Applied Ecology and Environmental Sciences, 2021, Vol. 9, No. 2, 177-185
Available online at http://pubs.sciepub.com/aees/9/2/9
Published by Science and Education Publishing
DOI:10.12691/aees-9-2-9

Analyzing the Effect of Environmental and Demographic Factors on COVID-19 Spread in India Using Statistical Methods: A Case Study

Veena Ghuriani[*,#], Jyotsna Talreja Wassan[*,#]

Department of Computer Science, Maitreyi College, University of Delhi, India
[#]These authors contributed equally to this work.
*Corresponding author: vghuriani@maitreyi.du.ac.in, jtwassan@maitreyi.du.ac.in

Received December 29, 2020; Revised January 06, 2021; Accepted January 15, 2021

Abstract The study focuses on analyzing the factors associated with novel coronavirus (COVID-19) which is reported as a pandemic by the World Health Organization (WHO) using the statistical methods. The main objective of this work is twofold: i) predicting the spread of coronavirus across regions by analyzing the growth rate, death rate, and recovery rate in different states of INDIA and ii) correlating the coronavirus and environmental metadata conditions. This study aims to analyze how environmental metadata such as temperature, humidity and AQI and population demographics can affect the COVID-19 cases across INDIA, a land of natural variety. The study considers COVID-19 cases, recovery, and deaths in March-July 2020 along with the data collected regarding environmental conditions of temperature, humidity, and air quality. The role of environmental data in spreading of COVID-19 in the highly populated country INDIA has potential of being studied for making policies to curb COVID-19. In this paper, we propose use of statistical methods of Spearman correlation and Linear regression to conduct different analyses related to the coronavirus. Our results indicate that varied temperatures, high humidity, and the air quality could serve as drivers for COVID-19 transmission and recoveries. For example, a high correlation of 0.98 was observed between humidity and COVID spread in southern Indian regions of Karnataka and West Bengal with tropical semi-arid steppe climate. Recovery rate found to be related to the environmental conditions such as air quality. Additionally, the population density proved to be one of the major factors affecting COVID-19 spread in INDIA with rural and urban populations. Keeping this in viewpoint, prevention measures accordingly be adopted in Indian States and Union territories for reducing the COVID-19 spread and prepare INDIA as medically and socially a better country amidst this global pandemic.

Keywords: *COVID-19, temperature, humidity, regression analysis, spearman correlation*

Cite This Article: Veena Ghuriani, and Jyotsna Talreja Wassan, "Analyzing the Effect of Environmental and Demographic Factors on COVID-19 Spread in India Using Statistical Methods: A Case Study." *Applied Ecology and Environmental Sciences*, vol. 9, no. 2 (2021): 177-185. doi: 10.12691/aees-9-2-9.

1. Introduction

COVID-19 is an infectious disease resulting in public health emergencies with its epicenter located in Wuhan City, China [1,2]. On 30 January 2020, after the first case was reported in India but the number of cases shot up near March 2020 and still are on progressing waves [2,3]. The increase in COVID-19 cases in INDIA could be attributed to multiple factors, including population density, social aspects, environmental determinants such as temperature, humidity, air quality, etc., and governmental policies. The effects of environmental factors on the COVID cases to support decision-making about disease mitigations in humid and warmer locations such as INDIA are under investigation. Since absolute humidity which is a measure of water content in ambient air, is an important environmental factor for viral spread [4] along with temperature patterns, we examined the variability in temperature with absolute humidity and the COVID-19 spread across India. Understanding the relationship between environmental metadata and COVID-19 spread is a useful factor to foresee the intensity of COVID. Such analysis is still under investigation and the subject of this paper over COVID cases. Additionally, demographics have potential to affect COVID cases across INDIA. However, a silver line in Indian context is the better recovery rate of COVID cases. How the better environmental conditions contribute to recovery of COVID patients is also an interesting area of study. In the current study we analyzed COVID related data sourced from a public dataset available at *https://www.mohfw.gov.in* and other public websites storing India's environmental conditions. The purpose of this study is to see the effect of external factors on COVID dynamics.

2. Background

Previous study by [5] analyzed the effect of environmental indicators on COVID-19 in the USA and stated that temperature serves as a driver for the COVID-19 spread. However, there is no generalization yet of this analysis of environmental factors affecting a diverse country like INDIA [6]. The authors in [7] positively correlated COVID cases with temperature and negatively correlated with humidity. The work in [8] highlighted that high temperature and humidity can reduce the transmission of COVID-19. In [9], authors further studied environmental effects on COVID-19 in Brazilian cities and concluded that the observation of high temperature and high humidity reducing the transmission of COVID-19 could not be applied to tropical regions with great transmission rates. Overall, from state-of-the-art [5,6,7,8,9,10], the effect of any environmental based COVID spread predictions cannot be generalized or correlated with the number of new cases in different countries. Since the effect of environmental conditions over COVID-19 is not apparent, this study aims to see the effect of environmental-based spread in INDIA. In this study we characterized distribution of COVID-19 cases reported between March to July 2020 across all INDIAN States/Union territories marked with diverse environmental conditions. Through the correlation analysis, we were able to identify how environmental parameters are affecting COVID-19 cases, deaths, and recovery in INDIA against state-of-the-art worldwide findings. The analysis of the spread of the COVID-19 cases in different INDIAN States/Union territories is affected by population density. Based on our results, we anticipate our findings to be useful for managing the spread of COVID-19 in INDIA.

