, the UK, and Spain. Around 80 000 patients were recorded in China as of March 1, 2020, while 7 200 people were spotted elsewhere in the world [3]. The coronavirus outbreak first appeared in China, but by March 15, 2020, the region's population had overtaken that of Europe and the USA. There are currently more deaths in a few areas than there are overall China. Positive indicators numbered more than 3,000,000 in the world by the end of April 2020, having over 210 000 fatalities [4].

A zoonosis called COVID-19 that starts in creatures and can spread to people through reconnecting, variation, and adaptation. Despite the fact that Covid-19 and the flu appear to be similar [5], there are important differences between the two viruses in the pace of transmission. Compared to the Covid- 19 virus, influenza has shorter serial incubation time and midway time interval [5]. In this example, During the gestation stage, the amount between being exposed to the disease and the

onset of indicators, whereas the a series interim is the interval among patients that progress. While the influenza virus typically has a serial interim of three days, the COVID-19 serial interim is estimated to range between two and fourteen days. This suggests that the influenza virus spreads more slowly than COVID-19. The virus-filled beads may also land on surfaces that people may touch and become contaminated if they then touch their eyes, snout, or maw. According to the WHO and analysers, the virus may readily survive for up to three days on a hard, surface that is impermeable like plastic and for almost a day on an unpleasant surface like cardboard [8]. It is believed that the 6-foot social distance restrictions protect people from infectious respiratory particles. These viral beads can occasionally spread quickly, therefore this distance is also insufficiently safe [5]. The fact that people with modest symptoms and those with no side effects might disseminate COVID-19 is another factor in its extensive distribution. The main factor causing their prevalence is these asymptomatic instances. Due to its rising fatality rate, the corona virus has established deep roots and may be highly deadly. A few times, the onset of symptoms could lead to death from severe alveolar injury and respiratory failure. A packed metro vehicle, a rally, or a concert are just a few examples of heavily populated areas where the infection can spread quickly. To be able to stop an infection from spreading,, social distancing refers to the practice of keeping a 6 foot space from other people whenever it is practical. According to an examination of 56000 individuals, Signs include less breath, 5% vomiting or diarrhoea, 5% nasal blockage, and 4% loose stools [9]. Central Agency for Disease Control recognizes the temperature, less breath, and hack as typical side effects, but others have been added to the list, such as loss of aroma and taste, these side effects could manifest for up to 14 days. Roughly 81% of the afflicted victims experienced some problems, according to several reports put together by health professionals. Around 14% of those who experience significant side effects also experience major side effects. Individuals prior medical conditions, such as weakened immune system, or age-related frailty, are also linked to critical cases and fatality rates. The Covid-19 outbreak causes a surge in critical patients, a decrease in ventilator availability, and paramedical services shortages in a number of locations. The WHO has recommended that communal and commercial activities be suspended in the majority of the affected regions of the world to combat the crisis. The stock market and global economic activity are significantly impacted by COVID-19 [10]. To remedy the situation, the ideas of virtual working and virtual learning environment are initiated. For underdeveloped and debt-ridden economies, another catastrophe has been sparked by drop in prices of oil and the closure of industrial and manufacturing prerequisites [11].

Algorithms for Artificial Intelligence and Machine Learning are frequently employed for pattern recognition, predictive modelling, and forecasting [12, 13]. These formulas are applicable with COVID-19 data to foresee and predict COVID-19 ample, demand for medical equipment, a shortage of ventilators, and the need for life-saving medications. Hospitals and health departments can better prepare for emergencies by using the COVID-19 cases prediction. LSTM, a form of recurrent neural networks, has been employed in this paper to anticipate COVID-19 cases using a reliable forecasting model. To predict upcoming instances, A trained LSTM model is used for actual data that was gathered from Pakistan's Information Portal for the National Covid-19 [14]. This article can aid in making better preparations for the eventuality of an epidemic. In order to give patients the essential and appropriate treatment they need, hospitals and related industries can manage their facilities, personnel, and supplies. To better comprehend and estimate the geographically specific

patterns of illness propagation, area wise prediction can be used. This might be useful for implementing and planning smart lockdowns. Severe patients can also predict mortality cases.

