State-wise prevalence of COVID 19 in India by using machine learning approaches

Article · September 2020

DOI: 10.31838/ijpr/2020.SP2.295

CITATIONS
3

READS
349

2 authors, including:

Vijayalakshmi Kakulapati
Sreenidhi Institute of Science & Technology
114 PUBLICATIONS **95** CITATIONS

SEE PROFILE

ISSN 0975-2366
DOI: https://doi.org/10.31838/ijpr/2020.SP2.295

Research Article

State-wise prevalence of COVID 19 in India by using machine learning approaches

B. SREE KEERTHI MEGHANA[1], V. KAKULAPATI[2]
[1]Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, Telangana-500075,
[2]Sreenidhi Institute of Science and Technology, Yamnampet, Ghatkesar, Hyderabad, Telangana-501301.
Email ID: meghanasharma224@gmail.com, vldms@yahoocom
Received: 30.07.20, Revised: 29.08.20, Accepted: 26.09.20

ABSTRACT
Coronavirus is one of the world's most critical issues, till date. Comprehension of causative variables such as mellitus, heart-related issues, asthma, blood pressure, etc., including the intrinsic transmission mechanisms of the disease, COVID 19 and its eradication are important for neurological investigation. Hence, the advance of appropriate modeling approaches and methods applied to current corona information on the pervasiveness of the pandemic and other serious illness aspects, is taking consideration. The prevalence of COVID 19 in India has reached epidemic proportions, and this disease is becoming a significantly increasing case in India. In this work, polynomial regression analysis methods employ to to forecast the number of COVID 19 corona patients. In this, we described a decision tree, polynomial and random forest classification of disease in COVID 19 incidences modelling and forecasting in India and a predicted prevalence of high level of confidence.

Keywords: polynomial, decision, prevalence, model, COVID 19, classification

INTRODUCTION

The COVID-19, the infectious disease caused by coronavirus, is currently a concern in the world. Before the outbreak in Wuhan, China, in December 2019, this new viruses and disease were unknown. Healthcare providers and WHO (world health organization) have declared a considerable number of cases in China. Gradually, the cases are reporting in several countries, frequently spread by individuals going outside of China. The pandemic doesn't have any clinical diagnosis and vaccine or diagnosis to date. The treatment and analysis of COVID-19 patients are not predefined. With the collaboration of different pharma companies in Indian states are working with the support of the central Government.

In India, coronavirus outbreak, also named as COVID-19 or SARS-CoV-2, occurred in the state of Kerala in late January 2020 where three students travelled from Wuhan town to the virus, and the positive case of coronavirus was registered. At the same time, several other cases were detected in other parts of the country, most of which were linked to people with a travel history to affected countries. Infections increased quickly since March, with significant growth in testing. In June 2020, India's casualty rate for positive COVID-19 cases [1] remained at about 2.8 percent. Nonetheless, as in India's different states, this does exclude unreported diseases or deferrals from ailment to death. However, it is important to evaluate and direct the interest in a health setting in all reported cases and conceivable new cases. Numeric and observable demonstration systems to carry out short- and prolonged case tests to evaluate the additional materials and properties that the flare-up is required to address. Average disease weight assessment is important to ensure sustainable and convenient clinical consideration and different assets to defeat the pandemic in general welfare authorities. In addition, the powers and intercessions required to lighten the flaring up will coordinate this gage [2].

Currently, in India COVID-19 outbreak is on the pinnacle and spread exponentially, because of that the financial, well-being administrations, business, little ventures are enduring, and all individual's life is in stakes. In this situation, government officials, health workers, and doctors are working day and night to deal with the circumstances. The police and physicians are giving maximum effort to save the public. Investigators and scientists are busy working in the direction of medicine and vaccine development to help healthcare providers: physicians, nurses, professional, and paramedic staff. In this work, we analyze COVID 19 cases, such as confirmed, recovered, and fatality analysis of state-wise and predicting the pandemic outbreak's prevalence. For implementation purposes, polynomial regression, decision tree, and random forest classification are analyzed.

RELATED WORK

Nowadays, machine learning models demonstrated that efficient predictions and accurate consequences for different diseases in

health care. A systematic review of approaches [3] provided by used to forecast the dynamics of influenza pandemic. They have reviewed research papers based on determine frame action models, regression, prediction rules, Bayesian network, ARIMA forecasting model etc. Current COVID-19 surveys mainly include a review of the existing inadequate data [4-6].

