Prediction of COVID-19 Cases Using Machine Learning Techniques

S. Anil Kumar[1], G. Naveena[2]

[1]Assoc. Prof, Dept of CSE, Tirumala Engineering College, Narasaraopet, Guntur, A.P., India

[2]PG Scholar, Dept of CSE, Tirumala Engineering College, Narasaraopet, Guntur, A.P., India

ABSTRACT - COVID-19, a new flu virus in 2020-21, created a widespread epidemic that heavily damaged the global economy and people's health; however, the societal consequences are unknown. COVID-19 can be detected using two types of tests: a viral test that identifies infected patients and an antibody test that enables us to detect if patients have had a previous infection. The following assays use reverse transcription, polymerase chain reaction (PCR), lateral flow immunochromatographic testing, and ELISA-like immunoassay procedures. For this article, we've developed and deployed a system that uses disruptive technologies like artificial intelligence and intelligent computing to detect increases in Covid-19 cases. Big Data and other information-science advances, such as the ability to quickly and accurately identify relevant molecular, cellular, and biochemical factors for the early detection of COVID-19, can help the scientific community.

Keywords: Covid-19, Machine learning, Forecasting

I. INTRODUCTION

Diseases such as the common cold and Middle East Respiratory Syndrome (MERS) are among the illnesses caused by the Coronavirus, a group of viruses that is also responsible for Severe Acute Respiratory Syndrome (SARS) [6]. Coronaviruses termed MERS-CoV and SARS-CoV are responsible for the spread of these two diseases. SARS was initially discovered in China in 2002, while MERS first appeared in Saudi Arabia in 2012 [8]. SARS-COV-2, a new virus found in Wuhan, China, causes the corona virus.

Wuhan, China, diagnosed with pneumonia of unknown origin on December 31, 2019, was first reported to the World Health Organization (WHO) Country Office in China [1]. At the same time, the cases of the Coronavirus and the resulting fatalities have been rising. The virus, which moves quickly, spread to the whole country in about 30 days [5]. The WHO called it COVID-19 on February 11, 2018[5]. Electronic devices powered by AI can play a significant role in stopping the spread of this infection by transmitting it from person to person. As the responsibility of epidemiologists in healthcare has increased, so has the spread of electronic health data [13]. Healthcare research and application are given a big boost by increasing electronic health data available [4]. Such data can train artificial intelligence algorithms, which can then be employed to increase AI's ability to make accurate predictions about the disease. According to statistics, as of May 16, 2020, an astounding 44,250,485 cases of COVID-19 had been recorded, and 3,020,059 people had died as a result [3]. Around 213 nations and territories are affected by COVID-19. The hospital systems had to bear a significant weight as more and more people contracted the Coronavirus, a pathogen they had inadequate medical resources to combat. Medical therapy is regularly hampered by a lack of resources in hospitals and the time lag in getting the results of medical tests. Due to the rising demand for coronavirus testing, it is no longer practicable due to time and money constraints. We want to employ machine learning techniques in our study to forecast the coronavirus infection of patients.

II. RELATED WORK

The previous work was carried out worldwide with an old dataset broken into three sub-datasets: confirmed, death, and recovered. Each of these sub-datasets had attributes such as country, province/state, and dates. Machine learning models like linear regression, SVM, and exponential smoothing were used to model the data. Despite previous models that were effective in their outcomes, Es stood out among them and performed well for the next few days' forecast and verified results.

Nanshan Chen and his colleagues at Jinyintan Hospital in Wuhan, China, conducted a retrospective, single-center examination of various patients' data. They also wrote about epidemiological data, including signs and symptoms, CT findings, and clinical outcomes in this project[16]. Although it does not focus on the COVID-19, this initiative has a positive impact on clinical outcomes.

Wang et al. found CT changes of people suffering from COVID-19 in China through their study of CT pictures. He uses deep learning approaches to design a new diagnostic tool to draw out COVID-19's graphic properties from CT scan pictures. CT pictures of verified COVID-19 patients, as well as those who have pneumonia, have been obtained. From their study, the proof-of-principle for AI COVID-19 prediction[7] may be found. In contrast to our study, which relies on clinical features and laboratory results, this project uses CT Scan images.

The authors of this paper, all from Zhongnan Hospital in Wuhan, China, have described the epidemiological, demographic, clinical, laboratory, and treatment data from their institution. Data was captured and analysed to serve as a tracking tool for disease outbreaks[6]. The author provides insight into our prognosis of COVID-19 by looking at the radio-logical and therapy data.

