Time Series Analysis Of COVID-19 Occurrence In Different States Of India: A Periodic Regression Analysis

Kuralayanapalya Puttahonnappa Suresh[1], Chandan Dharmashekar[2], Chandan Shivamallu[2], Chaitra N[3], Mallikarjun S Beelagi[2], Preethi R Bhat[3], Darshan Hebbuluse Veerabhadrappa[1], Pooja Anudhar G[4], Snehalatha Basavaraju[5], Santosh Kumar SR[6], Shiva Prasad Kollur[7], Ashwini Prasad[5], Chandrashekar Srinivasa[8], Sharanagouda S Patil[1,*]

[1]ICAR-National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), Yelahanka, Bengaluru-560064, India.
[2]Department of Biotechnology and Bioinformatics, Faculty of Life Sciences, JSS Academy of Higher Education & Research, Mysuru-570015, India.
[3]Division of Medical Statistics, Faculty of Life Sciences, JSS Academy of Higher Education & Research, Mysuru-570015, India.
[4]Department of Nutrition and Dietetics, Faculty of Life Sciences, JSS Academy of Higher Education & Research, Mysuru-570015, India.
[5]Department of Microbiology, Faculty of Life Sciences, JSS Academy of Higher Education & Research, Mysuru-570015, India.
[6]Department of Food Technology, Davangere University, Shivagangotri, Davangere –577007, Karnataka, India.
[7]Department of Sciences, Amrita School of Arts and Sciences, Mysuru, Amrita Vishwa Vidyapeetham, Karnataka – 570 026, India;
[8]Department of Studies in Biotechnology, Davangere University, Shivagangotri, Davangere Karnataka-577 007, India.

***Correspondence:** Dr. Sharanagouda S Patil, Principal Scientist, ICAR-National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), PBNo-6450, Yelahanka, Bengaluru-560064
Email: chandans@jssuni.edu.in (CS); sharanspin13@gmail.com

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

Abstract- COVID-19 is the deadliest pandemic, with over 18.2 million people infected with the SARS-CoV-2 virus by August 2, 2021 resulting in human deaths and economic losses. A number of countries have formulated control measures in order to prevent the spread of the virus. However, it is unknown when the outbreak will subside in different countries around the world. The role of predicting the COVID-19 trend is extremely difficult. Indian government has made disease outbreak analysis a priority in order to implement necessary healthcare measures to reduce the impact of this deadly pandemic on human health and country's economics. The time series data for COVID-19 disease was collected from the website www.covid19india.org and were analyzed using a periodic regression model using the data from 22nd Janaury March 2020 to 01st Febraury 2021 the estimated number of cases until 27 July, 2021 was predicted to develop a stochastic model using periodic regression and were documented in top 10 highly infected states in India. The analysis revealed a increasing pattern for the number of reporting cases in the early days of prediction and decreasing trend for the number of reporting cases in the later days of prediction, which could decrease in future days in Karnataka, West Bengal, Uttar Pradesh, Telangana, Bihar and Haryana states. However, in Madhya Pradesh, Andhra Pradesh, Maharashtra and Tamil Nadu states showed a rapid phase of rise in disease incidence, which is likely to infect a larger population and suggests the disease's pandemic existence over a duration. Our model emphasizes the importance of ongoing and continuous efforts that are in place in all states to minimize occurrence of new cases of infections, so as to potentially improving India's economic wealth with the available resources.

Keywords: COVID-19, Time series, Outbreak, Periodic regression, Disease.

Introduction

According to WHO reports, COVID-19 disease has been reported in over 210 countries worldwided (Sexton et al., 2016). As a result of its rapid spread through many countries, the international committee on taxonomy of viruses (ICTV), are responsible for official classification of viruses and viral taxonomy, later called the causative virus as SARS-CoV-2 (Liu et al., 2020). COVID -19 was first found in bats and then affect human health. It can also be

