2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/Confluence51648.2021.9377075

COVID-19 Outbreak: An Epidemic Analysis using Time Series Prediction Model

Raghavendra Kumar[1]
[1]Department of Information Technology
KIET Group of Institutions
Delhi NCR Ghaziabad (INDIA)
raghavendra.dwivedi@gmail.com

Anjali Jain[1]
[1]Department of Information Technology
KIET Group of Institutions
Delhi NCR Ghaziabad (INDIA)
jainanjali4u@gmail.com

Arun Kumar Tripathi[2]
[2]Department of Computer Applications
KIET Group of Institutions
Delhi NCR Ghaziabad (INDIA)
mailtoaruntripathi@gmail.com

Shaifali Tyagi[1]
[1]Department of Information Technology
KIET Group of Institutions
Delhi NCR Ghaziabad (INDIA)
shaifalityagi1809@gmail.com

***Abstract*- Coronavirus (COVID-19) epidemic affects public health infrastructure across the world. The outbreak is considered as third major Coronavirus epidemic after SARS (Severe Acute Respiratory Syndrome) in the year 2002-2003 and MERS (Middle East Respiratory Syndrome) in 2015 since past 2 decades. It has been observed that the nature of growth of coronavirus is exponential. It has been tough to control and analyze the situation with limited human resource and treatment process must be carried for the large number of patients within an appropriate time. So, it has become obligatory to work on an automated model, grounded on computing approach, for curative measure. This paper concludes a Time Series Forecasting model and analyze the COVID-19 epidemic occurrence to check whether these numbers are going to be increased or decreased in near future. Statistical pattern analysis and data visualization is performed with widely accepted time series approaches as Auto-Regressive Integrated Moving Average (ARIMA) and its constituents Moving Average (MA) and Auto Regressive (AR). Finally, time-dependent parameters can enlighten the trends of the outbreak COVID-19 in India.**

Index Terms- COVID-19, Time Series Data, ARIMA

I. INTRODUCTION

Wuhan, the capital city of Hubei province, known as a transportation hub of China. The Wuhan city is famous for wholesale market for trading of seafood, wet animals and lives animals across the world. In late December 2019, a group of adult patients reported in the city hospital with severe pneumonia symptoms. Initially, medical reports diagnosed the common symptoms of potential virus outbreak. In addition, sample reports taken from a surveillance system were sent to an etiologic investigation lab for further examination [1,2,3]. Based on the findings, China notified this virus decease to World Health Organization (WHO) on December 31st ,2019 and sealed the Wuhan market completely next day, on January 1st, 2020. WHO (World Health Organization) identified the virus as Corona virus enveloped as RNA viruses that ranging between 60nm to 140nm in diameter. The virus appears under electronic microscope, projects like a crown on surface therefore it is named Corona virus. On January 11th, 2020, it is further renamed as COVID-19 by WHO (World Health Organization) [5,6,7]. Further investigation diagnosed that, increasing exponentially cases are not limited to the place from where COVID-19 was originated. So far, COVID-19 facts were proven that community transmission was occurring [8]. In the context of India and its neighboring countries (i.e. Pakistan, Srilanka, Bangladesh, Nepal, etc.) the growth rate of infected people is still believed to be in early stage. The very first case of India in COVID-19 epidemic is revealed on 30th of January 2020 in Kerala. However, the outbreak transmission was escalated in mid of March 2020. Mostly, the reported cases are linked to people with travel history to affected countries. Despite being a nation of billion people population as per Indian government reported record [Fig-1], the confirmed cases, recovered cases and death cases are 10197, 1344 and 392 respectively as on date [9].

Since the COVID-19 outbreak has exponential growth as evolutionary nature therefore clinical process and treatment analysis are next to impossible to handle such a huge number of patients. So, it is compulsory to construct an automated model, based on computing approach, for curative measure. In this study, a Time Series Forecasting model is proposed to understand the course pattern of the COVID-19 outbreak occurrence and to check whether these numbers are going to be like in near future days. Deep Learning algorithms and Machine Learning are amply accepted mechanism to analyze epidemic time series data. Such mechanism helps to find out epidemic pattern so that spreading syndrome can be controlled and corrective measures can be taken [10,11].

Noticeably, in a very less time span most of the models are not accurate to predict the epidemic next stage and it has been a challenge for research community and govt. and private agencies. Remaining part of this paper is organized as follows. Section-2 includes related works and the brief discussion of existing models. Section-3 briefly discuss the methodology includes

prediction model based on ARIMA. Section-4 reports the result and discussion about dataset and obtained effective outcome. Section-5 finally concludes the paper and enlightens the future prospect of the study.