2. RELATED WORK

One of the worst problems of this century is the Covid-19 pandemic. It immediately caused havoc in all areas of life and had an impact on economic activity as well as social norms; as a result, COVID-19 presents new difficulties for virtually every sector of society, including commerce, manufacturing, and industry. However, Covid-19 makes it particularly difficult for medical professionals and support employees to do their jobs. This sickness was more harmful because it was contagious.Covid-19 put the entire world on a lockdown considerable amount of time. Although Covid-19 survived the lockdown, humanity did not. The term "smart lockdown" was created to help in this case by Organization for World Health and the majority of the afflicted nations. Only those regions are sealed in smart lockdown that are experiencing a sharp increase in Covid-19 instances. Finding patterns in these locations is a difficult endeavour. Pattern recognition and prediction are two areas where machine learning excels. Researchers used a variety of strategies to forecast the patterns in Covid-19 various demographic groups. Tiwari and others[15] .'s attempt used Covid-19 pandemic tendencies in China to forecast A Covid-19 distribution in India is typical. By comparing the patient's symptoms, AI-based methods can aid medical personnel in the early diagnosis of disease. in light of openly accessible(January 22 through April 3) figures show that authors suggested a forecasting methodology for estimating affirmed, recouped, and passing cases in 2020. The time-series forecasting technique was used by the forecasting model. Covid-19 was expected to have effects throughout the third and fourth weeks dated April 2020. WEKA was utilised to create the forecasting model [15]. Similar efforts were done by Marino et al. [16] to forecast A Covid-19 dissemination Italy's pattern is one of the first nations hit. The pattern of the Covid pandemic in Bangladesh was predicted by Rahman et al. [17]. This infectious disease was more difficult to control in densely populated locations, such the nations of South Asia. The difficulties using Covid-19 in South Asia were thoroughly discussed by Sharma et al. in their paper from [18]. Machine learning can aid in the finding of impacted Covid-19 sufferers in addition to forecasting spread patterns. The significance based on AI strategies in identifying, following, likewise tracking COVID-19 instances was explored by Vaishya et al. [19]. Health departments can take proactive steps to stop the spread with the use of real-time surveillance. A forecast made right away of the dissemination Singapore's Covid-19 pattern was put forth by Tariq et al. [20]. Time-series approach that examined the quantity with Covid-19 events and their frequency distribution was put forth by Deb and Majumdar [21]. They worked on several ways and carried out statistical studies to examine the trends in flare-ups to be able to highlight the epidemiological stage at which a location and identify the many strategies to deal with the Covid-19 that is prevalent in many countries. In light of the current situation, it is essential to plan and manage successful safety measures by understanding the early transmission patterns of illness cases [21]. By combining several data sets, Kucharski and others [22] developed a simple Model for SARS-CoV2 transmission to analyse the Covid-19 status in Wuhan and its environs. The potential distribution of Covid-19 is examined by this model [22]. Recent investigations regarding

the Covid-19 epidemic outbreak have been undertaken utilizing exploratory data analysis (EDA), and other scientists are attempting working to combat the Covid-19 outbreak.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

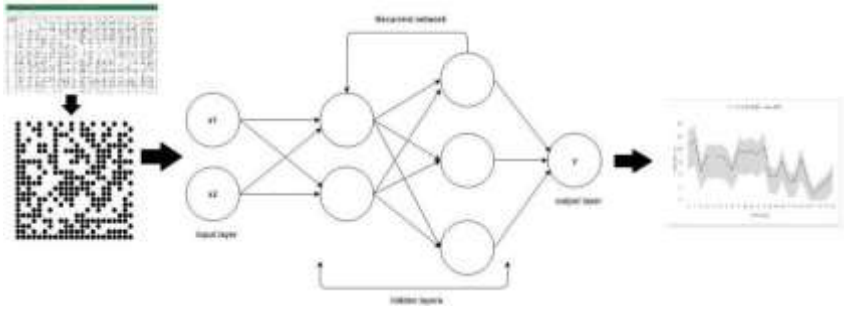

Fig. 1. Structure of the LSTM.

The major goals of these research are to comprehend the great hazards and future planning for such locations by focusing information on confirmed cases, fatalities, and recovered cases in Wuhan around the globe [23]. The Robust Weibull model with iterative weighting was used by Tuli et al. [24] to fit the generalised Inverse Weibull distribution. They made predictions about Covid-19's growth patterns and compared the outcomes to those anticipated by the Gaussian model. The aforementioned discussions demonstrate how Covid-19 cases and their accompanying information have been identified/detected using AI and machine learning-based techniques. Covid-19 remains in the forecast the initial stages. The efficacy of machine learning-based forecasting models needs to be continuously improved due to the changing nature of disease symptoms and the information they are connected with.

3. SUGGESTIVE METHODOLOGY

A. Data Translation and Extraction

Data is taken from the Information about Covid-19 online platform for predictive analysis [14]. Pakistan's Ministry of Health, updates this website daily with information on new cases, cases that have been recovered, deaths, and critical cases. The official source for the province-level statistics for all the previously mentioned factors is this Covid-19 data source [14]. We trained and tested our model using the set of data that was accessible as of the piece's submission date. For forecasting cases daily, There is a predictive model created LSTM is used, as shown as Evidence. On the basis of temporal data that covers the period from March 11, 2020, until May 31, 2020, the LSTM is trained. During the initial 24 days of June 2020 data which were accessible up before the deadline for delivering this work in order to evaluate the LSTM model. The daily verified caseload percentage for the total count of daily analysis is used by the model as input for training. Data are divided up into two partitions: grounding data and evaluation data after the necessary preparation has been done on date-containing data columns using Pandas' time of day functions [25]. Before pattern-series data can be incorporated into a supervised learning model, it must be transformed toward straightforward both input and output structures elements. For this conversion, a generator of time series offered from Keras [26] is employed. With the help of this time-series generator module, temporal data batches can be produced.