transmitted from person to person. SARS and MARS, two other pandemics similar to COVID, first reported in Guangdong, China and then in Middle East (Acter et al., 2020). In 2019, SARS-CoV-2 was discovered in Wuhan. It has spread to over 200 countries since its current outbreak. As of July 31, 2020, there are about 34,968 death cases and 15,83,792 were infected across India since the outbreak initiated (Raji et al., 2020). Machine learning is a technique that allows computers to learn without having to be programmed. There are majorly two forms of machine learning: supervised learning and unsupervised learning. The mapped by analyzing the training data's input output relationship is an example for supervised learning (Yang et al., 2021). Covid-19 outbreak is predicted using auto regression method, SEIR model, and a seasonal periodic regression model (Mwalil et al., 2020). Dew to the pandemic of virus outbreak, scientists across the world started conducting research on virus spreading. Lin et al. (2020) proposed Susceptible– Exposed–Infectious–Removed (SEIR) model to study the spread of the infection in China. The SIRD Model is used to calculates epidemiological parameters such as reproduction number, recovery rate, mortality rates and infection. Gated Recurrent Unit and short-term memory are using to regular neural network which are also used to predict restored, negative and death rate (Arun kumar et al., 2021). More than 210 countries and territories have been affected by the epidemic, with the United States accounting for about one-fifth of all global outbreaks. In late January 2020, India experienced an outbreak of the virus, as when three Indian students travelled to Kerala from Wuhan, China, the epicenter of the outbreak (Vaman et al., 2020). COVID-19 was found in all three of them, indicating a local outbreak. Several other cases were discovered in other parts of the world at the same time, the majority of which were attributed to people who had travelled to the affected countries previously. Since March, the number of infections has risen dramatically, along with a major increase in research. Kerala was praised for acting quickly to prevent the virus from spreading further (Chathukulam et al., 2021). Thousands of people were routinely put in quarantine at home or in institutions, where they were monitored for signs and infections. Despite ramping up in recent months, India had one of the lowest testing rates for the virus compared to other countries (Shams et al., 2020).

In India, livestock disease outbreaks have a negative impact on the economy of animal husbandry farmers. Analysis of recent disease outbreaks will aid in future disease prevention preparation and successful preventive measures (Cirrincione et al., 2020). The livestock diseases can occur in a seasonal and cyclical pattern in animals; if this information is obtained through statistical analysis methods, it will aid in the efficient use of resources in organizing the preventive steps, using all available tools. To determine the pattern and potential prediction of outbreaks, periodic regression is used (Thyagaraju et al., 2020). Several types of disease data exhibit periodic/cyclic characteristics and appear to fluctuate at frequent time intervals. Periodic regression curve corresponds to few variables in time and is frequently repeated at fixed time intervals for predicting the spreading of COVID-19 disease (Chaurasia et al., 2020). Disease modeling and incidence analysis can aid in forecasting disease probability and controlling early disease preparedness through effective control measures. For predicting the spread of COVID-19 disease, a periodic regression curve corresponds to a few variables in time and is often performed at fixed time intervals (Yan et al., 2008). Since there is no well-test and effective viral vaccines developed against this deadly virus, Thus, the present study provides a key role in controlling the pandemic and flattening the disease curve. It will contribute to a larger picture of aggressive and timely controlling measures in infrastructure, service facilities, infectious vaccinations, and effectively controlling related epidemics in future (Krishnamoorthy et al., 2019).

Materials and Methods

A curve relating certain variable to time as well as repeating at fixed intervals of time is a periodic type. This model type has the ability of representing time series using a minute number of parameters which is highly significant, specifically when time series is not monotonic and stationary and includes a non-linear trend, cyclical and seasonal components having distinct periodicity. In the periodicity analysis, the definition of the three principle parameters is if of great importance: cycle length or fundamental period length; its amplitude or the range from the minimum to the maximum response and the phase angle or angular point in time during the period when the response is maximum. Estimation of these parameters is easy by employing any statistical software (Bliss et al., 1970).

A time series Y_t (t=1… N) detected at equal time intervals can be represented as

$Y_t = \hat{Y}_t + \varepsilon_t$, $\hat{Y}_t$ is definite unobserved value at time t. and $\{\varepsilon_t\}$ is a random error sequence with identical and independent distribution having 0 as the mean and σ^2. For determining if the time series variability contains periodic components, approximation of the series is done by the finite Fourier series having the form, if number of data is even: N = 2n

$$\hat{Y}_t = A_0 + 2\sum_{m=1}^{n-1}(A_m cos2\pi m f_1 t + B_m sin2\pi m f_1 t) + A_n cos2\pi n f_1 t$$

or

$$\hat{Y}_t = A_0 + 2\sum_{m=1}^{n-1}(A_m cos2\pi m f_1 t + B_m sin2\pi m f_1 t)$$

If the number of data is odd: N = 2n-1

Here, $R_m = \sqrt{A_m^2 + B_m^2}$ is the amplitude, $\emptyset_m = arctg(\frac{B_m}{A_m})$ is the ith component phase. The function $\hat{Y}_t$ is a linear combination of the functions sinus and cosinus with frequencies being proportional to the fundamental frequencies f_1=1/N, so it is a linear multiple regression where the functions sinus and cosinus are the regressors.
Further, to test the goodness of fit, performance of F-test and t-test can be useful.