3. Data and Methods

The primary dataset used in the current study is related to the number of confirmed cases, deaths, and cured cases every day across States/Union Territories of INDIA. Indian Government has collected the data and made it available for the public (*https://www.mohfw.gov.in/*). The data from March 2020-July 2020 is used in the current study. Additionally, we have collected information on environmental factors (temperature, population, air quality index, humidity) from the sources mentioned in Table 1 for the period that we presume to affect COVID cases/deaths/recoveries. The information on current factors is collected from the various sources mentioned in Table 1 to create a novel dataset with compiled information.

In this paper, COVID-19 cases were studied state-wise to understand the transmission and the growth patterns. The relationship of cases in different states (lying in different weather zones) with population, population density in both rural and urban areas was determined using the correlation analysis. The relationship of COVID cases and environmental conditions was studied on the different Indian States to see the effect of environment on the COVID cases spread. For a block of March-July 2020, the total number of confirmed cases and average temperature, humidity and air quality across a state have been observed for studying the weather-based COVID-19 transmission.

Table 1. Sources Used to Extract Environmental and Demographic Conditions of March-July 2020

S.No.	Environmental Factor	Data Source
1	Temperature	https://www.accuweather.com
2	Population Demographics	https://censusindia.gov.in/census_data_2001/india_at_glance/density.aspx https://www.niti.gov.in/niti/content/population-density-sq-km
3	Air Quality Index	https://air.plumelabs.com/air-quality/india https://app.cpcbccr.com/AQI_India/# http://www.cpcb.nic.in/AQI_Bulletin.php
4	Humidity	https://www.timeanddate.com/weather/india/new-delhi/climate

In this paper, the following three different tasks are performed: -

Task I) Trends of COVID-19 Dynamics in India.

The first case of COVID-19 in India was reported in [1]. The Ministry of Health and Family Welfare (MoHFW) confirmed a total of 697,413 cases, 424,432 recoveries and 19,693 deaths in the country as on July 6, 2020. In this task we have tried to study the trend of spread of COVID-19 in various States of India. We have collected a dataset about COVID 19 spread in different States/Union Territories in India from March - July 2020. This data set comprises information about confirmed cases, deaths, cured date-wise and state-wise. Analysis of the data was performed to study the patterns in the data. Firstly, we calculated the recovery rate, mortality rate and growth rate. And thereafter we analyzed the increase of cases daily, the growth rate, recovery rate and the mortality rate in different states and compared them. The data has been represented graphically for better understanding.

Task II) Correlating the coronavirus and metadata conditions. The environmental conditions that affect the spread of COVID 19 ~~is~~ are not same for all regions. The main objective of this task was to attain the count of confirmed cases of Indian states date wise with temperature, humidity, air quality index and population demographics. To understand the correlation between the COVID spread and environmental conditions Spearman correlation method was used as it is independent of the underlying data distribution. Spearman's correlation is a measure of strength of association between two variables and the direction of their relationship with values ranging between +1 and greater than −1. A value +1 or −1, means a perfect correlation between two variables [11]. For the Spearman correlation~~,~~ is a nonparametric test that does not carry any assumptions about the distribution of the data and is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal. Following is the formula which is used to calculate the Spearman's correlation.

$$\textit{Spearman Correlation} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (1)$$

where n= number of data points of the two variables and di= difference in ranks of the "i th " element. To perform this task, Spearman correlation was found between humidity, temperature, air quality, population density, gender ratio and confirmed COVID-19 cases in six different regions of INDIA. The correlation tables for temperature, humidity, air quality, population density and gender ratio and total confirmed COVID-19 cases were plotted. It was observed from the correlation chart, the confirmed cases are driven by these metadata factors.

Task III) Estimating the regression fit of temperature and humidity with the Recovery Rate. There has been a great improvement in India's COVID-19 recovery rate since the lockdown initiation on March 24, 2020. The number of confirmed COVID-19 cases has seen a steep rise. However, the number of patients recovering has also embarked a great increase correspondingly. Since there was no vaccine for the virus in the assumed period, it was important to see recovery rate trends. The main objective of this task is to predict recovery rate based on its linkage with the environmental conditions. Regression analysis [12] is used to estimate the recovery rate w.r.t to environmental factors of temperature and humidity playing an important role. The regression model reports coefficients of regression and *p*-value of *t*-statistics involved in the regression statistical test [13]. The *p*-value is used as an alternative to the rejection point to provide the least significance at which the null hypothesis would be rejected. In this study the null hypothesis is assumed to be "*No relationship between an environmental factor and COVID-19 recovery rate*".

p-value < 0.05 indicates the null hypothesis is rejected. The model also reports the Coefficient of Determination R^2, which is used as a measure of the goodness of fit. If R^2 is 0.94, it shows that 94% of the values fit the regression analysis model (i.e. 94% of the dependent variables (y-values) are explained by the independent variables (x-values)).

Standard Error is another goodness-of-fit measure indicating the precision of regression modelling. The smaller the error, the more certain is the regression. Standard Error shows the average distance on which the data points fall from the regression line unlike R^2 which depicts the percentage of the dependent variables variance as determined by the regression model. We carried out experimentations considering Y as recovery rate and *Xi*'s as environmental factors of temperature and humidity. From our experimentation we received an indication of a relationship between environmental factors and the recovery rate of COVID-19 by modelling of data with Least Squares method (LSM) [14]. LSM establishes a relationship between the dependent and independent variable along a linear line. The least squares regression line is the one data fitting line that has the smallest possible value for the sum of the squares of the residuals amongst all the possible linear fits (Figure 1).

