How to Cite:
Thilaga, P. J., Kavipriya, S., & Vijayalakshmi, K. (2022). Predictions of confirmed and death caused by COVID-19 in India. *International Journal of Health Sciences*, *6*(S8), 5401–5407. Retrieved from
https://sciencescholar.us/journal/index.php/ijhs/article/view/13463

Predictions of confirmed and death caused by COVID-19 in India

P Jothi Thilaga
Assistant Professor, Department of Computer Science and Engineering, Ramco Institute of Technology, Rajapalayam, India
Corresponding author email: jothithilaga@ritrjpm.ac.in

S Kavipriya
Associate Professor, Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, India
Email: urskavi@mepcoeng.ac.in

K Vijayalakshmi
Professor & Head, Department of Computer Science and Engineering, Ramco Institute of Technology, Rajapalayam, India
Email: vijaylakshmik@rirjpm.ac.in

***Abstract*---**COVID-19 is spreading within the sort of a massive epidemic all over the world. This epidemic affects a lot of individuals in India. The World Health Organization states that COVID-19 could be spread from one person to another at a rapid manner through contact and respiratory spray. On these days, India and all countries worldwide should rise to an effective step to investigate this disease and eliminate the effects of this epidemic. The proposed work presents about the detailed forecasting model and prediction of the number of confirmed, recovered, and death cases in India caused by COVID-19 using machine learning algorithms. The multiple linear regressions and correlation coefficients have been applied for prediction and auto-correlation and auto-regression have been used to improve the accuracy.

***Keywords*---**COVID-19, machine learning, linear regression, prediction, correlation.

Introduction

COVID-19 is a most infectious respiratory disease caused by one of the coronavirus family members, has led to a pandemic situation worldwide in 2020.

At first, the virus was identified in Wuhan city of China in December 2019. This infectious disease has taken minimum of three months' time period to widely spread all over the world.In India, the first person affected by coronavirus disease 2019 (COVID-19) was announced on 30th January 2020. Thisvirus spreads to the whole of India (in their differentdistricts) till April 2020 end. During the spread of the epidemic, it is a greatest significance to use the effective data for forecasting and prediction through machine learning model for the analysis of the entire epidemic [3].

The virus ofCOVID-19 reflects immense similarity withthe Severe Acute Respiratory Syndrome (SARS) andMiddle East Respiratory Syndrome (MERS) coronavirusas investigated by pathologists. It can spreadbetween humans and easily transmit from human tohumanthrough droplets of coughs or sneezing by an infected person [5]. Major symptoms of COVID-19 are fever,cough, diarrhea and breathing trouble.The major problem in the case ofthis disease is that its symptoms generally come outwith the symptoms of only after one to two weeks if an individual gets infected. This phase isknown as the incubation period and the mean incubationperiod is approximately seven days. An infected personmay also infect the numerous healthy persons during theincubation phase. These patients are asymptomaticandmajor dispute is to recognize them. The rate at whichone infected person transmits this disease into othersis termed as transmission rate [8].

Machine learning algorithms act an important part in disease prediction and analysis. Though COVID-19 is a new infectious disease still works have done in COVID-19 prediction using machine learning. Continuous contribution is going on in this field of research [1]. Also an LSTM, whichisa special type of recurrent neural network (RNN) that usesmemory units to serve as short-term memory gives an outstanding performance on speech recognition, handwriting recognition, and pattern recognition, can be applied for the prediction of this infectious diseases. LSTM's prediction performs well on temporal data [16].

Related Works

Nanning Zheng et al. [14] developed a hybrid AI model for the timely prediction of the Covid-19 death, recovery and spreading cases in China. The time distribution of the contagioussources of newly confirmed cases must be determined toexamine the infection law of an epidemic. The major task of this article is to analyze the spread laws and development trend of an epidemic by modeling new confirmed data. Safynaz Abdel et al. proposed a predictive analysis of Covid-19 severity using machine learning techniques [17]. Early predictions of severity of patient will helps in decrease the continual death of patients by providing early medication actions.The rate of reproduction in virus among people is related to many factors, comprises the average number of contacts a person has, number of days a person is infected, from the day of exposure to the infection, the number of active cases, values of index of stringency, testing capacity and positivity, and several other factors [2]. The outcome includes the reproduction rate, time curve, and future values cannot be satisfactorily estimated by the probability functions of distribution alone.

