

DOI: <http://dx.doi.org/10.5281/zenodo.4677470>

# Use of statistical analysis to monitor novel coronavirus-19 cases in Jammu and Kashmir, India

Digvijay Pandey <sup>1</sup>, Tajamul Islam <sup>2\*</sup>, Juniad A. Magray <sup>2\*</sup>, Aadil Gulzar <sup>3</sup>, Shabir A. Zargar <sup>2</sup><sup>1</sup> Department of Technical Education, IET, Lucknow, India<sup>2</sup> Department of Botany, University of Kashmir, Srinagar-190006, J & K, India<sup>3</sup> Department of Environmental Science, University of Kashmir, Srinagar-190006, J & K, India\* Corresponding author: E-mail: [islamtajamul66@gmail.com](mailto:islamtajamul66@gmail.com); [junaidmagray786@gmail.com](mailto:junaidmagray786@gmail.com)

Received: 12 March 2020; Revised submission: 30 March 2021; Accepted: 09 April 2021

<https://jbrodka.com/index.php/ejbr>Copyright: © The Author(s) 2020. Licensee Joanna Bródka, Poland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>)

**ABSTRACT:** Coronavirus disease (COVID-19) has been increasing slowly and steadily in all the districts of Jammu and Kashmir, India. It is essential for the government and health management system to monitor the districts affected due to COVID-19. The main objective of this study is to ascertain and categorize the COVID-19 affected districts into real clusters based on similarities within a cluster and differences among clusters in order to imply standard operating procedures (SOPs) policies, decisions, medical facilities, etc. could be improved for reducing the risk of infection and death and optimize the deployment of resources for preventing subsequent outbreaks.

**Keywords:** Clusters; Coronavirus; SARS-CoV-2; Districts; Jammu and Kashmir; Similarities.

## 1. INTRODUCTION

The ongoing outbreak of pandemics caused by a novel coronavirus was originated from the local seafood market, Wuhan of Hubei Province, China, in late December 2019 [1-3]. Within a short period, an infection spread all over the world [4-6]. On 30th January 2020, the WHO declared this outbreak as a Public Health Emergency of International Concern [7-8]. Genetically coronaviruses (CoVs) are a diverse group with positive ssRNA as genetic material, enveloped with a protein coat (capsid) [9]. Several organ systems are affected by the coronavirus like respiratory, enteric, hepatic, etc., with varying severity among humans and animals [10-12]. There are some variants in CoVs such as HCoV-OC43, HCoV-229E, HCoV-NL63, and HCoV-HKU1, which may cause mild respiratory illness [13-15]. In the past two decades, two nCoVs: SARS-CoV and MERS-CoV, have emerged, which causes more severe human respiratory infections [16, 17]. During the previous outbreak of epidemics caused by SARS-CoV, there were cases of about 8000 people worldwide with nearly ca. 800 deaths, representing a mortality rate of around 10%. Whereas in MERS-CoV, the fatality rate was ca. 35%, caused 334 deaths in 857 officially infected cases. SARS-CoV is the seventh member of the family of CoVs that is zoonotic and infects humans. Fever, fatigue, and cough are the main symptoms of COVID-19, similar to SARS-CoV and MERS-CoV diseases. There is distinctness in the etiology and pathogenesis of these CoVs, which causes severe diseases in humans [18].

On 16th March 2020, the first case of COVID-19 was reported in Jammu and Kashmir (J&K), which had a travel history from Saudi Arabia. On the day when the first patient was tested positive for the novel coronavirus, the state government declared the disease as an epidemic in the summer capital (Srinagar-SGR) of J&K and closed commercial and educational establishments. The district magistrate banned the assemblage of people and some other measures. On March 18, 2020, the government imposed section 144 and lockdown in the summer capital of Union Territory (J&K) due to an increase in the number of COVID-19 cases. Later on 22nd March 2020, the government of India (GoI) declared a 14-hour public curfew and ordered the closure of all educational institutions and commercial offices. Further, on 24th March 2020, the government announced the nationwide lockdown (phase I) for 21 days and after the completion of phase I lockdown, the government extended it up to 3rd May 2020. To stop the spread of coronavirus disease (COVID-19), many steps were taken by the state and UT's government [19]. This study aims to classify and categorize the COVID-19 affected districts by clustering based on similarities in confirmed, active cured and death cases. This will help the government recognize the most or less affected Jammu and Kashmir districts to optimize the screening, lockdown, curfews, and other legal steps in severely affected districts or areas. Furthermore, the study will be beneficial in understanding the status of increment in COVID-19 cases across various districts and will insight the government, doctors and NGO's to improve their policies which will be helpful to improve the various medical facilities such as ventilators, testing kits and masks that will ultimately reduce the spread of infection across the region.