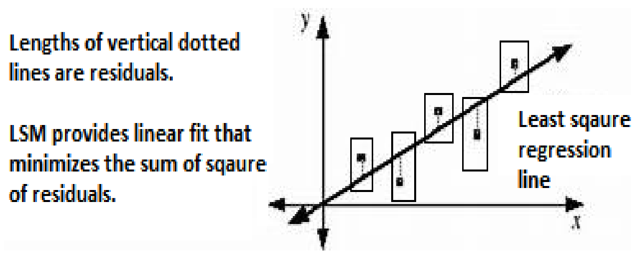

Figure 1. Linear Regression Modelling with LSM

4. Results

Task I) Trends of COVID-19 Dynamics in India.

COVID-19 virus speeded at immense rates amongst countries all over the world. It is important to understand how the virus is spreading among different regions covering various States and Union territories in INDIA. To predict how many persons are infected each day, records for the number of confirmed cases across a variety of states were taken. To prepare the dataset for this task firstly, the total number of confirmed deaths and recovered cases per day was calculated by subtracting the total number of cases a day before to the total number of cases on the current day. The number of COVID-19 cases has seen a steep rise from 19 March 2020. On March 19, every confirmed case in India was transmitting the virus to 1.7 people on average and the number had risen to 1.81 by March 26, 2020. However, the growth rate declined greatly in July 2020. In March 2020, the daily growth rate was approximately 30 % whereas subsequently on May 31 it was 4.82% and which was further reduced in July to 3.24%. Looking at the vast population of India, this growth rate of COVID cases is very low. It is worth to see how the environmental conditions played role in severely affected areas with COVID-19 cases. The number of patients recovering has also seen a substantial increase correspondingly. The distribution of recovered cases across Indian states followed the same order as the total cases. Maharashtra, Delhi, and Tamil Nadu have been the worst COVID-19 hit states across India. Other majorly impacted states are Karnataka, Uttar Pradesh, West Bengal, Gujarat, Rajasthan, and Assam. Interestingly these belong to regions with different environmental conditions. The growth in COVID cases, death rate and recovery rate in the states are pictorially shown in Figure 2. Maharashtra recorded the highest number of COVID-19 cases with the sharpest rise in COVID-19 deaths (Figure 2). Tamil Nadu, although witnessed a great surge in COVID-19 cases but had the least number of mortalities. We have tried to analyze the rate of increase in confirmed cases, deaths, and recovery in the most affected states. It is observed that although the cases are increasing with time, the recovery rate is significantly higher in all the states (Figure 2). The mortality rate is below 3% in most of the states except for Maharashtra, Gujarat, and West Bengal.

Task II) Correlating the coronavirus and metadata conditions

The study indicates empirical estimations of environmental indicators on COVID cases, deaths or recovered cases. The environmental metadata affecting the spread of COVID 19 is not the same for all states of INDIA. Across different states, it is important to analyze which environmental conditions mostly affect the COVID-19 spread. The main objective is to get data regarding the count of confirmed cases of different Indian states date-wise with temperature, humidity, and air quality index and how these factors are affecting the COVID-cases. The environmental indicators included in this study are average temperature, highest temperature, average humidity, and air quality index (AQI). To perform the task of associating environmental indicators of different states of INDIA, data were sourced from different websites as mentioned in the Table 1 giving weather information, which consisted of humidity and temperature for the months of March 2020-July 2020. Spearman rank correlation tests [11] were conducted for analysing relationships as these do not assume normal distribution of COVID data.

India is primarily divided into 6 zones based on the location: 1) North most, 2) Northeast, 3) North 4) North west 5) Southern India and 6) India West Coast (Figure 3). Humid Subtropical climate is usually observed in north; Tropical Monsoon climate in east, west coast and warm arid climate in the Northwest areas; tropical semi-arid steppe climate in southern India areas such as Tamil Nadu. Amidst these different environmental conditions, COVID spread is being affected differently. To analyse the effects of correlated environmental conditions and total positive cases, deaths and recoveries different statistical correlation charts were plotted for different states covering each zone. As the regions vary in their environmental conditions, these may influence the spread of COVID-19 differently. The environmental conditions were drier and warmer across North and North west Zones relatively. In the state of Rajasthan, New Delhi, Uttar Pradesh the maximum temperature was observed to be 46°C , 45°C, and 44°C, respectively, which is different from coastal tropical areas of Maharashtra and Karnataka where the temperature was observed to be 38°C and 36°C, respectively. Even in subtropical areas of West Bengal and Assam, maximum temperature was observed as 39°C and 34°C. COVID Cases seem to get affected differently in varied temperature regions. In dry regions with high temperatures, COVID cases seemed to increase with rising temperatures; however, in tropical/sub-tropical areas COVID cases seemed to increase with lowering of temperatures. The less temperature posed as a risk factor for states which have a subtropical climate. Furthermore, combining humidity as an environmental factor with temperature may pose more risks for the spread of COVID-19. The correlation analysis is performed from March-July 2020 for COVID cases in most affected states of INDIA studying their relationship with environmental factors in the form of correlation heatmaps depicting the relationship between average temperature, average humidity and COVID cases (Figure 4 (a)-(i)).