Halgurd S. Maghdid et al. have offered a new method for detecting coronavirus sickness using smartphones' onboard sensors. A set of machine learning models was created to collect sensor data to predict the grade of pneumonia and infection risk of disease [26]. A methodology that uses CT scan images as the primary way for predicting COVID-19 is being presented. The system uses multithreading from several sensors relevant to COVID-19 symptoms.

As an alternative diagnosis option to COVID-19, Ali Narin et al. created an automatic detection system. The study mentioned in this paraphrase ("three different convolutional neural network-based models (ResNet50, InceptionV3, and Inception-ResNetV2) have been proposed for the detection of coronavirus pneumonia infected patient using chest X-ray radiographs [12]") relates to how three different models were developed to detect coronavirus pneumonia. He also elaborates on the accuracy of the CNN classification results for the three models.

III. PROPOSED ARCHITECTURE

A paraphrase: Polynomial and Prophet have implemented and tested accessible alternatives to past work. Polynomial functions are analogous to linear functions. But it has successfully dealt with the problem of exponential growth, and Facebook's prophet modelling has grown to feature a new development of predicting models. This is a time series addiction model that can be adjusted to meet weekly, yearly, and daily trends, and so on via assigning value to object (periods). Several features are accurate and perform well, making them a good choice for implementation. Figure 1 depicts the suggested architecture.

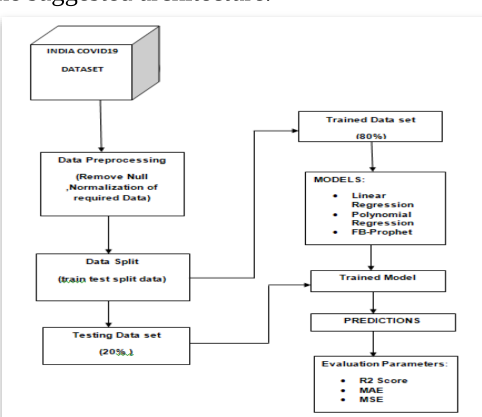

Figure 1. Proposed architecture

IV. RESULTS AND OBSERVATION

This work aims to create a machine learning system that can estimate the amount of future COVID-19 patients. Information about new COVID-19 India cases, recoveries, and fatalities was obtained from daily reports, and it was incorporated into the dataset utilized for the study. Confirmed cases and fatality rates are rising daily, making the situation for India extremely alarming. No one knows how many people were impacted by the COVID-19 outbreak in India. This investigation aims to identify the number of people at risk of contracting new infections and suffering from additional deaths, as well as the number of projected recoveries in the following 31 days. Four machine learning models—LR, Prophet, and Random Forest—were utilized to predict the number of newly infected patients, deaths, and recoveries.

The purpose of the study is to predict future time-series data. A time series is a series of measurements acquired over time, usually at equal intervals. Series of Events -> Times Because time series data are utilized in several projects, and forecasting has been an essential element. Time series data in medical, weather forecasting, biology, supply chain management, and stock price forecasting, among other fields, can be found in many places. The ability to influence forecasting is an essential element for Covid19 since it helps in the decision-making process by determining how much influence various future values could have on good results. Soon.