II. RELATED WORK

COVID-19 researchers proposed their exceptional strategies and models to early prediction of conditions around the world listed in TABLE-1. Shawni D. et al. proposed Long short-term memory and Gated Recurrent Unit on data set Covid-19 on the duration of cases containing from period being 20th of January 2020 to 12th of March 2020 and predicted LSTM and GRU models that are also finally picked up for instructing the dataset and the reports predicted by clinical doctors matched with the predicted results. The results that are predicted get confirmed against the original data based on few preset measures [15].

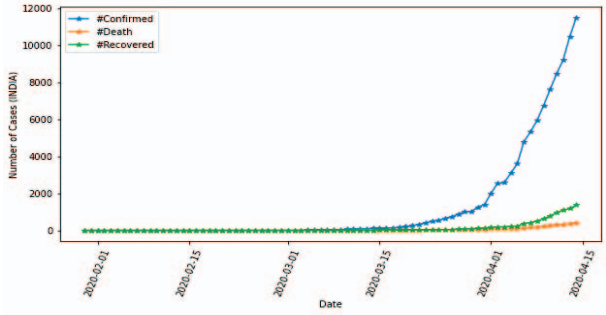

Fig.1 Growth of all cases in India

Domenico B et al. proposed Moving Average (ARIMA) model on the epidemiological data of Johns Hopkins for forecast of 11th February 2020 and 12th February 2020 predicted the epidemiological course of the occurrence of COVID-2019. For future work and comparison of real time data must be managed. In this study ARIMA parameters are considered as (1,2,0) and (1,0,4) respectively [16].

Soudeep D et al. proposed a model to obtain the occurrence of COVID-19 outbreak on dataset accumulated from JHU-CSSE (2020), the repository of GitHub which is continuously maintained by the Johns Hopkins University Center which predicted that the structure of the model is concise, and using various measures with appropriate diagnosis, they successfully captured the incidence pattern of the disease by figuring out a time-dependent quadratic trend [17].

Pavan K et al. works on machine learning approach on data set that contains COVID-19 cases having confirmed death and recovery from John Hopkins Coronavirus resource center (https://coronavirus.jhu.edu/) reported in period 21 January 2020 to 26 March 2020 predicting some track of COVID-19 in the upcoming time (until April 30, 2020) using the model ARIMA and VAR (Vector Auto Average). Proposed model is basically used to capture the spatial extinct for constructing GIS map across the world on three different variables while using remote sensing data [18].

Yunlu W et al. uses Respiratory Simulation Model (RSM) with GRU to classify six patterns of respiration. It was measured by the depth camera and examined by real-world data with 605 instances [19].

Ashutosh S et al. examined SIR model in April 2020, to capture the pattern before their respective lockdown conditions and concluded that they are working continuously each and every day for finding the parameters of the model and track the possible situations. They differentiate the exposure for the rate of infection occurred, which results in levels of quarantine.

However, the parameters of the model look much consistent [20].

Jinyu Z et al. proposed CT scans on COVID-CT dataset, which is publicly available, with 275 CT scans. These instances are obtained from (https://github.com/UCSD-AI4H/COVID-CT) of 30 March 2020. On this dataset a deep convolutional neural network is trained and measured F1 as 0.85 to prove outstanding efforts [21].

Jianpeng Z et al. proposed a new model having deep anomaly detection on data set https://github.com/ieee8023/covid-chestxray-data on 27 March 2020. Deep anomaly detection model obtains sensitivity as 96.00% and specificity as 70.65%, thus promising the proposed model for its reliable COVID-19 screening [22]. Raju V et al. recognized a set of seven remarkable applications. AI based applications help in detecting the cluster of cases and predict where this virus will affect in future with possible analysis [23].

III. METHODOLOGY

Time series data considers massive amount of data, high dimensional and update continuously. The continuous updates in numerical sequence maintain the dependency in dataset [12]. Therefore, previous value is interrelated with the upcoming values and so on. Time series analysis keeps time interval in specific period into consideration while deal with data. Stipulated Time interval generates behavior and pattern in data series which help to design a forecasting model. There are well known time forecasting models like AR, MA and ARIMA to be considered as linear model [13,14]. Proposed time series model is identified COVID-19 dataset from John Hopkins Coronavirus resource center. Data instances are considered between February 2020 to April 2020 for the statistical analysis and prediction. The aggregate dataset is used precisely for time series prediction using AR, MA and ARIMA models. Time dependent parameters can illustrate the COVID-19 outburst patterns in India.