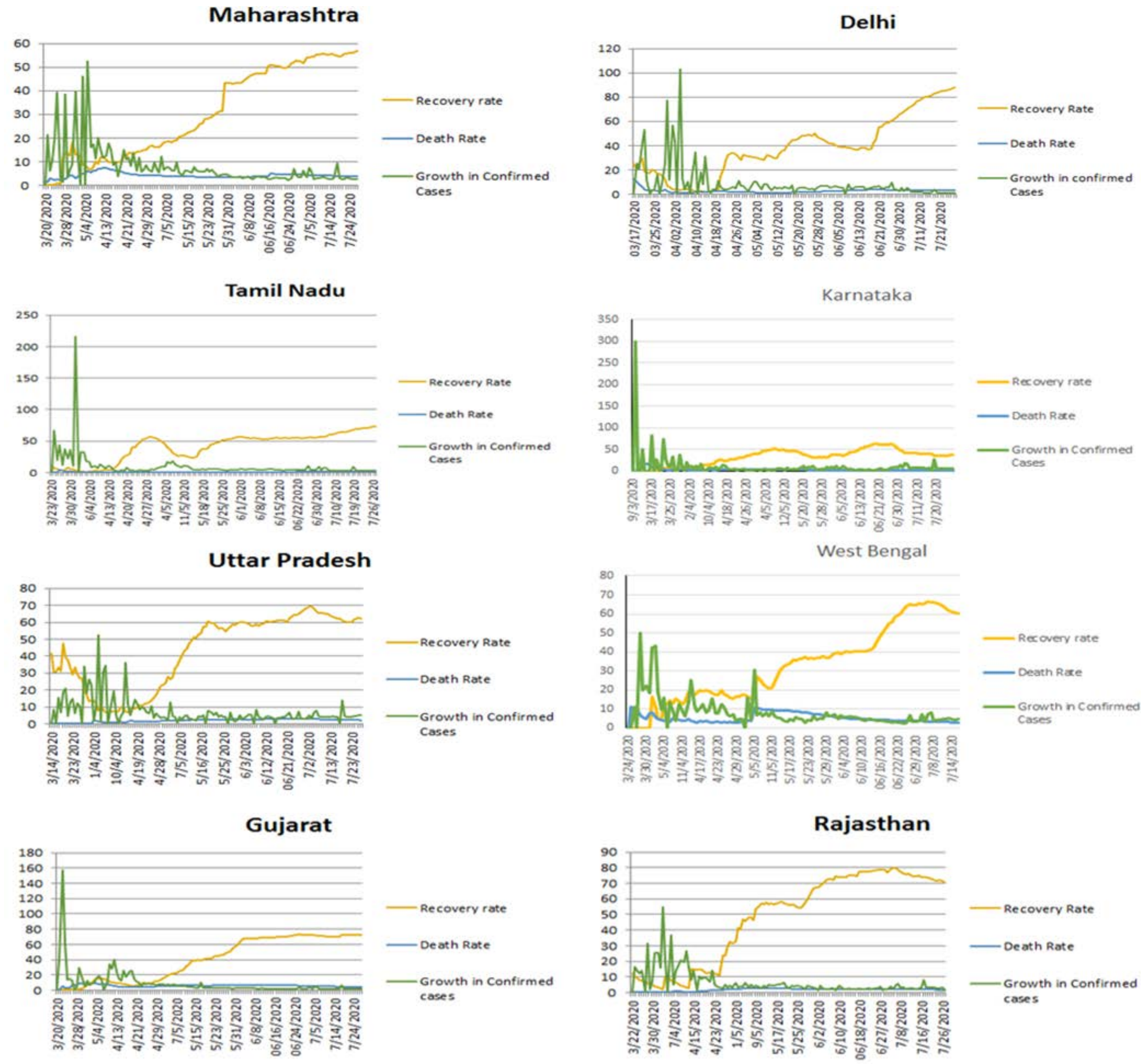

Figure 2. Growth, Death and Recovery Rate amongst most affected states due to COVID-19