B. Long Short-Term Memory

The artificial neural network variant known as LSTM excels at classification and regression tasks. Short-term memory is used by LSTM, a particular kind using recurrent neural

networks (RNN), which employs memory units. for this purpose. In the areas of Speech [27], handwriting [28], and pattern recognition [29] recognition, LSTM has demonstrated exceptional performance. As a contender for pattern recognition, it is also strong. On temporal data, LSTM prediction performs well. Fig. 1 presents introduction to the LSTM model. The activation function and the hidden nodes define the LSTM model. The amount of neurons utilised to train the system is represented by hidden nodes. The number of nodes affects how long it takes to train the model. The input is converted into an output using the activation function. That is essay, There are employed a linear activation function corrected. In accordance with the criteria, Steps in time are provided as guidelines. These are used to specify the network's memory size. The Loss and optimizer operations must to be initiated once the model has been defined. The learning rate is calculated by an optimizer.

4. EXPERIMENTS AND RESULTS

Data taken from the Source of Covid-19 data is used to train the LSTM model [14]. There are utilized the % of Patients with Positive Results (PPP)provided as an alternative to using the daily patient count (PC) to train the LSTM model and predicting PC in the following period.

$$PPP = \frac{\text{Daily Positive Tests Percentage}}{\text{Total Number of Tests Per Day}}$$

A small In Pakistan, several PCR COVID-19 tests are being performed., hence PPP is being used rather than PC. Pakistan can only import a certain quantity of COVID-19 test kits for PCR because to the great demand for them across the globe. Direct PC prediction is therefore not a valid consideration. To achieve the greatest performance, The simulation was trained using various LSTM concealed nodes, the quantity as well as the batch size. The following section discusses these variables in great detail. Different experimental values are shown in Table I together with their associated mean absolute percentage inaccuracy (MAPE).The prediction accuracy is computed using MAPE. It shows the average difference between the model's predictions and its related outputs. The MAPE is a typical error indicator standard for assessing forecast performance. According to Lewis' definition [30], the forecast is deemed erroneous if the MAPE score is higher than 50%. The forecast is reasonable when its value is between 50% and 20%. It is regarded as accurate predicting if the value is greater than 10% and less than 20%. A MAPE value of less than 10% demonstrates excellent forecasting outcomes.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

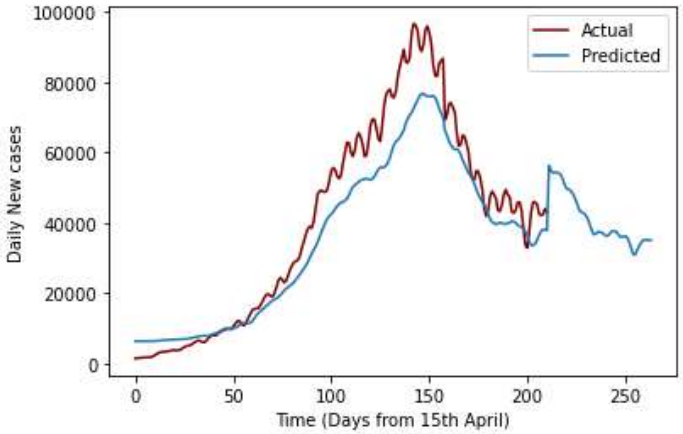

Fig. 2. Univariate analysis

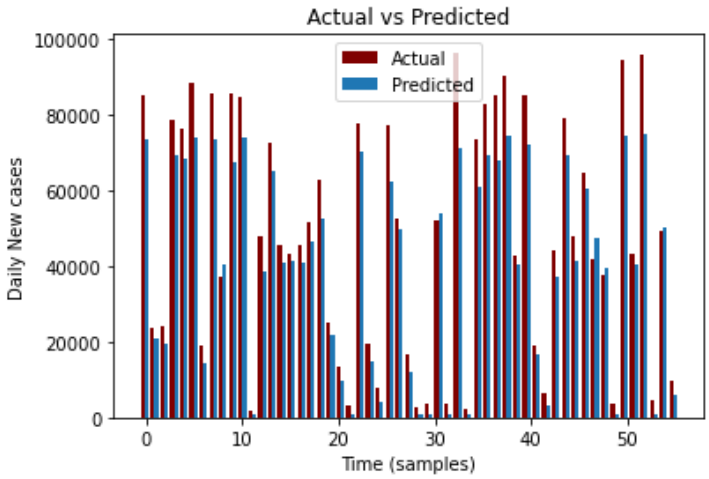

Fig. 3. Multivariate analysis

5. CONCLUSION

In this article, the COVID-19 was predicted using the LSTM model. We are unable to forecast PC directly since PCR COVID-19 test kits are not widely available. Instead, we estimated PC by forecasting PPP and the total number of tests per day independently. Results showed that between June 1, 2020, and June 24, 2020, our anticipated patient count is substantially same as real patient count. The suggested method also forecast COVID-19 cases locally, which aids in informed lockdown choices and local sampling. We intend using death cases to train the model in the future and attempt to predict transience rates and their relationships with severe instances.

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-2019K.
[7] 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 opti- mization 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 intelli- gence (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 net- works," 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://covidsi m.org/v2.20200706/?place=Pakistan.