As per study number of confirmed cases (C_c), number of recovered cases (C_r) and number of death cases (C_d) are considered to compute the recovery rate (R_r) and mortality rate (M_r).

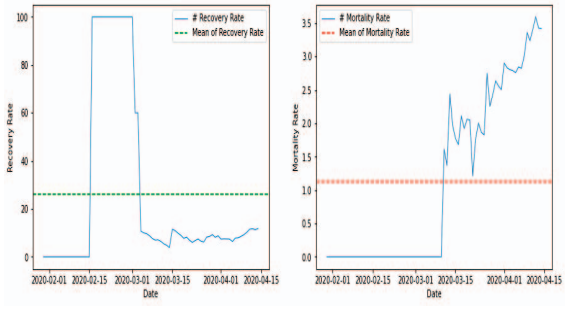

Fig.2 Recovery Rate and Mortality Rate of India over Date

TABLE-1 EXISTING WORK

Article [Ref]	Country/Dataset	Dataset Duration	Classifiers/ Method(s)	Performance Measure(s)	Findings
Bandyopadhyay S. et al. (2020) [15]	COVID-19 (South Korea)	20th Jan 2020 - 12th March 2020	LSTM, GRU	Accuracy, Precision, Recall	COVID-19 Cases: Positive, Negative, Death, Release
Domenico B. et al. (2020) [16]	CSSEGIS and Data. (France)	20thJan2020-10th Feb 2020	ARIMA	Accuracy, Recall	Epidemiological trend
Soudeep Deb et al. (2020) [17]	CSSEGIS and Data. (France)	20thJan2020-10th Feb 2020	Time series	Accuracy, Precision, Recall	Time-dependent quadratic trend
Pavan K et al. (2020) [18]	John Hopkins Coronavirus resource center	21st Jan 2020- 26th Mar 2020	ARIMA	Precision	Prediction of spatial extinct
Yunlu Wang et al. (2020) [19]	Real-world data	Dec 2019 - Feb 2020	Bidirectional and attention mechanisms	Accuracy	Classified 6 clinically respiratory patterns
Ashutosh S. et al. (2020) [20]	Covid-19 (Europe, India)	January 2020 - April 2020	Stochastic SIR model	Precision, Recall	Prediction
Authors Jinyu Z. et al. (2020) [21]	UCSD-AI4H /COVID-CT	Dec2019 - March 2020	Transfer Learning and Data Augmentation	Accuracy, Precision, Recall	Prediction
Jianpeng Z. et al. (2020) [22]	UCSD-AI4H/COVID-CT	Jan 2020 - March 2020	Deep anomaly detection model	Accuracy, Precision	Prediction
R . Vaishya et al. (2020) [23]	Pubmed, Scopus and Google Scholar	Dec2019 - Apr 2020	Coronavirus and Artificial Intelligence	Accuracy, Precision, Recall	Prediction

$$\text{Mortality Rate } (M_r) = (C_d/C_c) \times 100 \quad (1)$$

$$\text{Recovery Rate } (R_r) = (C_r/C_c) \times 100 \quad (2)$$

Initially, recovery rate was higher when the number of confirmed cases was low due to testing rate was too low.

The higher side of mortality rate and dropped recovery rate has been an upsetting factor for India. In the same way, increasing mortality rate and recovery rate is convincing indication for increase in number of recovered cases. Recovery rate and mortality rate are still under control since the confirmed cases are low due to lower rate of corona testing in the country particularly on hotspot areas. However, testing rates are less due to being in first stage of COVID-19 outbreak in India. COVID-19 prediction model (Fig-3) illustrates the complete process. The aggregate COVID-19 dataset instances are divided in 80%-10%-10% as training, testing and validation dataset respectively used precisely for time series prediction model. Data splitting ensure the model to overcome from over fitting and under fitting problem. In addition, COVID-19 dataset is normalized with z-score normalization. Missing values imputation is done with mean values of previous and next day confirmed cases of outbreak.

IV. RESULT & DISCUSSION

Finally, dataset is preprocessed and get ready for experiments performed with the help of python libraries and in-build packages. ARIMA model enhance the feature of Auto Regressive (AR) and the Moving Average (MA) model. ARIMA model (Fig-3) illustrated prediction of the confirmed cases in India. Finally, model is evaluated with root mean square error (RMSE) to check forecasting accuracy.