Results and Discussion

The major 10 states contributing for more than 50% of the total Indian reported covid-19 cases were included in the comprehensive study of periodic regression analysis for the prediction of covid-19 incidences that may probably occur in the upcoming period of six months depicted in Table 1. Highly significant values ($p<0.001$) corresponding to the values of intercept, x, sinus and cosinus were observed for all the Indian states (See Table 1). The analysis of covid-19 disease dataset aided in showcasing the upcoming trends of coronavirus infection for a six-month interval (up to 27 July 2021) in the respective states. The curve of periodic regression analysis exhibited the outbreaks' baseline, upper bound line that appears at 95% confidence interval from the baseline, also the observed line indicating the definite infections that have occurred in the duration chosen for the study, which thereby helps in the estimation of probable cases that might arise in the next six months duration in the 10 Indian states.

The curves representing the current and future COVID-19 infections in the highly effected 10 major Indian states can be observed in the Figure 1. From the trend analysis of Kerala, the observed outbreaks indicate a gradual increase in the disease infections along with the tendency of a rapid rise of infections the future crossing the upper bound line. This is a clear indication of the disease approaching a critical stage in the upcoming days if no necessary and strict control measures are adopted. From the trend analysis of Delhi, we can observe an increasing and decreasing pattern over a time period for the outbreaks, implying that COVID-19 incidences have a tendency to follow beneath the upper bound line and the number of incidences may decrease in the future with the strict and effective implementation of control measures.

The trend analysis of West Bengal and Tamil Nadu disclosed that although there were rise in the observed incidences till late October and mid of July respectively, with the cases not traversing the upper bound line, a drastic reduction in corona infections were seen till mid of August whereas in case of states like Uttar Pradesh and Rajasthan even though the outbreaks were found to traverse beyond the upper bound line during the mid-September and late December respectively, a dramatic reduction was observed later on. Also, it is evident that COVID incidences fall below the baseline for West Bengal as well as Rajasthan from the graphs. In states like Andhra Pradesh, Maharashtra, Karnataka and Odisha, the line depicting the observed outbreaks showed the tendency of following the baseline, with the outbreaks of Andhra Pradesh nearly approaching 0. Thus, in these states, less severity of infections was expected in the upcoming days. The primary reason behind the reduction of reported coronavirus incidences in these states might be the population adaptation in accordance with the rules imposed by the Indian government for the effective mitigation of the disease. However, negligence of the population and the relaxation of rules imposed regarding the population and travellers movements from distinct states may increase the infection severity in the impending days. In this study we have utilized R-software version 3.6.3 CRAN (Comprehensive R Archive Network) for computing impending trend of COVID-19 incidences for the next six months by incorporating periodic regression model for obtaining regression curves and predicted the future disease outbreaks for the period from 2020-06-22 to 2021-07-27.

Table 1. Periodic regression analysis values of COVID-19 disease outbreaks in the top 10 highly infected states in India

States	Parameter	Estimate	Standard deviation			R^2	Adj.R^2
Maharashtra	Intercept	4183.935	253.545	16.502	<0.001**	0.68025	0.67691
	x	13.485	1.341	10.056	< 0.001**		
	c1	1191.839	189.861	6.277	< 0.001**		
	s1	-4813.72	208.355	-23.103	<0.001**		
Karnataka	Intercept	1889.438	139.551	13.539	< 0.001**	0.78252	0.78024