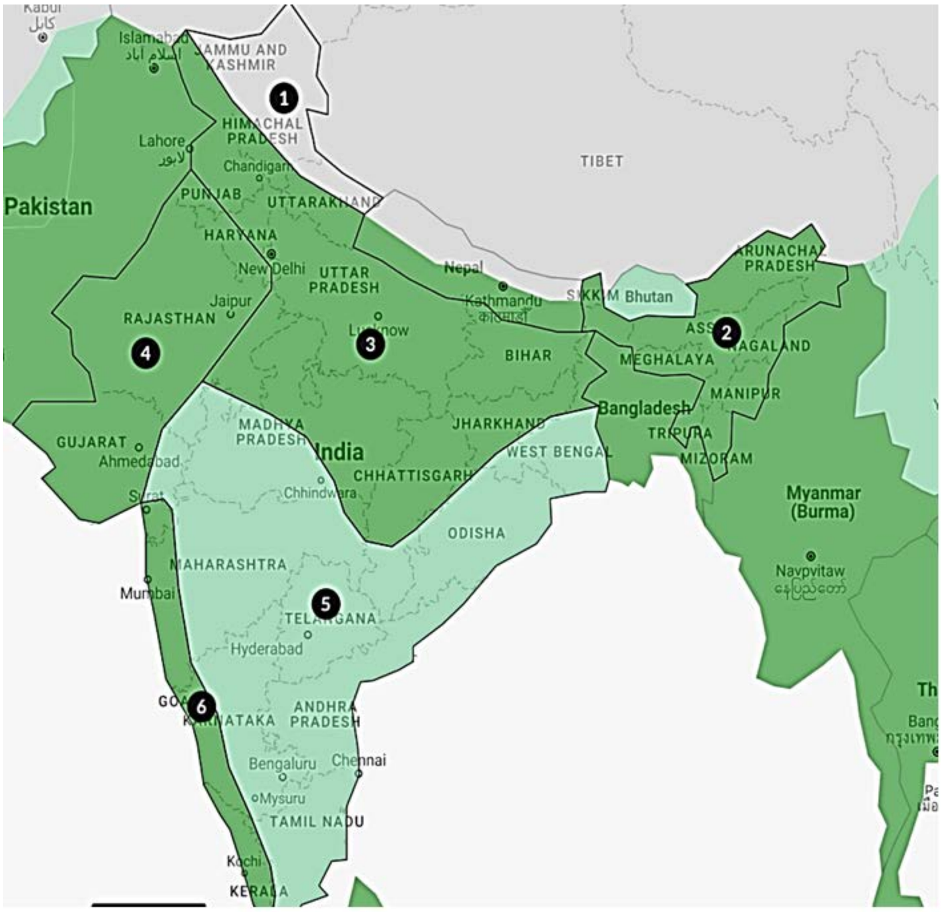

Figure 3. Indian States in different Zones (1-6) based on location and climate

Delhi's climate is an overlap of humid subtropical and semi-arid conditions. The humidity levels are low to moderate. It is observed from the correlation chart, the number of confirmed cases, mostly affected by temperature and humidity in Delhi (a state in Zone 3) with relatively high modulus of coefficient value as 0.55 and 0.54 respectively (Figure 4(a)). Delhi's extreme weather conditions and vast population makes it prone to spread of COVID cases. The climate of Uttar Pradesh is classified as humid subtropical, it also experiences extreme temperatures in summers and winters like Delhi. The correlation chart (Figure 4(b)) indicates the number of confirmed cases, mostly affected by temperature and humidity in Uttar Pradesh (a state in Zone 3) with coefficient value as 0.38 and 0.55 respectively (Figure 4(b)). In Zone 4 (Figure 3) with warm arid temperature, state of Rajasthan also indicates from the Spearman test, average temperature is positively related to the total number of COVID positive cases (with a correlation of 0.56), and humidity with cases (a high correlation of 0.69) (Figure 4 (c)). The predominantly high temperature is a key factor in the spread of COVID cases.

The state of Gujarat has varied climatic conditions. The northern part adjoining Rajasthan is hot and dry, and the southern part has sub-tropical climate with very high humidity. The Spearman test indicates the average temperature is negatively related to the total number of COVID positive cases (with a correlation of 0.34), and humidity with cases (a high correlation of 0.97) (Figure 4 (d)) in Gujarat. Humidity plays a driving role in the spread of COVID in Gujarat. The state of Maharashtra has a typical monsoon climate. The humidity level is very high throughout the year. The correlation table indicates a very high correlation of humidity with the spread of COVID-19 (correlation coefficient of 0.96) and negatively associated with temperature (Figure 4 (e)) in Maharashtra.

The weather in Karnataka (Zone 5) ranges from arid to humid to tropical due to its varying geographical conditions. It was observed that spread of COVID-19 is associated with humidity (with a very high correlation coefficient of 0.98) and negatively associated with temperature (-0.77) (Figure 4(f)). Like all coastal regions, humidity has played a key role in the spread of COVID in Karnataka. West Bengal has a hot and humid climate (lies in Zone 5 shown in Figure 3). West Bengal (Figure 4 (g)), also indicates a high positive relation between humidity and COVID spread (0.98), however it depicts temperature to be negatively associated with COVID-spread.

Assam has a tropical monsoon climate, which is characterized by heavy monsoon rains. In the state of Assam in Northeast of India (Zone2) (Figure 3), humidity found to be positively correlated with COVID spread (correlation of 0.92) but temperature found to be negatively correlated indicating as temperature increases, COVID cases decreases in number (Figure 4 (h)). In the state Jammu & Kashmir (J&K), North most of India (Zone1) (Figure 3) with Steppe and cold weather zone, Humidity found to be negatively correlated with COVID spread (correlation of -0.26) but temperature found to be positively correlated indicating as temperature increases, COVID cases increases in number in such areas (Figure 4 (i)).

