Third International Conference on "Materials, Computing and Communication Technologies"
in association with International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Covid-19 Patient Count Prediction Using Machine Learning Algorithms

Dr. D. Ganesh[1], Mr. K. Arun Kumar[2], Ms. K. Manvitha[3], Ms. K. Tejaswi[3], Ms. K. Sony Priya[3], Mr. Pawan Kalyan.L[3]

[1]Associate Professor, Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India

[2]Assistant Professor Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India

[3]UG Scholar, Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India

ABSTRACT

More than 200 locations have been affected by the pandemic known as COVID-19 since it first emerged in Wuhan, China in 2009. Infected countries began to take the required actions to stem the spread of disease, such as providing the finest medical care to those infected and implementing preventative measures to curb the spread. Because the virus extend exponentially, infection spread patterns had to be simulated in order to calculate the patient volume. Estimating the number of people who will be affected is essential if local governments are to take the necessary steps to stop the spread of the disease, minimise the number of patients who will be admitted to the hospital, and allocate resources. The quantity of COVID-19 patients in India was predicted using long short-term memory (LSTM). The LSTM is a classifier, predictor, and regression recurrent neural network (RNN). Using India's COVID-19 data, we developed an RNN model to predict the percentage of COVID-19-positive patients in the U.S. in June 2020. We must also calculate the model's accurateness of unit, batch size, and epoch prediction accuracy by using mean absolute error (MAPE). For same time period, the patients are also matched to a prediction model, and it is shown that the findings are similar to the actual data collected. The goal of this work is to forecast Indian statistics. For model training and testing, we used machine learning algorithms and the data set that was available at the time this article was submitted. A prediction model was built using LSTM and RNN.

Keywords: Covid-19,LSTM,RNN,MAPE

I. INTRODUCTION

It started in Wuhan (China) in December 2019, with 2873 deaths first reported in China exclusively, with 104 deaths outside the country being reported. The death rate rose by 2.7 per cent and percentage points, respectively, until January 2021 as a result. Despite its origins in China, the virus quickly spread over the world, especially to the U.s, Italy, the U.k., and Spain[1]. A total of 80,000 patients in China and 7,200 patients around the world has been documented as of March 1, 2020. As of February 15, 2020, the incidence of corona virus infections had surpassed those in Europe and the Usa[2]. There has been a significant increase in the number of deaths in a few places.

Mortality, serious cases, and the availability of ventilators and paramedics have all increased because of the COVID-19 outbreak. The World Health Organization (WHO) has proposed that most affected areas around the world restrict social, commercial, and educational activities to combat the condition[3]. COVID-19 has a major impact on stock prices and global economic activities. Work-from-home opportunities and distance learning are being promoted as solutions to the problem. Due to the sharp decline in oil prices and resulting closures of manufacturing and industrial facilities, developing and heavily indebted nations are now facing a new crisis.

II. LITERATURE SURVEY

2.1. 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.[4]

An interactive online dashboard has been built to track and monitor cases reported of corona virus infection 2019 (COVID-19) in live time in response to the current public health emergency (COVID-19). Online portal DXY aggregates local press and government sources in China and other countries to offer COVID-19 case totals at the provincial and country levels in near real time. A tractor trailer living streaming data technique was employed instead of relying on people to update the data, because doing so would be unsustainable.

LIMITATIONS:

Despite the use of a semi-automated living data stream technique, some data is still manually submitted.

2.2. F. Al-Obeidat, Á. Rocha, M. S. Khan, F. Maqbool, and S. Razzaq

in recommendation engines, intelligent agents, and knowledge bases, prediction is a converter problem[5].

Several problem-solving techniques have already been developed. Tensor factorization is one approach to solving the link prediction problem. RESCAL is a well-known tensor factorization method for solving large-scale problems in a short amount of time and space.

LIMITATIONS:

The primary disadvantages of RESCAL are that it is inefficient to train due to its quadratic run-time and poor scaling to large knowledge graphs.

2.3. S. Razzaq, F. Maqbool, and A. Hussain

Clustering is regarded as a serious issue in terms of efficiency. They developed a modified cat swarm optimization method[6]. a CSO model that replicates cat behavior to solve collection development problems Optimization of heuristic modified cat swarms MCSO is used instead of the user-specified k to obtain initial collections.

LIMITATIONS:

The results of developing a modified cat breed are less accurate than the results of other methods.

2.4. S. Deb and M. Majumdar

Time-series forecasting is the practise of producing factual predictions based on previously moment data. Based on previous data, it creates models that forecast future observation and strategic decisions based on those models' predictions. They propose a time series approach to evaluating the trend pattern of COVID-19 outbreak incidence[7].Using this simple to develop model, we demonstrated that a time-dependent quadratic trend accurately captures the disease's incidence pattern using acceptable diagnostic data.