## 2. MATERIAL AND METHODS

### 2.1. Study area

Jammu and Kashmir (J&K) is a union territory of India that lies to the north of Himachal Pradesh and Punjab and to the west of Ladakh. It has a Mediterranean type of climate. As per Census 2011, J&K (including Ladakh) has a population of 1.25 Crores. The Jammu and Kashmir consists of two divisions, each comprises of ten districts: Srinagar (SGR), Anantnag (ANG), Bandipora (BD), Baramulla (BR), Ganderbal (GD), Budgam (BG), Kulgam (KL), Pulwama (PL), Kupwara (KP), Shopian (SP) are districts of Kashmir division. While as Jammu (JM), Ramban (RB), Reasi (RS), Udhampur (UD), Kathua (KT), Kishtwar (KW), Poonch (PN), Rajouri (RJ), Samba (SM), Doda (DD) forms the Jammu division. There are 217 Tehsils, 558 Niabats and 7055 Villages in the UT (including Ladakh as per census 2011).

### 2.2. Methodology

The present study has been divided into three parts. Part I consists of data collection; part II consists of a statistical analysis of COVID-19 data set using cluster analysis (Brey-Curtis); and part III consists of analysis using radar charts to depict the number of confirmed, active, cured and death cases in each district.

#### 2.2.1. Part I. Data collection and exploratory analysis

The data of all cases (confirmed, active, cured and death cases) related to COVID-19 have been retrieved from March 16, 2020, to January 5, 2021, from the website of "COVID-19 Monitoring Dashboard maintained by the Ministry of Health and Family Welfare Government of India (GoI) [20].

From the website of COVID-19 monitoring dashboard, data of all COVID-19 affected districts: ANG, BD, BR, BG, DD, GD, JM, KT, KW, KL, KP, PN, PL, RJ, RB, RS, SM, SP, SGR, UD have been collected. The data consist of four variables: the number of confirmed, active, cured/discharged and death cases. The total number of confirmed, active, cured and death cases during the above-mentioned period are 121786, 2684, 117211 and 1891, respectively as given in Table 1.

**Table 1.** The total number of confirmed, active, cured and death cases from March 16, 2020, to January 5, 2021.

District	Confirmed cases	Active cases	Cured cases	Death cases
Anantnag (ANG)	4823	108	4632	83
Bandipora (BD)	4660	58	4542	60
Baramulla (BR)	8004	120	7712	172
Badgam (BG)	7644	108	7426	110
Doda (DD)	3398	45	3289	64
Ganderbal (GD)	4500	81	4375	44
Jammu (JM)	24058	677	23024	357
Kathua (KT)	3214	42	3123	49
Kishtwar (KW)	2722	13	2688	21
Kulgam (KL)	2666	57	2556	53
Kupwara (KP)	5571	123	5357	91
Poonch (PN)	2462	41	2397	24
Pulwama (PL)	5600	151	5361	88
Rajouri (RJ)	3846	128	3664	54
Ramban (RB)	2113	22	2070	21
Reasi (RS)	1628	12	1601	15
Samba (SM)	2775	202	2534	39
Shopian (SP)	2524	107	2378	39
Srinagar (SGR)	25482	537	24495	450
Udhampur (UD)	4096	52	3987	57
Total	121786	2684	117211	1891