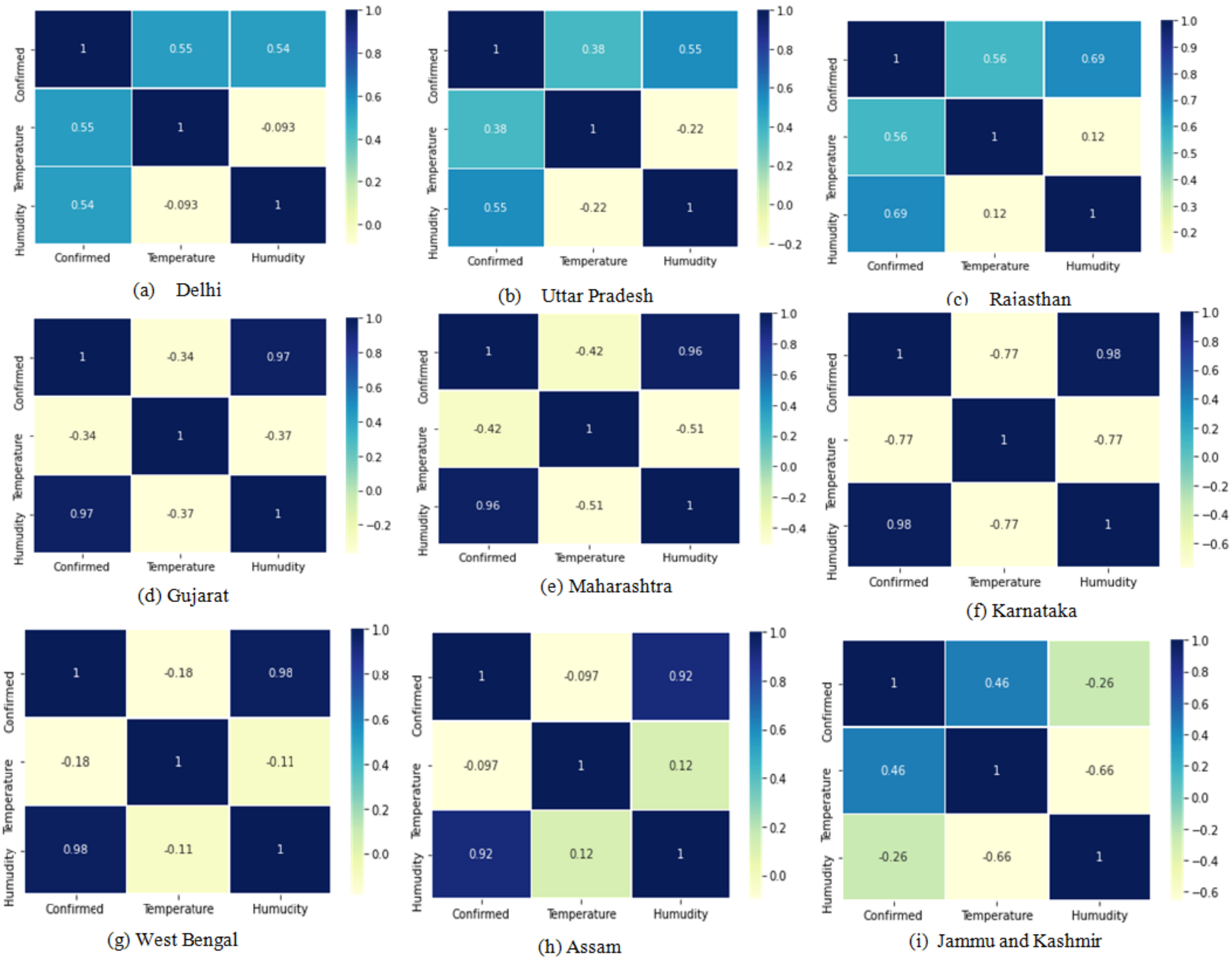

Figure 4(a)-(i). Correlation tables for temperature, humidity, and COVID19 cases in Delhi, Uttar Pradesh, Rajasthan, Gujarat, Maharashtra, Karnataka, West Bengal, Assam, and Jammu and Kashmir

The overall analysis shows that humidity and COVID confirmed cases in INDIA affect each other. It is observed that the number of COVID cases increases with an increase in humidity across states in different locality zones. Too much humidity could proliferate the viral spread as in tropical-subtropical areas, airborne droplets may bear the corona virus for longer periods. Also, it is observed that the number of positive cases decreases with an increase in temperature in tropical climate states of INDIA, however in the warm/arid States, COVID-19 cases increase with increasing temperatures.

Unlike other studies [8,15] in India, higher temperatures in warm arid/semi-arid areas combined with humidity posed more risk of the COVID spread. High humidity across all over INDIA, favored COVID-19 transmission except in J&K which is a hilly cold region (Zone 1). Comparatively low temperatures, as observed in tropical regions of INDIA (areas at verge of Zone 5 and 6; e.g., in Karnataka and Maharashtra), combined with high humidity, can favors the transmission. Average temperature with high humidity favors the transmission of COVID-19 in states of INDIA as was observed in [16] in the states of Brazil.

In the next steps of analysis, we noticed the effect of AQI overgrowth in cases of COVID and recovery rate in different states of India being affected by the air quality. The lockdown imposed in India due to COVID-19 had significantly improved the air quality across all cities in India. According to the report released by Central Pollution Control Board of India, the air quality improved due to restricted industrial activities and vehicular emissions [17]. We also tried to find correlation between AQI and recovery and death rate of COVID cases across the 10 states we have considered above. It can be observed in the Figure 5 that there is positive correlation of 0.23 between AQI and the recovery rate. It implies that the improved air quality during the lockdown period has helped in the recovery of the COVID patients. Also, an important observation is the silver line of recovery rate been significantly higher than the death rate (as seen in Figure 2).

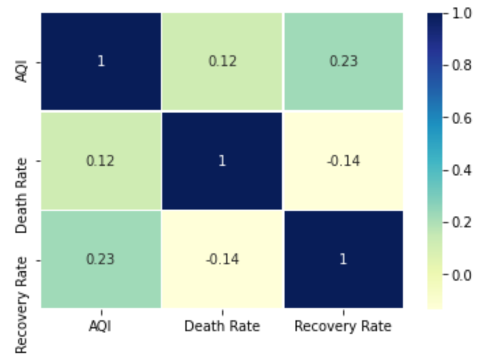

Figure 5. Correlation between AQI, Recovery rate and Death rate across INDIA

In addition to the environmental meta-factors, COVID cases are also influenced by the high population of INDIA and variety in Indian States. Different States of INDIA have different population sizes and population densities. Considering that the population size significantly, the infection spread rate of different areas are different. Therefore, it will be helpful to distinguish most affected areas even after studying the role of environmental factors on COVID-19.

Observing the overall summary from the sourced dataset Maharashtra, Delhi, Gujarat, Rajasthan, Tamil Nadu, and Uttar Pradesh were observed to be the most affected states in terms of the COVID spread. There are highly populated regions, and Delhi and Maharashtra are the urban hubs posing as most popular states. Highly populated regions (Delhi, Maharashtra) have catalyzed the spread of COVID-19. The states with higher population density are more responsive to COVID prevalence. Figure 6 indicates that cases and deaths due to COVID are highly correlated with Urban population (a correlation of 0.82) and so is the number of cases (a correlation of 0.81). However, comparatively a smaller number of deaths and cases were observed in rural populations (a correlation of 0.63).