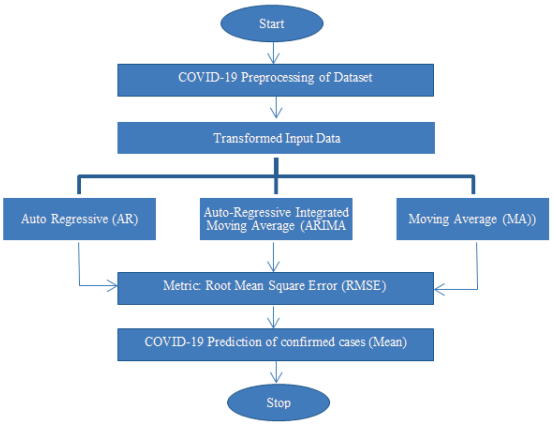

Fig.3 COVID-19 Prediction Model

The objective to apply ARIMA, AR and MA model is to find the time trend that changes after certain interval. Table-2 illustrates the five entries predicted for the duration of 17-21 April 2020.

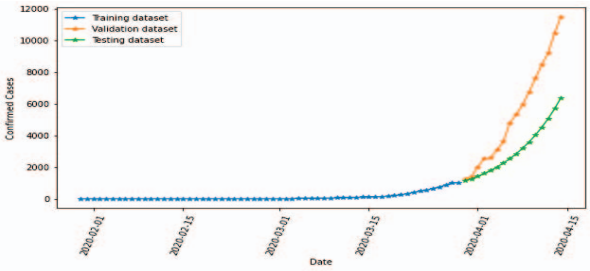

Fig.4 ARIMA Prediction Model

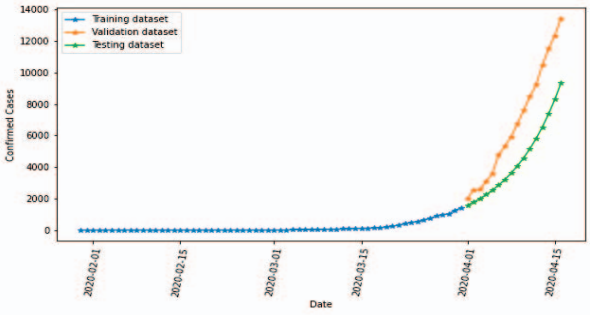

Fig.5 AR Prediction Model

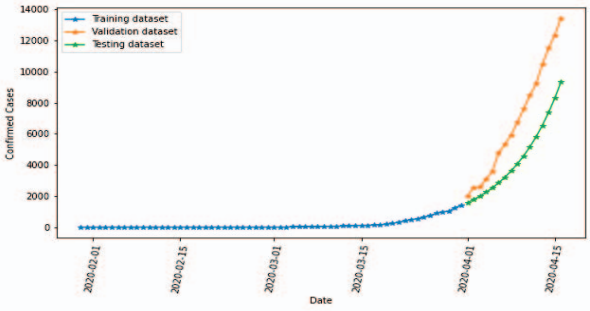

Fig.6 MA Prediction Model

TABLE-2 COVID-19 PREDICTION AS ON DATE

Date	AR Model	MA Model	ARIMA Model	Mean of Predictions
April 17 2020	14681	14562	14691	14644
April 18 2020	16583	16443	16595	16540
April 19 2020	18732	18568	18746	18682
April 20 2020	21159	20967	21175	21100
April 21 2020	23901	23676	23919	23832

RMSE (Root Mean Square Error) is considered as performance metrics for prediction error that is illustrated on Table-3. The average RMSE to mentioned model obtained 1096.975.

$$RMSE = \sqrt{\sum_{i=1}^{n}(Xi - X'i)^2/n} \quad (3)$$

RMSE is a frequently used error measure that takes Root of MSE [13]. It obtains error gap analysis between the observed (X_i) and predicted values($X`_i$) mentioned in equation-3.

TABLE-3 COVID-19 RMSE AS ON DATE

Performance Measure	AR Model	MA Model	ARIMA Model	Mean of RMSE
RMSE	1083.366	1128.500	1079.058	1096.975

V. CONCLUSION

COVID-19 epidemic has evolved as a challenge to the world's public health infrastructure, medical facilities, economical and societal aspects. The imposed outbreak challenges are likely to be greater in developing countries like India due to limited community measures. This study proposed to utilize time series model for COVID-19 epidemic analysis using John Hopkins Coronavirus resource center. Time series model imposed Auto-regressive integrated moving average (ARIMA), Moving Average (MA) and Auto Regressive (AR) model to predict confirmed cases in India. Time series models are capable to estimate the current situation of COVID-19 outbreak. Proposed model can assist to public, private and government agencies as supplementary method to design and act for decision making policies. The model can help to minimize the pandemic affects in country by providing accurate results which further be extended and validated with other machine learning and deep learning models. However, becoming unsocial, maintaining community distance and disciplined lockdown execution are the key factors to minimize and stop COVID-19 pandemic affects.