LIMITATIONS:

A perfect prediction does not exist.

It may take a long time and a significant amount of money to complete.

2.5. S. Tuli, S. Tuli, R. Tuli, and S. S. Gill

It is possible to keep track of the disease, predict its progress, and devise plans and laws to prevent spreading using ML and Cloud Computing (CC).[8]

An improved machine learning-based approach was used to forecast the potential threat of COVID-19 in countries around the world.

A data-driven method with increased precision, such as this one, can be extremely beneficial in allowing the government and the public to respond more quickly.

LIMITATIONS:

Big data sets which are both comprehensive and unbiased are necessary for machine learning to train on. When fresh data is generated, they will have to wait.

III. METHODOLOGY

3.1 DATA EXTRACTION:

Until the prediction is revised, data is gathered from the COVID-19 online platform. In addition to new and reclaimed cases, the data provides information on dates, crucial cases, and more. For train and test reasons, this data was acquired. For daily prediction of positive cases using LSTM, Bidirectional LSTM, and Encrypted LSTM, a predictive model employs LSTM. The model's training data spans 250 days. The entire number of positive results has been calculated by dividing the number of tests completed, resulting in a percentage.

3.2 LSTM:

Collection prediction problems can be handled using LSTM neural networks, which are recurrent neural systems capable of handling order dependence[9].Device translating, voice recognition, and the like are examples of very risky domain names for which this is a must..LSTM are a difficult deep learning topic. It may be difficult to understand what LSTM are and how terms such as bidirectional and collection-to-collection apply to the subject[10].The success of LSTM can be attributed to the fact that they were among the first implementations to overcome technical challenges and deliver on the promise of recurrent neural networks.

Long Short Term Memory is a recurrent neural network type. As it enters the modern era, the output of the final phase is sent into the RNN.

3.2.1 STRUCTURE OF LSTM:

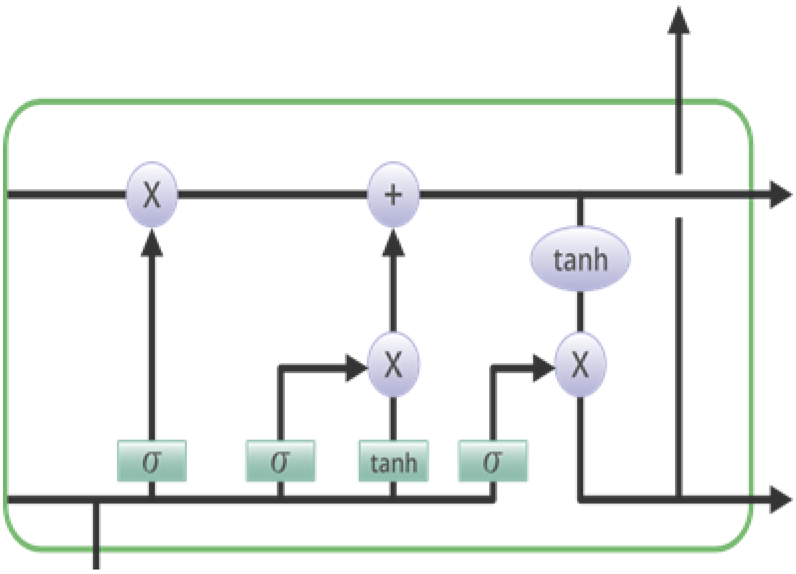

Fig 1: Structure of LSTM

3.2.2 BI- DIRECTIONAL LSTM :

The process of building a neural network that can store sequence information in both directions (backwards and forwards) is known as bidirectional long-short term memory (Bi-LSTM) (past to future).

The fact that our input travels in two directions distinguishes a Bi-LSTM from a standard LSTM. A normal LSTM can only accept input in one direction, either backwards or forwards. To preserve future and past information, we can make the input flow bi-directionally in both directions.

BI-LSTM is commonly used for sequence-to-sequence activities. This type of network can help with text categorization, speech recognition, and forecasting models.

3.3 DECONVOLUTION:

Deconvolution is a computational method that interprets the image as an estimate of the true specimen intensity and uses a point spread function formula to generate a better estimate of the intensity.In mathematics, deconvolution is the inverse of convolution. Both of these operations are used in signal and image processing. For example, using a high-precision deconvolution approach, it may be possible to recover the original signal after a filtering step (convolution). Because of the inaccuracy of the recorded signal or image measurement, it can be demonstrated that the lower the SNR, the worse the filter reversal will be; thus, inverting a filter is not always a good idea.