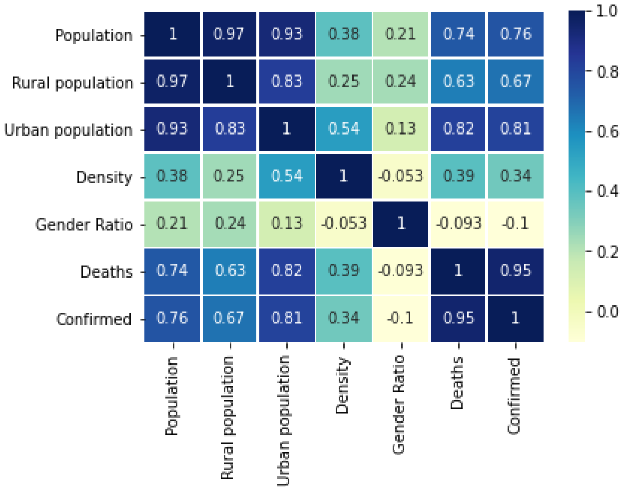

Figure 6. Impact of Demographic Indicators on COVID-19 Data obtained in INDIA (Correlation Heatmap depicting relationship of population density with COVID cases, death, and recovery cases)

The study also observed a factor of gender ratio. However, it reported that gender ratio (males to females) does not seem to have much effect on COVID spread. From the analysis it seems that Males and females are almost equally susceptible to the COVID spread (Figure 6).

Task III) Estimating the regression fit of temperature and humidity with the Recovery Rate.

Since, the silver line lies in better recovery rate of COVID cases in INDIA, it is interesting to study how temperature and humidity plays a role in recoveries of COVID-19. In the current study, Regression analysis was done to estimate the effect of temperature and humidity on recovery rate. The regression model reported coefficients of regression and *p*-values of t-statistic involved in the test. The model also reports the Coefficient of Determination R^2, which is used as a measure of the goodness of fit [18]. If R^2 is 0.94, it shows that 94% of the values fit the regression analysis model (i.e. 94% of the dependent variables (y-values) are explained by the independent variables (x-values)). Standard Error served to be another goodness-of-fit measure indicating the precision of regression modelling. The smaller the error, the more certain is the regression. Standard Error showed the average distance on which the data points fall from the regression line unlike R^2 which depicts the percentage of the dependent variables variance as determined by the regression model. We carried out experimentations considering dependent variable Y as outcome (confirmed, death or recovery rate) and independent variables Xi's as environmental factors of temperature and humidity in the equation of linear regression (Eq. 2).

$$Y_i = \propto + \beta_i X \qquad (2)$$

where the model can be used to draw inferences about any linear relationship between Y and X. From our experimentation we received an indication of a relationship between environmental factors and the recovery and death rate of COVID-19 by modelling data of different Indian states/territories. The significance level values ≤ 0.05 (*p*-value) for the $\chi 2$ test were used to select these canonical pairs. The statistical evidence indicated that average temperature, and humidity have a significant effect on the COVID-19 recovery rate and hence its spread across different states of INDIA. The results indicate, higher the average temperature, higher is the recovery rate. But lesser the humidity, more is the recovery rate. Temperature is positively related (coefficient in regression equation: 1.68) and humidity is negatively related (coefficient in regression equation: -0.42) with recovery rate. Figure 7 and Figure 8 represents the respective regression fit of average temperature and humidity with recovery rate.

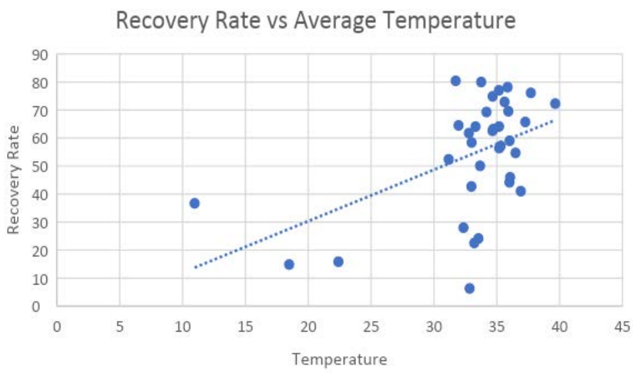

Figure 7. Relation of COVID Recovery Rate with Average Temperature across INDIA

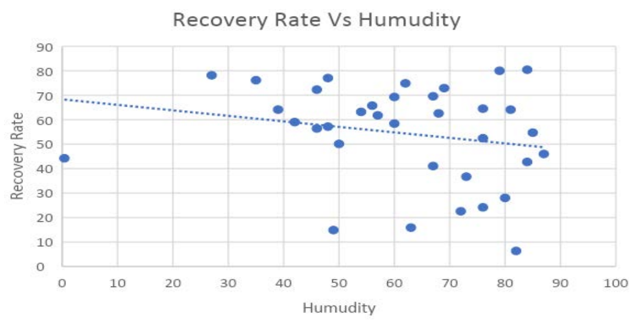

Figure 8. Relation of COVID Recovery Rate with Average Humidity across INDIA

5. Discussions

This study generated new findings on the COVID dynamics, environmental characteristics affecting COVID-19 in India.

A positive aspect is even with the rapid spread of COVID-19 across Indian states, the recoveries have been high, and deaths have been low. The COVID cases were observed to be more in the urban population compared to in rural populations. Also, it was observed that gender does not affect the COVID-19 spread. Both the males and females are equally susceptible to the COVID infection. The highly affected states seemed to be Maharashtra, and Tamil Nadu. The below 3% mortality rate observed in India is almost half of the average worldwide mortality rate of 6% [19]. Additionally, the disease spread has been predicted to be affected by the environmental conditions of India in contrast to the findings published in [6]. Our results showed that higher average humidity favored the COVID-19 spread which is in line with the observation conducted w.r.t the Brazilian cities studies in [9]. In tropical regions, high humidity seemed to be a cause of COVID cases with high correlation values. A high correlation of 0.98 was observed between humidity and COVID spread in southern Indian regions. In warm and arid regions, a correlation of 0.56 was observed between temperature and COVID spread. Temperature in warm, arid, or semi-arid areas seemed to be positively correlated with the COVID spread, however in wet/monsoon tropical regions, temperature found to be negatively correlated with the COVID-spread. Also, given the high recovery rates in India, average temperature is positively correlated with recoveries whereas negatively correlated with the humidity. Another interesting observation was the improvement in air quality post lockdown imposed due to COVID-19 and better air quality indicated better recovery rate of COVID cases in INDIA. High temperature also seemed to have a positive effect on recovery rate. However, humidity is negatively correlated with the recovery rate.