Similarly, Muhammad Iqbal et al projected a long short term memory (LSTM) model to forecast confirmed cases of COVID-19 [6]. It is highly precise that LSTM is a vast data-driven model, which is not appropriate for small sample learning problems and leads to a high risk of over fitting that is easy to take place in small sample learning. Karen Panetta at el proposed a work on detection of Covid-19 using shape-dependent Fibonacci-p patterns for efficiently categorize COVID-19, viral pneumonia and normal condition chest radiographs [4].

One of the complications for generating an accurate model forCOVID-19 prediction is arriving to the parameter valuerequired for the model to ensure unbiased and reliability. A complex model can be used with more complex data ofepidemiological and biological information [18]. SEIR is a basic mathematical model used for various infectious diseases. During this pandemic, higher number of researchers uses SEIR model to predict the basic reproduction number of the infectious cases. [7] SEIR model comprises the objects such as 'S'represents the suspected person, 'E' represents the expected person, 'I' represents the infected person and 'R'represents the recovered person.

Proposed Work

In this proposed system, a real dataset for the COVID-19 after the outbreak were used. Analysis of the COVID-19 dataset for coronavirus disease is performed on the basis of reported cases in India. The data from January 2020 to August 2021 were collected and used for prediction.For this prediction model, three different attributes of dataset were considered such as confirmed, recovered, and death cases. Target classes used in prediction dataset includes multiple discrete instances. These target classes are thefollowing:

(1) Confirmed cases: Number of cases confirmed at any particular date. It may be increased or decreasedbasedon next date, time, and location-specific to theIndian states only.
(2) Death cases: Number of death at any particular date in Indian states. It may be increased or decreasedaccording to next date, time, and location in Indian states only.
(3) Recovered cases: Number of recovered cases at anyparticular date. It may be increased or decreasedaccording to the next date, time, and specific state in India.

The proposed model is depicted in Fig. 1. Itis important to discover and compute the degree ofvariables in the dataset and this information is helpfulfor better preparation of dataset to meet the expectationsof machine learning algorithms. Recovery strategy and correlation analyses are performed on data using PythonSoftware. It reveals a statistical summary of confirmed, recovered, and death cases and also finds a strongrelationship among current data.

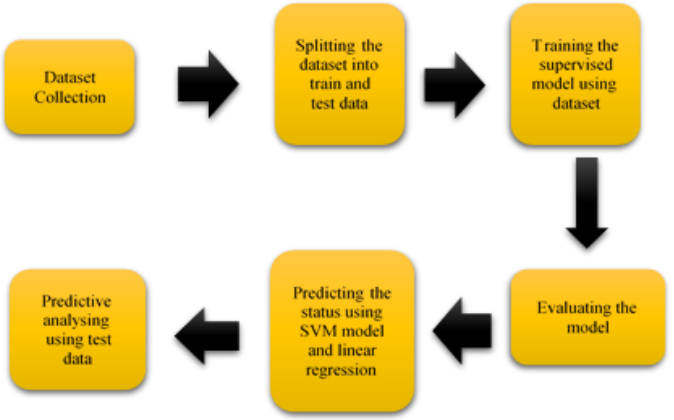

Fig. 1*:* Proposed model for predictions of confirmed, recovered, and death cases of COVID-19 in India

Implementing linear regression is straightforward, and it is simpler to understand the output coefficients. If the independent and dependent variable have a linear relationship, this type of algorithm is the best to use since it is less complex when compared with other algorithms. The major limitation of Logistic Regression is the assumption of linearity between the independent variable and the dependent variables. Non-linear problems could not be solved with logistic regression because it has a linear decision surface.

During the analysis of the complete dataset, it is observed that the model can use regression against itself and also able to use the autocorrelation plot to check the randomness within the data. Results prove that the forecasted range of time series is accurate. Now fit the model using the existing dataset and find the lag and coefficients. Based on the value of lag, different categories of confirmed, recovered, and death cases analysis is performed.