Even though, a way to understand the insight and analysis on the dynamics of COVID-19 spread, several recent studies on the Indian context [7] have also been published in the context of State perception examinations [8-10]. Concerning population size, population density, and geographical conditions, India's fundamental analysis does not offer an actual epidemic status. Thus, each of the states with enormous populations compared with other parts of the world must examine the spread of coronavirus separately.

In [11], a tool that could be useful in predicting the COVID-2019 region, conducting linear regression and expectation of COVID-19 Kaggle data in the forecasting of the epidemic trends of the COVID-2019 in India was recommended by a multilayer perceptron and vector autoregression technique. A significant learning based automated revelation [12] and portraying the portrayal model for fundus DR pictures. Assessment of the outbreak of COVID-19 and its difficulties and dependent on their outcomes made a few proposals to counter more episodes of the virus [13]. The role of inanimate surfaces in coronavirus outbreak was exploring in another study. On the basis of the result he made several recommendations for preventing further viral spread in the impact of surface disinfection [14]. The findings have shown that dynamic computational modeling is essential to predict virus behavior [15], and an overview of the actions of viruses is studied and measured the role of media coverage in the country. Their findings showed that inaccurate and partial media coverage could adversely affect individuals' mental health [16].

Predict the number of confirmed cases of coronavirus and analysed the trend of transmission of the virus from one another and recovery rates and introduced a numerical model. The results showed a useful presentation of the proposed model in anticipation of the cases confirmed [17]. Another research assessed China's trends in the trends of the COVID-19 case, which shows that rapid and dynamic methods can help reduce and force the emergency [18]. In order to minimise COVID-19 sequence, the effect of movement imperatives is estimating. The results showed that limitations of movement are incredibly useful in reducing the virus spreading [19].

PROPOSED METHOD

This study collects data from the official website from COVID-19 from affected Indians. The dataset includes the number of reported cases, active cases, retrieved cases, and death cases between 30 January and 4 August. There were overall 4846 records of eight independent variables. The creation and evaluation of various model regressions datasets are dividing into 70 percent for the train and 30 percent for the test set. Construct a better model with an accurate prediction, and pre-processing is needed. A classification algorithm uses organised mathematical and statistical principles to capture patterns in training data. A predictive model based on training data is developed from the algorithm to predict a class label in test data.

The objectives of this work

- Predicting of spreading coronavirus across Indian states.
- Analysis of coronavirus cases' growth rates and the types of mitigation across the country is calculating.

Polynomial regression: It is a type of linear regression that shows an nth degree polynomial relation between the independent variable y and x. A non-linear relation between the x estimate and the related mean of y for the containment is the polynomial regression, indicated by E(y).

Decision tree: It is a support tool for decision making that uses a tree-like decision model, including chance events,

resource costs and usefulness. One way is to demonstrate an algorithm that includes only conditions of control.

Random forest: It is also calling as random decisions that are ensemble methods for categorization, regression analysis, and multiple forests that grow a large number of trees to schedule a time for the classes or the goals of the students. Random forests are essential methods of learning for groups.

IMPLEMENTATION

COVID-19 data from the official website containing 4846 records of the affected Indians containing eight independent variables. This data set is 70% for the collection of trains and 30 percent for the set of tests as an input for developing and evaluating different regression models. Polynomial regression has been finding to have a low error rate and the highest R2 value both for training and for defining predictions. The graph also demonstrates how well a model predicts the number of input deaths for the test set; for example, the active cases recovered.