6. Conclusion

Although COVID cases increased in INDIA in between March to July 2020, a silver line is that the recovery rate is significantly better than the growth rate across varied states. Additionally, death rate is low in comparison to growth rate. Environmental indicators are playing a key in the spread of COVID-19 in INDIA. Our paper demonstrates that environmental factors affect the COVID-19 spread, and the recovery rate across INDIA. This study attempts to model the environmental factors of average temperature, average humidity and air quality with the COVID-19 spread in INDIA with varied climate across different regions. For the period Mar-July 2020, the COVID-19 cases in INDIA were observed to be favored by varied temperatures as well as the high average humidity across different states. COVID cases are directly related to high humidity in Indian regions. Also, the significance of air quality implies it could help in reducing the spread of infectious diseases and enhancing the recovery rate affecting respiratory systems such as COVID-19. It is also worth considering that there is disparity in terms of social-economic conditions in addition to environmental conditions in INDIA. In this sense, the rural and urban population are being impacted by COVID-19 in different ways. Mainly urban States are the ones where the COVID spread has been increasing rapidly. Henceforth, there is more need for prevention measures in these urban cities aiming to reduce transmission. Current study is of exploratory nature and can be enhanced in future with more environmental parameters such as change in weather conditions, carbon emission, wind speed, etc., and other demographics such as state wise testing, condition of hospitals, water waste management etc. with socioeconomic conditions of Indian states/territories to provide a comprehensive insight for the fight against COVID-19, which is an active area of research.

Declaration of Competing Interest

The authors declare that they have no known competing interests that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank *Summer Projects Research Team* at Maitreyi College, University of Delhi to motivate us for undertaking this study.

References

[1] M. A. Andrews *et al.*, "First confirmed case of COVID-19 infection in India: A case report," *Indian Journal of Medical Research*. 2020.

[2] Z. Y. Zu *et al.*, "a Coronavirus Disease 2019 (COVID-19): A Perspective from China. Radiology", 200490. *Radiology*, 2020.

[3] R. B. Bhagat, R. R.S., H. Sahoo, A. K. Roy, and D. Govil, "The COVID-19 , Migration and Livelihood in India," *Int. Inst. Popul. Sci. Mumbai*, 2020.

[4] W. Luo *et al.*, "The role of absolute humidity on transmission rates of the COVID-19 outbreak," pp. 1-7, 2020.

[5] M. F. Bashir *et al.*, "Correlation between climate indicators and COVID-19 pandemic in New York, USA," *Sci. Total Environ.*, 2020.

[6] S. Gupta, G. S. Raghuwanshi, and A. Chanda, "Effect of weather on COVID-19 spread in the US: A prediction model for India in 2020," *Sci. Total Environ.*, 2020.

[7] B. Oliveiros, L. Caramelo, N. C. Ferreira, and F. Caramelo, "Oliveiros et al 2020 Temperature and humidity in doubling time of cases," 2020.

[8] J. Wang, K. Tang, K. Feng, and W. Lv, "High Temperature and High Humidity Reduce the Transmission of COVID-19," *SSRN Electron. J.*, 2020.

[9] A. C. Auler, F. A. M. Cássaro, V. O. da Silva, and L. F. Pires, "Evidence that high temperatures and intermediate relative humidity might favor the spread of COVID-19 in tropical climate: A case study for the most affected Brazilian cities," *Sci. Total Environ.*, 2020.

[10] M. M. Sajadi, P. Habibzadeh, A. Vintzileos, S. Shokouhi, F. Miralles-Wilhelm, and A. Amoroso, "Temperature and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19," *SSRN Electron. J.*, 2020.

[11] P. Sedgwick, "Spearman's rank correlation coefficient," *BMJ (Online)*. 2014.

[12] C. Angelini, "Regression analysis," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 2018.

[13] P. Humphrey, “What is a p-value?” *Teach. Stat.*, 2012.
[14] J. T. Kilmer and R. L. Rodríguez, “Ordinary least squares regression is indicated for studies of allometry,” *Journal of Evolutionary Biology*. 2017.
[15] P. Mecenas, R. T. da Rosa Moreira Bastos, A. C. Rosário Vallinoto, and D. Normando, “Effects of temperature and humidity on the spread of COVID-19: A systematic review,” *PLoS One*, 2020.
[16] L. D. Martins, I. da Silva, W. V. Batista, M. de F. Andrade, E. D. de Freitas, and J. A. Martins, “How socio-economic and atmospheric variables impact COVID-19 and influenza outbreaks in tropical and subtropical regions of Brazil,” *Environ. Res.*, 2020.
[17] S. Gautam, “The Influence of COVID-19 on Air Quality in India: A Boon or Inutile,” *Bull. Environ. Contam. Toxicol.*, 2020.
[18] J. W. Gooch, “Coefficient of Determination,” in *Encyclopedic Dictionary of Polymers*, 2011.
[19] M. Palacios Cruz, E. Santos, M. A. Velázquez Cervantes, and M. León Juárez, “COVID-19, a worldwide public health emergency,” *Revista Clinica Espanola*. 2020.