REFERENCE

[1] C. Wang, Horby PW, Hayden FG, Gao GF "A novel coronavirus outbreak of global health concern," Lancet., vol.395, pp. 470-473, 15 Feb. 2020.

[2] A. Bogoch, Watts, A. Thomas-Bachli, C. Huber, M.U.G. Kraemer, K. Khan, "Pneumonia of unknown etiology in wuhan, China: potential for international spread via commercial air travel," mar. 2020.

[3] Coronavirus Outbreak. Available at: https://www.worldometers.info/coronavirus/. Accessed 14 April 2020.

[4] H. Lu, C.W. Stratton, Y.W. Tang, "Outbreak of pneumonia of unknown etiology in wuhan China: the mystery and the miracle," J. Med., Vol. 92, pp.401–402, April 2020.

[5] World Health Organization. Situation reports. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/. Accessed 22 Feb,2020.

[6] S. Zhao, Q. Lin, J. Ran, S.S. Musa, G. Yang, W. Wang, et al., "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak," International journal of infectious diseases. Vol. 92, pp.214–217, Mar.2017.

[7] D.D. Richman, Whitley RJ, Hayden FG., "Clinical Virology," 4th ed. Washington: ASM Press; 2016.

[8] C. Huang, Wang Y, Li X, “Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.”, Lancet. Vol.395, pp.497–506, Feb.2020.

[9] Coronavirus Outbreak. Available at: https://www.covid19india.org/. Accessed April 2020.

[10] Singh, R., Singh, R., Bhatia, A. “Sentiment analysis using Machine Learning technique to predict outbreaks and epidemics.”, vol.3, pp.19-24, March 2018.

[11] Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. “Application of the ARIMA model on the COVID-2019 epidemic dataset.”, vol.29, April. 2020.

[12] Box, G. E., Jenkins, G. M., Reinsel, G. C., &Ljung, G. M., “Time series analysis: forecasting and control,” John Wiley & Sons ,2015.

[13] Kumar R., Kumar P., Kumar Y., “Time Series Data Prediction using IoT and Machine Learning Technique,” Procedia computer science, Vol.167, pp. 373-381, 2020.

[14] Fu, T. C., “A review on time series data mining.” Engineering Applications of Artificial Intelligence. Vol. 24, pp.164-181, Feb. 2011.

[15] Bandyopadhyay, S. K., & Dutta, S., “Machine Learning Approach for Confirmation of COVID-19 Cases: Positive, Negative, Death and Release,” medRxiv, 2020.

[16] Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). “Application of the ARIMA model on the COVID-2019 epidemic dataset.” Data in brief., 2020.

[17] Deb, Soudeep, and Manidipa Majumdar. "A time series method to analyze incidence pattern and estimate reproduction number of COVID-19." arXiv preprint arXiv:2003.10655,2020.

[18] Kumar, Pavan, et al. "Forecasting the dynamics of COVID-19 Pandemic in Top 15 countries in April 2020: ARIMA Model with Machine Learning Approach." medRxiv, 2020

[19] Wang, Y., Hu, M., Li, Q., Zhang, X. P., Zhai, G., & Yao, N. “Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner.” arXiv preprint arXiv2002.05534.

[20] Simha, A., Prasad, R. V., & Narayana, S., “A simple Stochastic SIR model for COVID-19 Infection Dynamics for Karnataka after interventions–Learning from European Trends,” arXiv preprint arXiv:2003.11920, March 2020.

[21] Zhao, J., Zhang, Y., He, X., & Xie, P., “COVID-CT-Dataset: a CT scan dataset about COVID-19,” arXiv preprint arXiv:2003.13865, June 2020.

[22] Zhang, J., Xie, Y., Li, Y., Shen, C., & Xia, Y., “Covid-19 screening on chest x-ray images using deep learning-based anomaly detection.” arXiv preprint arXiv:2003.12338, Dec-2020.

[23] Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A., “Artificial Intelligence (AI) applications for COVID-19 pandemic,” Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol.14, pp.337-339, Aug.2021.