3.4 IMPLEMENTATION:

LSTM ALGORITHM:

- Input: Historical COVID-19 confirmed cases
- Output: Forecast for the next seven days for n Data Points.

- COVID-19 cases confirmed in the past
- Function pre-processing (confirmed cases, length of sequence)
- Cases that have been confirmed (Data set)
- Output: The data is summarized and divided into Train and Test windows. Data windows=windows(Confirmed cases, sequence length)
- normalized data =[] function normalize(Data windows)
- In Data window, for I:
- For p in I normalized window=[((float(p)/float(i[0]))-1)Standardized
- _data.append(Normalized window)
- Return normalized data
- Rows=90% of the normalized data shape
- Train data=[: row,: -1]
- Train label= [: row. -11/as sequence prediction is performed, the same data is divided to Test data= [row:.:-1]
- /train data, train label, test data, and test label
- Test labels [row:. -1]
- model(a, b, epoch, batch size)
- Input: a and b are the Train Data and Train Label, respectively.
- The trained model is obtained as an output.
- Network Sequential () Network
- add(LSTM(input, output, dropout)) /Input Layer of an LSTM RNN Network.
- add(LSTM(cells, activation, dropout))
- Network. add (LSTM(cells, activation, dropout))
- Network. add(LSTM(cells, activation, dropout))n /Hidden Layers
- Network. add(LSTM(output, output activation) /Output Layer
- Network.
- compile (loss function, optimizer) /Model Definition Optimization
- Network.fit(a, b, epoch, batch size, validation) /Training the model

3.5 ALGORITHM OF RNN:

- Pre-process the data using EMA to eliminate noise.
- Seven days of COVID-19 positive cases and deaths
- For each data collection, normalise it so that the values range from 0 to 1.
- Configure your network
- Specify the amount of RNN blocks to use and how they will be activated.
- Make the training window larger
- End-of-run predictions for the number of periods of history and batch size for the network

IV. RESULTS

4.1 UNIVARIATE ANALYSIS:

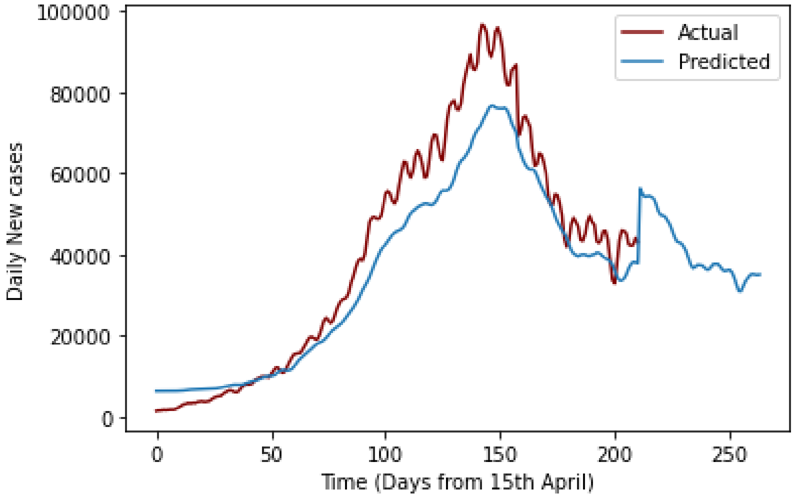

Fig 2: Uni variate Analysis

The diagram above depicts the univariate analysis.The analysis was carried out on data from all states across India.The analysis is carried out over a period of 250 days beginning April 15th.The predicted data is nearly identical to the actual data.