Table 1: Example of daily cases

S.No	Date	State/UT	Confirmed Indian	Confirmed Foreign
1	30-01-20	Kerala	1	0
2	31-01-20	Kerala	1	0
3	01-02-20	Kerala	2	0
4	02-02-20	Kerala	3	0
5	03-02-20	Kerala	3	0

4846 rows × 9 columns

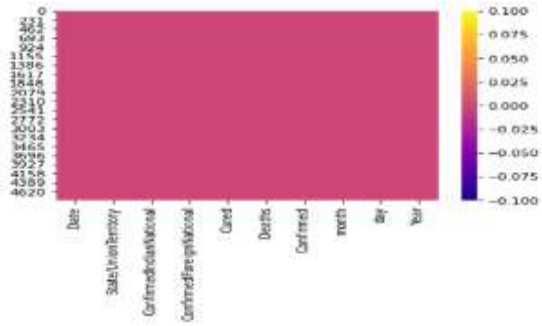

Fig.1: Day wise covid 19 cases (confirmed, recovered, and fatality)

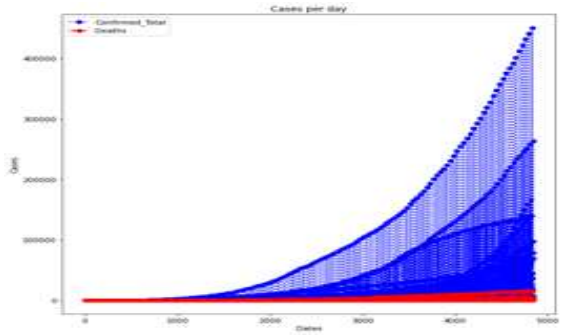

Fig.2: The increase of corona cases (confirmed, fatality and recovered)

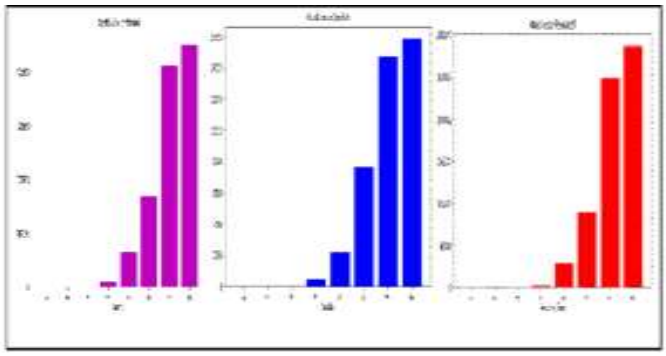

Fig.3: Day wise analysis of corona cases

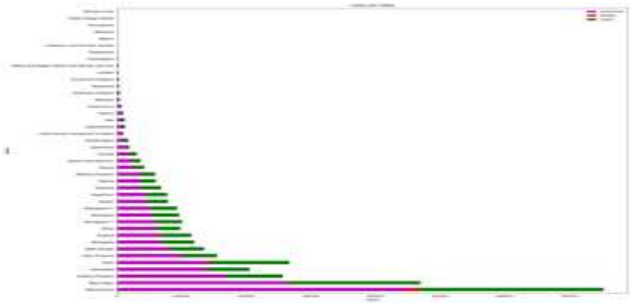

Fig.4: State wise corona case in India

Table 2: Daily recovered, death, and confirmed cases

	Date	Sno	Cured	Deaths	Confirmed
0	01/02/20	3	0	0	2
1	01/03/20	32	0	0	3
2	01/04/20	15776	144	41	1834
3	01/05/20	47824	9065	1152	34972
4	01/06/20	91458	91819	5394	190535
...	...	...	...	...	...
183	30/07/20	162890	1020582	34968	1583792
184	31/01/20	2	0	0	1
185	31/03/20	14434	124	35	1397
186	31/05/20	90162	86984	5164	182143
187	31/07/20	164115	1057805	35747	1638870

188 rows × 5 columns

1 60
2 90
3 120
4 150
5 180
6 210

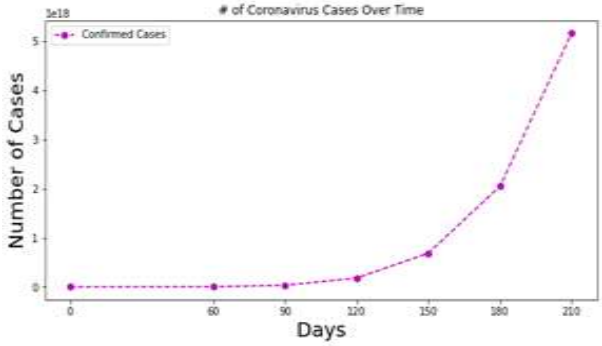

Fig.5: The percentage of Coronavirus cases over time by using polynomial regression