	x	8.573	0.743	11.544	< 0.001**		
	c1	1602.066	105.239	15.223	<0.001**		
	s1	-3248.76	111.725	-29.078	<0.001**		
Kerala	Intercept	-1517.06	78.187	-19.403	<0.001**	0.9199	0.91907
	x	24.668	0.432	57.123	<0.001**		
	c1	1045.139	61.502	16.994	<0.001**		
	s1	338.944	57.721	5.872	<0.001**		
Andhra Pradesh	Intercept	2294.367	158.907	14.438	<0.001**	0.75981	0.7573
	x	5.358	0.805	6.657	<0.001**		
	c1	1379.581	107.431	12.842	<0.001**		
	s1	-3682.45	134.41	-27.397	< 0.001**		
Tamil Nadu	Intercept	2500.286	128.366	19.478	< 0.001**	0.71203	0.70902
	x	3.622	0.643	5.631	<0.001**		
	c1	478.524	85.573	5.592	<0.001**		
	s1	-2713.76	107.111	-25.336	<0.001**		
Delhi	Intercept	884.812	130.152	6.798	< 0.001**	0.24368	0.23578
	x	4.635	0.724	6.398	<0.001**		
	c1	509.35	94.671	5.38	< 0.001**		
	s1	-495.198	98.886	-5.008	< 0.001**		
Uttar Pradesh	Intercept	1036.736	92.49	11.209	< 0.001**	0.67262	0.6692
	x	5.815	0.48	12.125	< 0.001**		
	c1	764.012	65.815	11.608	< 0.001**		
	s1	-1594.18	78.278	-20.366	< 0.001**		
West Bengal	Intercept	445.383	90.681	4.912	< 0.001**	0.66088	0.65734
	x	8.494	0.505	16.815	< 0.001**		
	c1	869.97	69.902	12.446	< 0.001**		
	s1	-859.362	67.008	-12.825	< 0.001**		
Odisha	Intercept	527.072	51.861	10.163	< 0.001**	0.741	0.73829
	x	3.504	0.274	12.784	<0.001**		
	c1	-989.65	42.586	-23.239	<0.001**		
	s1	-989.65	42.586	-23.239	<0.001**		
Rajasthan	Intercept	229.937	46.826	4.91	<0.001**	0.65836	0.65479
	x	4.466	0.259	17.248	<0.001**		
	c1	519.282	34.423	15.085	< 0001**		
	s1	-349.681	35.429	-9.87	< 0.001**		