### 2.2.2. Part II. Cluster analysis (CS)

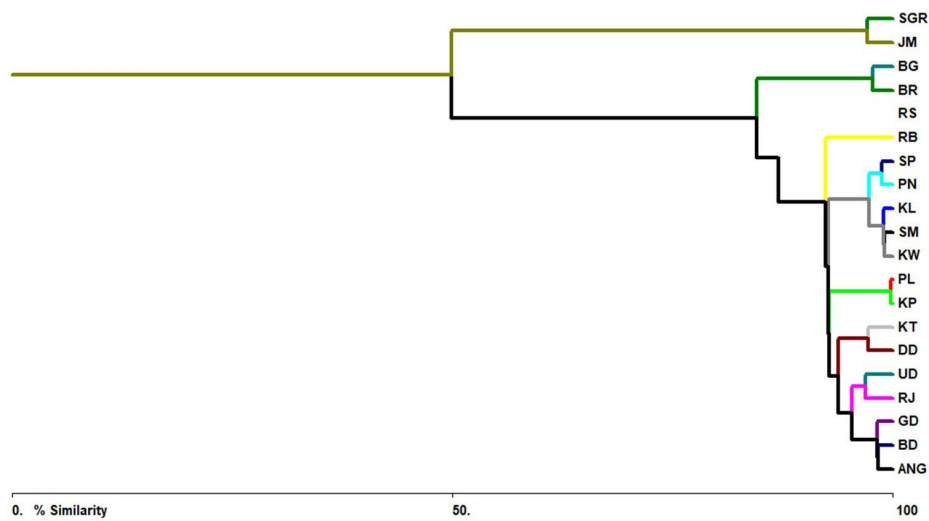
Cluster analysis is one of the best data analyzing techniques by which the sample variables are clustered into groups based on their similarities within a group and dissimilarities among different groups [21, 22]. The Bray-Curtis method is the robust and most conventionally used method that does not require prior postulation and uses variance analysis to calculate similarities among different clusters [23, 24]. This study used the OriginPro software (version OriginPro 2019b-64 bit) to accomplish the cluster analysis. The data set has been scaled properly before executing the cluster analysis.

### 2.2.3. Part III. Analysis using a radar chart

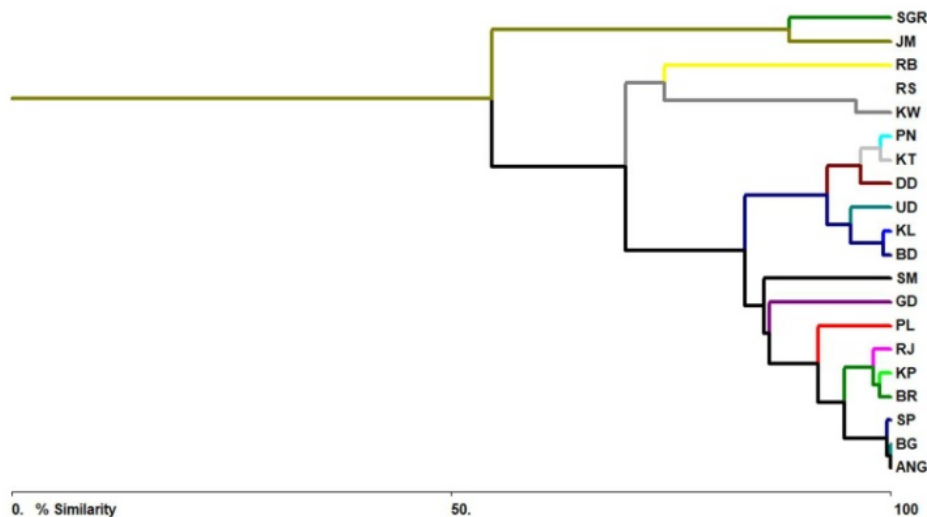
To show the increment of all the variables (cases), we plot the radar charts by analyzing the data set statistically using the PAST software (version 3). The radar charts are given in Figures 5-8. It is well known that these plots are easy to understand the values/increments of each variable.