Table 3: Coronavirus cases month and daily wise

	Confirmed	month	day
0	2	1	2
1	3	1	3
2	1834	1	4
3	34972	1	5
4	190535	1	6
...	...	...	...
183	1583792	7	30
184	1	1	31
185	1397	3	31
186	182143	5	31
187	1638870	7	31

188 rows × 3 columns

Random forest accuracy: 0.9853963265185706
Decision tree accuracy: 0.9978774578934049

CONCLUSION

In this study, we analyse the COVID 19 cases state-wise across India. And apply the polynomial regression, decision tree, and random forest classification to predict the coronavirus's growth rate in India. Moreover, because of the trend, the number of cases would undoubtedly increase. The doctors, health professionals, and those who offer vital care must safeguard in compliance with medical requirements prescribed. Owing to the

carelessness of individuals and organisations, the number of cases will rise exponentially in future. The peak is still to reach; therefore, the Government must be extra cautious and introduce strict measures. Also, medical services must be actively improving across the country.

FUTURE ENHANCEMENT

In the future, we intend to explore the methodology of prediction by using the dataset modified and apply the most precise and effective ML prediction methods. A pipeline will develop to integrate CXR-scan machine view models with these kinds of data processing models in demographics and healthcare. These models are then incorporating into applications that support mobile healthcare development. It may be a step towards a semi-autonomous mechanism for diagnosing, monitoring and preparing possible outbreaks quickly in COVID-19 affected regions.

REFERENCES

1. https://www.statista.com/topics/6135/coronaviruses-covid-19-outbreak-in-india/.
2. S Zhang et al., 2020 International Journal of Infectious Diseases "Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: a data-driven analysis", pp. 201-204,
3. Nsoesie, et.al., 2014. A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza and other respiratory viruses, 8(3), pp.309-316.
4. Pirouz, B., et.al., 2020. Investigating a Serious Challenge in the Sustainable Development Process: Analysis of Confirmed cases of COVID-19 (New Type of Coronavirus) Through a Binary Classification Using Artificial Intelligence and Regression Analysis. Sustainability, 12(6), p.2427.
5. More, et.al., 2020. A serological survey of canine respiratory coronavirus in New Zealand. New Zealand Veterinary Journal, 68(1), pp.54-59.
6. Wu, C., et.al., 2020. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA internal medicine.
7. P. Kumar et. al., Forecasting COVID-19 impact in India using pandemic waves Nonlinear Growth Models, medRxiv 2020: https://doi.org/10.1101/2020.03.30.20047803
8. P. Ghosh et. al., COVID-19 in India: State-wise Analysis and Prediction, medRxiv 2020: https://doi.org/10.1101/2020.04.24.20077792
9. N. Poonia and S. Azad, Short-term forecasts of COVID-19 spread across Indian states until 1 May 2020, https://arxiv.org/abs/2004.13538
10. S. Mondal et. al., Possibilities of exponential or Sigmoid growth of Covid19 data in different states of India, medRxiv 2020: https://doi.org/10.1101/2020.04.10.20060442.
11. Sujatha R, et.al., A machine learning forecasting model for COVID-19 pandemic in India. Stoch Environ Res Risk Assess. (2020) 34:959–72. doi: 10.1007/s00477-020-01827-8.
12. Sakarkar G, et.al., Comparative study of ambient air quality prediction system using machine learning to predict air quality in smart city. In Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), NITTTR Chandigarh, India. Singapore: Springer (2020). p. 175–82.
13. Lai, C, et.a., Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int. J. Antimicrob. Agents 2020, 55, 105924.
14. Kampf, G. Potential role of inanimate surfaces for the spread of coronaviruses and their inactivation with disinfectant agents. Infect. Prev. Pract. 2020, 2, 100044.
15. Telles, C.R. COVID-19, an overview about the epidemic virus behavior. MediArXiv 2020.
16. Wen, J, et.al., Effects of misleading media coverage on public health crisis: A case of the 2019 novel coronavirus outbreak in China. Anatolia 2020.
17. Chen Y.C, et.al.,. A time-dependent SIR mode for COVID-19. arXiv 2020, arXiv:2003.00122.
18. Li, Q, et.al., Trend and forecasting of the COVID-19 outbreak in China. arXiv 2020, arXiv:2002.05866.
19. Chinazzi, M, et.al.,. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020.