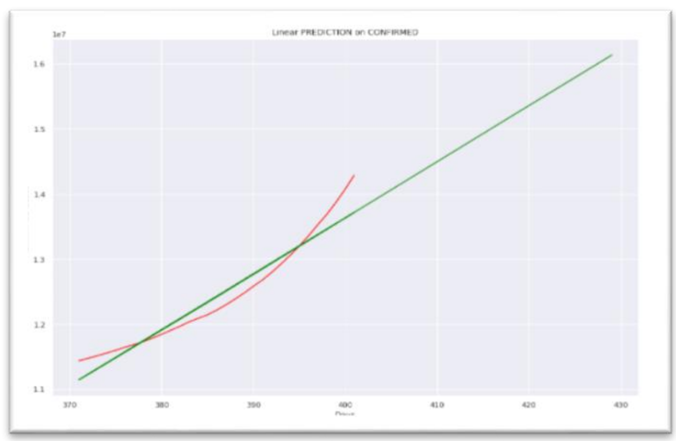

Figure 2. Predicted Results of Linear Regression Model

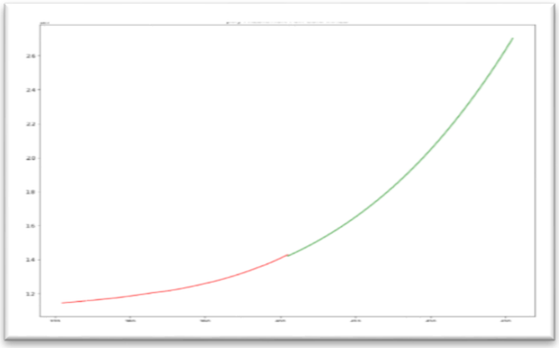

Figure 3. Predicted Results of Poly Regression Model

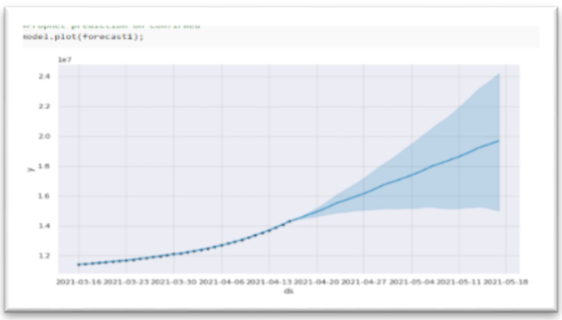

Figure 4. Predicted Results of Prophet Model

Table 1. Performance Comparison of Models

Model	R2 Score	MSE	MAE
New Confirmed Cases			
LR	0.94	47928802355.67	184958.24
Poly Regression	0.98	2455500935.20	44445.27
Prophet	0.99	6494037295.70	49185.08
New Death Cases			
LR	0.92	1905508.24	1029.85
Poly Regression	0.99	27146.15	145.07
Prophet	0.99	165738.70	299.93
New Recovered Cases			
LR	0.93	11571144694.69	93100.18
Polynomial	0.96	1221354937.30	30322.40
Prophet	0.99	1549967303.91	27903.44

V. CONCLUSION

To predict the future spread of the covid19 outbreak in India, two ML techniques were used in this work. The analysis shows that prophet outperforms all other models tried thus far, and polynomial is somewhat effective. Based on historical observations, the study suggests that mortality and recoveries will both decline. To boost the accuracy of our prediction methodology, we plan to update our data collection methods and employ machine learning to enhance the appropriateness of our model. Our future endeavours will involve real-time live forecasting and subgrouping customers by age.

VI. REFERENCES

[1] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," PLoS ONE, vol. 13, no. 3, Mar. 2018, Art. No. e0194889.

[2] G. Bontempi, S. B. Taieb, and Y.-A. Le Borgne, "Machine learning strategies for time series forecasting," in Proc. Eur. Bus. Intell. Summer School. Berlin, Germany: Springer, 2012, pp. 62–77.

[3] F. E. Harrell Jr, K. L. Lee, D. B. Matchar, and T. A. Reichert, "Regression models for prognostic prediction: advantages, problems, and suggested solutions," Cancer Treat. Rep., vol. 69, no. 10, pp. 1071–1077, 1985

[4] P. Lapuerta, S. P. Azen, and L. Labree, "Use of neural networks in predicting the risk of coronary artery disease," Comput. Biomed. Res., vol. 28, no. 1, pp. 38–52, Feb. 1995.

[5] K. M. Anderson, P. M. Odell, P. W. Wilson, and W. B.Kannel, "Cardiovascular disease risk profiles," Amer. heart J., vol. 121, no. 1, pp. 293–298, 1991.

[6] H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," Procedia Comput. Sci., vol. 83, pp. 1064–1069, Jan. 2016.

[7] F. Petropoulos and S. Makridakis, "Forecasting the novel coronavirus COVID-19," PLoS ONE, vol. 15, no. 3, Mar. 2020, Art. No. e0231236.

[8] G. Grasselli, A. Pesenti, and M. Cecconi, "Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: Early experience and forecast during an emergency response," JAMA, vol. 323, no. 16, p. 1545, Apr. 2020.

[9] WHO. Naming the Coronavirus Disease (Covid-19) and the Virus That Causes it. Accessed: Apr. 1, 2020. [Online]. Available: https:// www.who.int/emergencies/diseases/novel-coronavirus-2019/technic% alguidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virusthat-cau%ses-it

[10] C. P. E. R. E. Novel, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (Covid-19) in China," Zhonghua Liu Xing Bing Xue Za Zhi= Zhonghua Liuxingbingxue Zazhi, vol. 41, no. 2, p. 145, 2020.