4.2 MULTIVARIATE ANALYSIS:

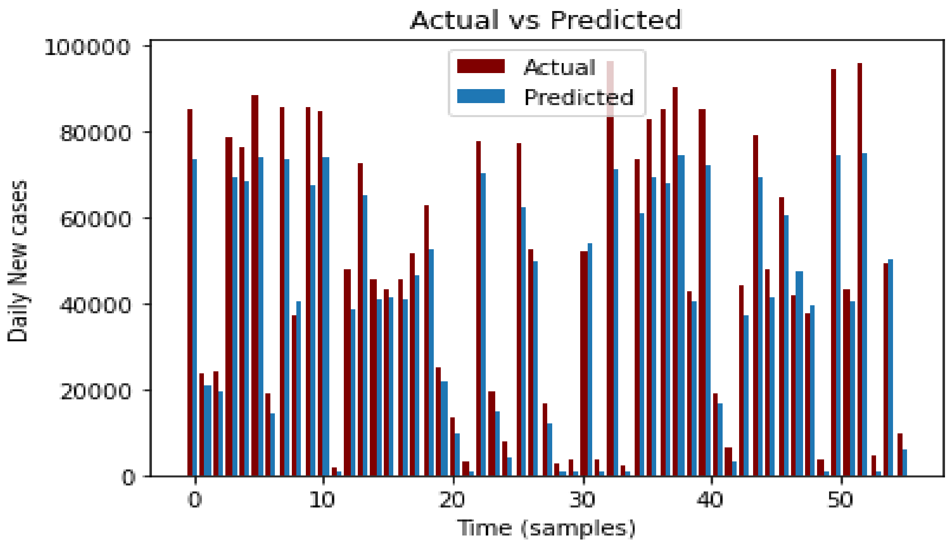

Fig 3: Multivariate Analysis

The above figure describes a multivariate analysis. Unlike the univariate analysis, this analysis was carried out state by state in India. The analysis is carried out for a period of 250 days beginning April 15th. The predicted data closely matches the actual data. The x-axis represents daily new cases, and the y-axis represents time.

4.3 SWAB TEST RESULTS:

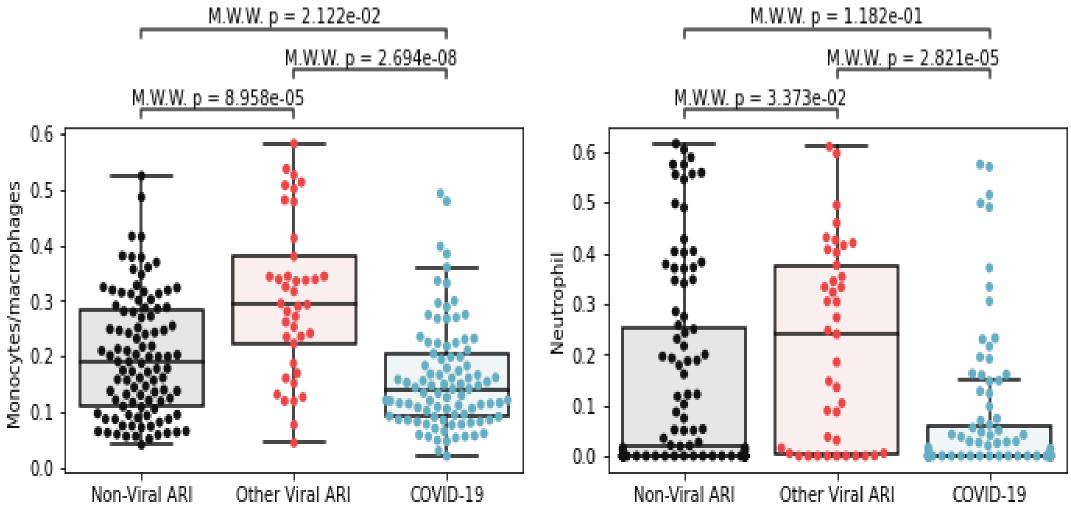

Fig4 : Swab Testing Results

The swab test is depicted above. The test is administered to three groups of people: COVID patients, other virus patients, and no-virus patients. This comparison assists in understanding how cells are structured, which aids in future vaccine research.

V. CONCLUSION

In this paper, we used the LSTM model to forecast the COVID-19 patient count. We used three models to increase the accuracy of this content.

The hospital will be fully prepared because it will know ahead of time how many patients to expect over a specific time period.

We also inserted the gene variable to learn more about how COVID-affected patients' and other people's cells are built, which will help vaccine research.

VI.REFERENCES

[1]. F. Wu et al., “A new corona virus 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 corona virus in Wuhan, China,” Lancet, vol. 395, pp. 497–506, May 2020.

[3]. Corona virus 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]. corona virus 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-corona virus2019

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

[8]. Natarajan, MS Kumar V. Anantha, and D. Ganesh Macha Babitha. "Machine Learning Based Identification of Covid-19 From Lung Segmented CT Images Using Radiomics Features." Biosc. Biotech. Res. Comm. Special Issue 14, no. 07: 350-355.

[9]. Devi, Bingi Manorama, Sandeep Vemuri, A. Chandrashekhar, C. Sushama, Praful Vijay Nandankar, and Pankaj Kundu. "Impact of COVID-19 pandemic and the diagnosis of the virus in the human body." World Journal of Engineering (2021).

[10].Ganesh D, M. Sunil Kumar, K. Balaji and T.P Kumar, "Implementation of Neural Networks For Reducing Primary User Emulation Attacks In Cognitive Radio Networks", Vidya bharati International Interdisciplinary Research Journal (Special Issue on Engineering Technologies and Management),2021, pp-176-182