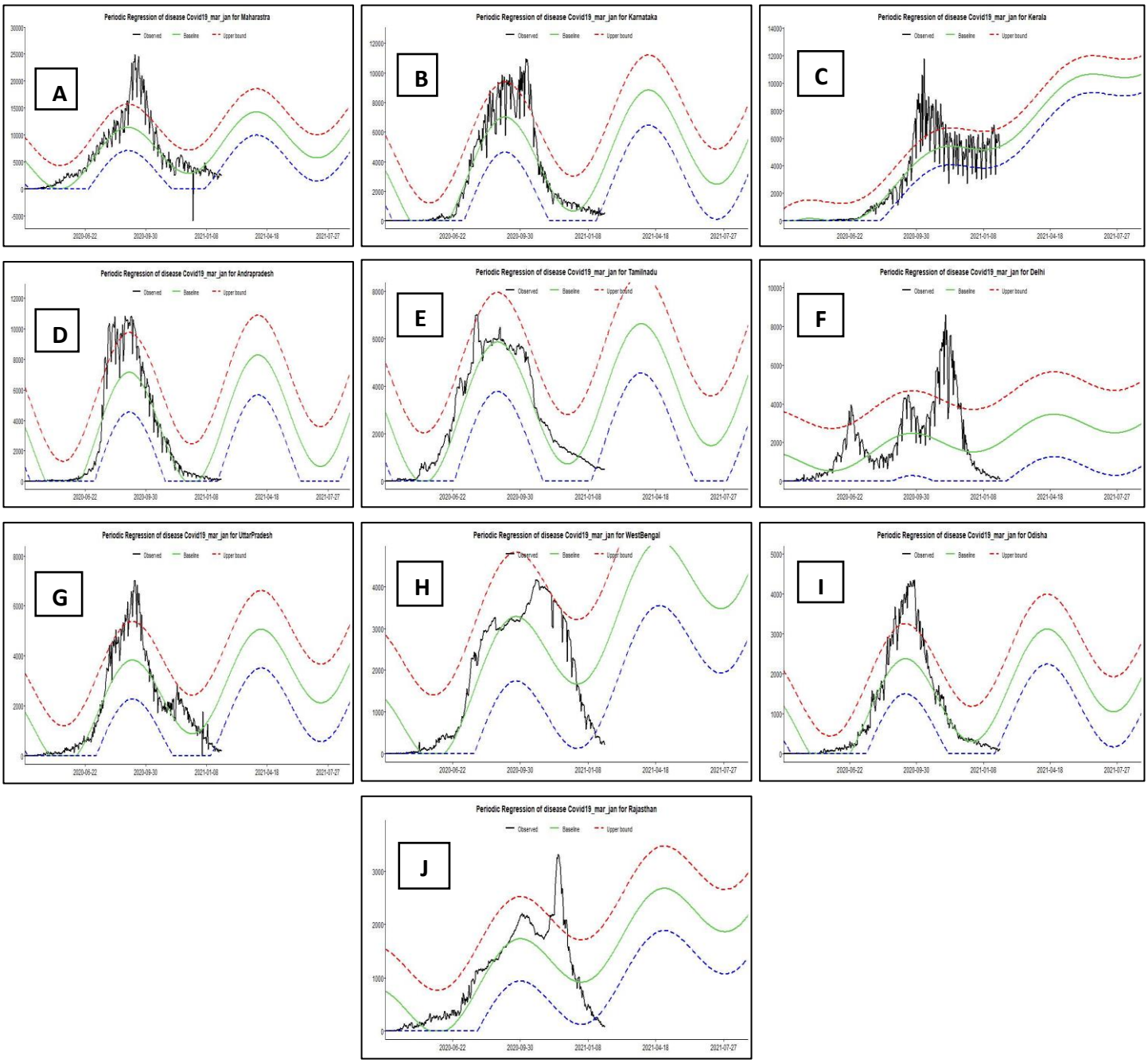

Figure 1. Periodic regression analysis of COVID-19 outbreaks in the top 10 highly infected states in India-Maharashtra (A), Karnataka (B), Kerala (C), Andhra Pradesh (D), Tamil Nadu (E), Delhi (F), Uttar Pradesh (G), West Bengal (H), Odisha (I) and Rajasthan (J)

Our current situation requires disease pattern prediction to follow in order to receive effective healthcare treatments. In future, we will be able to monitor the disease. The analysis predicts terrifying outcomes in India, especially in Karnataka, West Bengal, Madhya Pradesh, Uttar Pradesh, Andhra Pradesh, Telangana, Bihar, Haryana, Tamil Nadu and Maharashtra. Based on the results of our research, public health officials should adapt their proactive preparation and strategies to implement aggressive viral infection control strategies at hospital and community levels to limit the COVID-19 pandemic (Ghosh et al., 2020). These findings were close to those of a study that looked at the global level and collected data for the top 10 countries using Machine Learning approach and artificial Intelligence techniques (Ye

et al., 2003). In our research, we used a periodic regression model to forecast disease patterns for top 15 most infected states in India. Fortunately, India implemented stringent protection measures such as enforced quarantine, lockdown, curfews, and travel restrictions to limit the spread of infection during the early stages of a pandemic. But, as the lockdown is relaxed, the populace begins to mix without sufficient consideration, potentially resulting in a rapid increase in infectious transmission and disease spread. Low vaccination coverage, a lack of financial support, and a lack of medical infrastructure are all major barriers to controlling COVID-19 disease infection in developing countries like India, delaying the development of herd immunity (Mai et al., 2016). If immunity lasts longer than the disease outbreak, epidemic dynamics will not be disrupted by decreasing immunity (Omer et al., 2020) . Even if the government has implemented tight restrictions steps, the current trend indicates that there will be a geometric progression in the coming days, and there are certainly more chances of falling into exponential cases in the long run, according to the effects of periodic regression. The Government of India, on the other hand, hopes to reduce the disease curve with its new planning and strategies. Since the peak patterns to vary at regular time intervals and also shows the characteristic of periodic existence, the modified data at intervals of fifteen days must be used to carry out the study.

Conclusion

In this study, the statistical approach, periodic regression analysis has shown that both the increase and decrease of COVID-19 outbreak among the top 10 states of India in the upcoming days of July 2021. The trend analysis predicts that increase and tendency of rapid infection in Kerala. Whereas in West Bengal, Tamil Nadu, Uttar Pradesh, and Rajsthan the possibility of drastic reduction could be seen. In Delhi, the number of infection may act different compared to other states. Infection can play by increasing and decreasing due to the consistency of varience in the data. We believe that, this prediction of COVID-19 outbreaks will aid the future research community to understand the survivallence of the virus, and helps to the responsible citizens of nation to be more precautions.