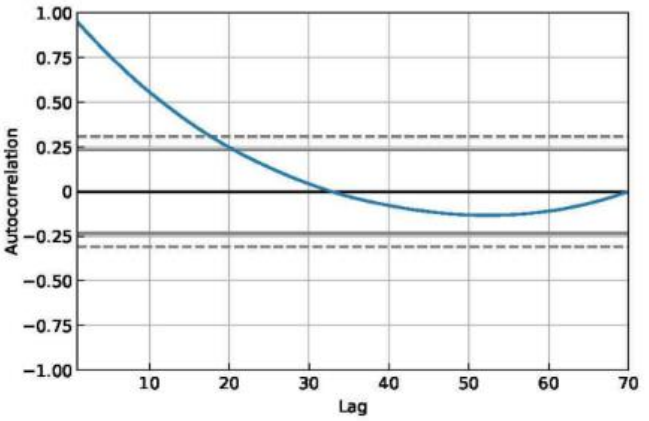

Fig. 2: Autocorrelation plotting of various cases

Results and Discussions

In this proposed study, a real dataset for the COVID-19 after the outbreak of the epidemic in India. The first case of the COVID 2019 epidemic was found in India on 30th January 2020 after that, things escalated in March, several cases were reported all over the country at the end of March leadsto radical loss of human lives. The distribution of data in the training and testing experiments has been set to 70% and 30%, respectively. The prediction of confirmed cases based the data from January 2020 to August 2021 were depicted in the Fig. 3. The technique of Decision tree learning is used to continuously split training and test data according to a specific parameter. It is a broadly accepted supervised learning approach that splitting the dataset based on certain conditions. It is equally suitable for classification and regression. This approach gives the most realistic class for each record for classification.

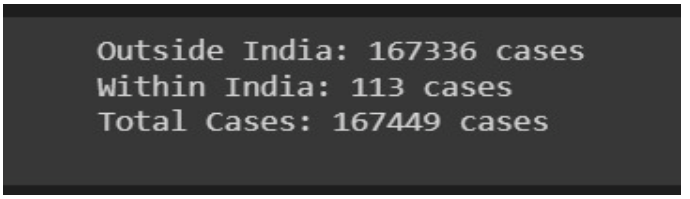

Fig. 3: Prediction of confirmed cases of COVID-19 in India

To find the effectiveness of the proposed work, various confirmed cases of worldwide is compared and analysed. Fig. 4 shows the histogram for the worldwide comparison of accuracy of confirmed cases, respectively using the random forest model. In further, Multiple linear regression and autoregression were used to predict the possible number of cases.

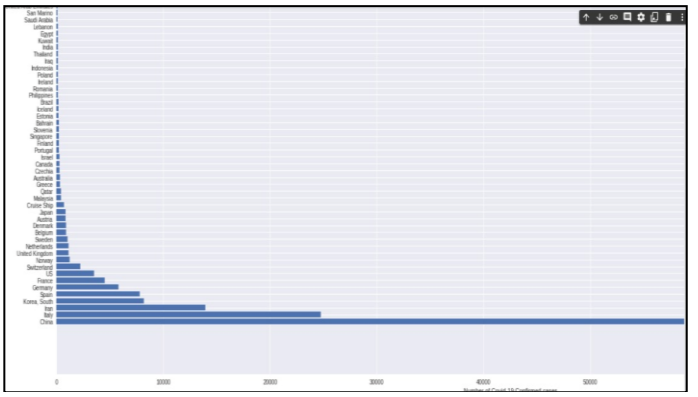

Fig.4: Worldwide plot of confirmed cases of COVID-19

Conclusion and Future Work

In this study, the spread of COVID-19 in worldwide, specifically in India was explored, and a model was put up to forecast the number of confirmed and death cases. The potential number of instances in the future was predicted using multiple linear regression and auto-regression. With adequate forecasting and rapid action, the loss of human life may be reduced. This prediction may be

useful in resource management, such as health services. To forecast when this pandemic will cease in a specific area, the proposed model may be expanded. This model can be used to forecast overall causality as well as total economic losses. In the future, the proposed model can be extended to implement the model on death cases and will try to forecast the death ratio and its correlations with the critical cases.