[11] L. van der Hoek, K. Pyrc, M. F. Jebbink, W. Vermeulen-Oost, R. J. Berkhout, K. C. Wolthers, P. M. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren, and B. Berkhout, ‘‘Identification of a new human Coronavirus,’’ Nature Med., vol. 10, no. 4, pp. 368–373, 2004.

[12] Johns Hopkins University Data Repository. Cssegisanddata. Accessed: March 27, 2020. [Online]. Available: https://github.com/ CSSEGISandData

[13] M. R. M. Talabis, R. McPherson, I. Miyamoto, J. L. Martin, and D. Kaye, "Analytics defined," in Information Security Analytics, M. R. M. Talabis, R. McPherson, I. Miyamoto, J. L. Martin, and D. Kaye, Eds. Boston, MA, USA: Syngress, 2015, pp. 1–12. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128002070000010

[14] H.-L. Hwa, W.-H. Kuo, L.-Y. Chang, M.-Y. Wang, T.-H. Tung, K.-J. Chang, and F.-J. Hsieh, "Prediction of breast cancer and lymph node metastatic status with tumor markers using logistic regression models," J. Eval. Clin. Pract., vol. 14, no. 2, pp. 275–280, Apr. 2008.

[15] R. Tibshirani, "Regression shrinkage and selection via the lasso," J. Roy. Stat. Soc., Ser. B, Methodol., vol. 58, no. 1, pp. 267–288, Jan. 1996.

[16] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 55–67, Feb. 1970.

[17] X. F. Du, S. C. H. Leung, J. L. Zhang, and K. K. Lai, "Demand forecasting of perishable farm products using support vector machine," Int. J. Syst. Sci., vol. 44, no. 3, pp. 556–567, Mar. 2013.

[18] F. Rustam, I. Ashraf, A. Mehmood, S. Ullah, and G. Choi, "Tweets classification on the base of sentiments for US airline companies," Entropy, vol. 21, no. 11, p. 1078, Nov. 2019.

[19] E. Cadenas, O. A. Jaramillo, and W. Rivera, "Analysis and forecasting of wind velocity in Chetumal, Quintana roo, using the single exponential smoothing method," Renew. Energy, vol. 35, no. 5, pp. 925–930, May 2010.

[20] J. Lupón, H. K. Gaggin, M. de Antonio, M. Domingo, A. Galán, E. Zamora, J. Vila, J. Peñafiel, A. Urrutia, E. Ferrer, N. Vallejo, J. L. Januzzi, and A. Bayes-Genis, ‘‘Biomarker-assist score for reverse remodeling prediction in heart failure: The ST2-R2 score,’’ Int. J. Cardiol., vol. 184, pp. 337–343, Apr. 2015.

[21] J.-H. Han and S.-Y. Chi, "Consideration of manufacturing data to apply machine learning methods for predictive manufacturing," in Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN), Jul. 2016, pp. 109–113.

[22] C. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Climate Res., vol. 30, no. 1, pp. 79–82, 2005.

[23] R. Kaundal, A. S. Kapoor, and G. P. Raghava, "Machine learning techniques in disease forecasting: A case study on rice blast prediction," BMC Bioinf., vol. 7, no. 1, p. 485, 2006.

[24] S. Baran and D. Nemoda, "Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting," Environmetrics, vol. 27, no. 5, pp. 280–292, Aug. 2016.

[25] Y. Grushka-Cockayne and V. R. R. Jose, "Combining prediction intervals in the m4 competition," Int. J. Forecasting, vol. 36, no. 1, pp. 178–185, Jan. 2020.

[26] N. C. Mediaite. Harvard Professor Sounds Alarmed on 'Likely' Coronavirus Pandemic: 40% to 70% of world Could be Infected This Year. Accessed: February 18, 2020. [Online]. Available: https://www.mediaite.com/news/ Harvard-professor-sounds-alarm-on-likely-%coronavirus-pandemic-40- to-70-of-world-could-be-infected-this-year

[27] BBC. Coronavirus: Up to 70% of Germany Could Become Infected— Merkel. Accessed: Mar. 15, 2020. [Online]. Available: https://www.bbc. com/news/world-us-canada-51835856