Reference

1. Acter, T., Uddin, N., Das, J., Akhter, A., Choudhury, T. R., & Kim, S. (2020). Evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as coronavirus disease 2019 (COVID-19) pandemic: A global health emergency. *Science of the Total Environment*, 138996.
2. ArunKumar, K. E., Kalaga, D. V., Kumar, C. M. S., Kawaji, M., & Brenza, T. M. (2021). Forecasting of covid-19 using deep layer recurrent neural networks (rnns) with gated recurrent units (grus) and long short-term memory (lstm) cells. *Chaos, Solitons & Fractals*, *146*, 110861.
3. Bliss, C. I. (1970). Statistics in biology. Vol. 2. *Statistics in biology. Vol. 2.*
4. Chathukulam, J., & Tharamangalam, J. (2021). The Kerala model in the time of COVID19: Rethinking state, society and democracy. *World Development*, *137*, 105207.
5. Chaurasia, V., & Pal, S. (2020). COVID-19 pandemic: ARIMA and regression model-based worldwide death cases predictions. *SN Computer Science*, *1*(5), 1-12.
6. Cirrincione, L., Plescia, F., Ledda, C., Rapisarda, V., Martorana, D., Moldovan, R. E., & Cannizzaro, E. (2020). COVID-19 pandemic: prevention and protection measures to be adopted at the workplace. *Sustainability*, *12*(9), 3603.
7. Ghosh, P., Ghosh, R., & Chakraborty, B. (2020). COVID-19 in India: Statewise Analysis and Prediction. *JMIR public health and surveillance*, *6*(3), e20341. https://doi.org/10.2196/20341
8. Krishnamoorthy, P., Kurli, R., Patil, S. S., Roy, P., & Suresh, K. P. (2019). Trends and future prediction of livestock diseases outbreaks by periodic regression analysis.
9. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., & Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England journal of medicine*.
10. Liu, Y. C., Kuo, R. L., & Shih, S. R. (2020). COVID-19: The first documented coronavirus pandemic in history. *Biomedical journal*, *43*(4), 328-333.
11. Mai, M. V., & Krauthammer, M. (2016). Controlling testing volume for respiratory viruses using machine learning and text mining. In *AMIA annual symposium proceedings* (Vol. 2016, p. 1910). American Medical Informatics Association.
12. Mwalili, S., Kimathi, M., Ojiambo, V., Gathungu, D., & Mbogo, R. (2020). SEIR model for COVID-19 dynamics incorporating the environment and social distancing. *BMC Research Notes*, *13*(1), 1-5.
13. Omer, S. B., Yildirim, I., & Forman, H. P. (2020). Herd immunity and implications for SARS-CoV-2 control. *Jama*, *324*(20), 2095-2096.
14. Raji, P., & GR, D. L. (2020). Covid Pandemic Analysis using Regression. *medRxiv*.

15. Sexton, N. R., Smith, E. C., Blanc, H., Vignuzzi, M., Peersen, O. B., & Denison, M. R. (2016). Homology-based identification of a mutation in the coronavirus RNA-dependent RNA polymerase that confers resistance to multiple mutagens. *Journal of virology*, *90*(16), 7415-7428.
16. Shams, S. A., Haleem, A., & Javaid, M. (2020). Analyzing COVID-19 pandemic for unequal distribution of tests, identified cases, deaths, and fatality rates in the top 18 countries. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *14*(5), 953-961.
17. Thyagaraju, B. P. C., Gowda, S., Patil, S., Srikantiah, C., & Suresh, K. P. (2020). Future trends of COVID-19 disease outbreak in different states in India: a periodic regression analysis. *Highlights in BioScience*, *3*.
18. Thyagaraju, B. P. C., Rajamani, S., Veerabhadrappa, D. H., Patil, S., Roy, P., Chandrashekar, S., & Amachawadi, R. G. Coronavirus (COVID-19) forecasting in India: Application of ARIMA and periodic regression models.
19. Vaman, R. S., Valamparampil, M. J., Ramdas, A. V., Manoj, A. T., Varghese, B., & Joseph, F. (2020). A confirmed case of COVID-19 among the first three from Kerala, India. *The Indian Journal of Medical Research*, *151*(5), 493.
20. Yan, P. (2008). Distribution theory, stochastic processes and infectious disease modelling. In *Mathematical epidemiology* (pp. 229-293). Springer, Berlin, Heidelberg.
21. Yang, D., Xu, Z., Li, W., Myronenko, A., Roth, H. R., Harmon, S., & Xu, D. (2021). Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan. *Medical image analysis*, *70*, 101992.
22. Ye, Q. H., Qin, L. X., Forgues, M., He, P., Kim, J. W., Peng, A. C., & Wang, X. W. (2003). Predicting hepatitis B virus–positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nature medicine*, *9*(4), 416-423.