References

1. Choujun Zhan, Chi K. Tse, Ying Gao, and TianyongHao, "Comparative Study of COVID-19 Pandemic Progressions in 175 Regions in Australia, Canada, Italy, Japan, Spain, U.K. and USA Using a Novel Model That Considers Testing Capacity and Deficiency in Confirming Infected Cases", IEEE Journal of Biomedical and Health Informatics, Vol. 25, No. 8, June 2021.
2. FurqanRustam, Aijaz Ahmad Reshi, ArifMehmood, SaleemUllah, Byung-Won On, WaqarAslam, And Gyu Sang Choi, "COVID-19 Future Forecasting Using Supervised Machine Learning Models" IEEE Access, DOI: 10.1109/ACCESS.2020.299731. June 2020.
3. HamzaTurabieh and Wahiba Ben AbdessalemKaraa, "Predicting the Existence of COVID-19 using Machine Learning Based on Laboratory Findings", International Conference of Women in Data Science at Taif University, 2021.
4. J. Christopher Clement; VijayaKumarPonnusamy; K.C. Sriharipriya; R. Nandakumar, "A Survey on Mathematical, Machine Learning and Deep Learning Models for COVID-19 Transmission and Diagnosis", IEEE Reviews in Biomedical Engineering, Vol. 15, 2022.
5. J. RelinFrancisRaj, K.Vijayalakshmi, S.KaviPriya, "Medical image denoising using multi-resolution transforms", Measurement, Vol. 145, October 2019, Pages 769-778.
6. Jayakumar Kaliappan, Kathiravan Srinivasan, Saeed Mian Qaisar, Karpagam Sundararajan, Chuan-Yu Chang and Suganthan C, "Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction Rate" Frontiers in Public Health, Volume 9, Article 729795, September 2021.
7. Karen Panetta, Foram Sanghavi, Sos Agaian, and Neel Madan, "Automated Detection of COVID-19 Cases on Radiographs using Shape-Dependent Fibonacci-p Patterns", IEEE Journal of Biomedical and Health Informatics, Vol. 25, No. 6, June 2021.
8. M. Jayalakshmi, LalitGarg, K. Maharajan, K. Jayakumar, Kathiravan Srinivasan, Ali Kashif Bashir and K. Ramesh, "Fuzzy Logic-Based Health Monitoring System for COVID'19 Patients", vol.67, no.2, CMC, 2021.
9. Muhammad Iqbal, Feras Al-Obeidat, FahadMaqbool, SaadRazzaq, Sajid Anwar, AbdallahTubaishat, Muhammad Shahrose Khan, and Babar Shah, "COVID-19 Patient Count Prediction Using LSTM", IEEE Transactions on Computational Social Systems, Vol. 8, No. 4, August 2021.
10. Muhammad Iqbal;Feras Al-Obeidat;FahadMaqbool;SaadRazzaq;Sajid Anwar;AbdallahTubaishat;Muhammad Shahrose Khan;Babar Shah, "COVID-19 Patient Count Prediction Using LSTM", IEEETransactions on Computational Social Systems, Vol. 8, Issue 4, 2021.
11. Nanning Zheng, Shaoyi Du, Jianji Wang, He Zhang, Wenting Cui, Zijian Kang, Tao Yang, Bin Lou, Yuting Chi, Hong Long, Mei Ma, Qi Yuan, Shupei

Zhang, Dong Zhang, Feng Ye, and JingminXin, “Predicting COVID-19 in China Using Hybrid AI Model”, IEEE Transactions on Cybernetics, 2020.

12. Prajoy Podder, M. Rubaiyat Hossain Mondal, “Machine Learning to Predict COVID-19 and ICURequirement”, 11th IEEE International Conference on Electrical and Computer Engineering (ICECE), 2020. DOI: 10.1109/ICECE51571.2020.9393123
13. PratimaKumari, DurgaToshniwal, “Real-time estimation of COVID-19 cases using machine learning and mathematical models – The case of India”, 15th (IEEE) International Conference on Industrial and Information Systems (ICIIS) 2020.
14. Rajani Kumari, Sandeep Kumar, Ramesh Chandra Paonia, Vijander Singh, Linesh Raja, VaibhavBhatnagar, and Pankaj Agarwal, “Analysis and Predictions of Spread, Recovery, and Death Caused by COVID-19 in India”, Big Data Mining and Analytics, ISSN 2096-0654 01107 pp65-75 Vol.4, No.2, June 2021 DOl: 10.26599/BDMA.2020.9020013.
15. Safynaz Abdel-Fattah Sayed;Abeer Mohamed Elkorany;Sabah Sayed Mohammad, “Applying Different Machine Learning Techniques for Prediction of COVID-19 Severity”, IEEE Access, Vol.9, 2021.
16. Vishan Kumar Gupta, Avdhesh Gupta, Dinesh Kumar, and Anjali Sardana, “Prediction of COVID-19 Confirmed, Death, and Cured Cases in India Using Random Forest Model”, Big Data Mining and Analytics ISSN 2096-0654 06/07 pp116–123 Vol 4, No. 2, June 2021. DOI: 10.26599/BDMA.2020.9020016.
17. Yongbei Zhu;Shuo Wang;Siwen Wang;Qingxia Wu;Liusu Wang;Hongjun Li;Meiyun Wang;MengNiu;YunfeiZha;JieTian, “Mix Contrast for COVID-19 Mild-to-Critical Prediction”, IEEE Transactions on Biomedical Engineering, Vol. 68, Issue 12, 2021.
18. Yunxiang Liu, Yan Xiao, “Analysis and Prediction of COVID-19 in Xinjiang based on Machine Learning”, 5th IEEE International Conference on Information Science, Computer Technology and Transportation (ISCTT), 2020.