### 3. RESULTS AND DISCUSSION

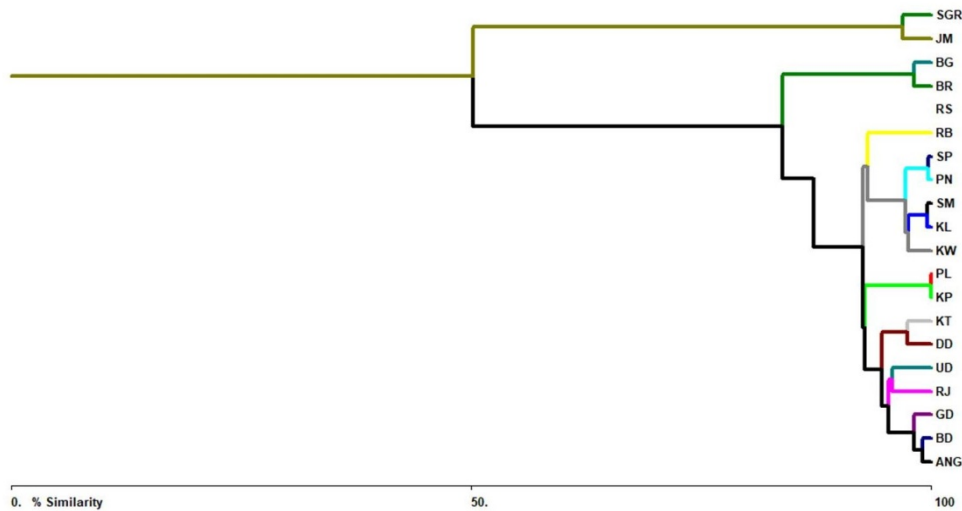
Results obtained from the current study suggested four dendrograms (Figures 1-4) for each variable (confirmed, active, cured and death cases). For the visual representation, these dendrograms of cluster analysis calculated separately from all the variables of the COVID-19 data set. For confirmed cases, districts like ANG, BD and GD; BG and BR; SGR and JM; SM and KW; PL and KP; KT and DD; UD and RJ; SP and PN. For active cases, districts like SGR and JM; RS and KW; PN and KT; KL and BD; KP and BR; BG and ANG. For cured cases, districts like SGR and JM; BG and BR; SP and PN; SM and KL; PL and KP; KT and DD; UD and RJ; BG and ANG. While as for death cases, districts like SGR and JM; RB and KW; SP and SM; RJ and KL; UD and BD; PL and KP forming clusters based on their similarity in COVID-19 cases as shown in Figures 1-4.



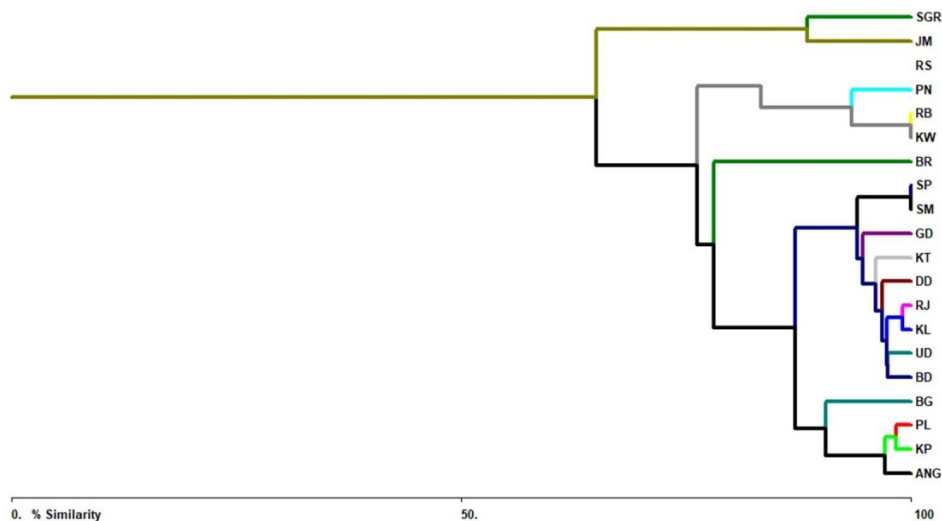
**Figure 1.** Dendrogram showing clustering of districts for confirmed cases of coronavirus disease (COVID-19).



**Figure 2.** Dendrogram showing clustering of districts for active cases of coronavirus disease (COVID-19).



**Figure 3.** Dendrogram showing clustering of districts for cured cases of coronavirus disease (COVID-19).



**Figure 4.** Dendrogram showing clustering of districts for death cases of coronavirus disease (COVID-19).

ANG - Anantnag; SGR - Srinagar; BD - Bandipora; BR - Baramulla; GD - Ganderbal; BG - Budgam; KL - Kulgam; PL - Pulwama; KP - Kupwara; SP - Shopian; JM - Jammu, RB - Ramban; RS - Reasi, UD - Udhampur; KT - Kathua; KW - Kishtwar; PN - Poonch; RJ - Rajouri; SM - Samba; DD - Doda.

All the districts of J&K have a high burden of confirmed cases. Districts like ANG, BR, BD, JM, KP, PL, RJ, SM, SGR and SP still have a good percentage of active cases. However, ANG, BP, BR, BD, JM, SGR, PL and KP have high rates of cured cases. While ANG, BP, BR, JM, SGR, PL and KP show a high rate of mortality. The trend shown in radar charts (Figures 5-8) for all the variables (confirmed, active, cured and death cases) related to COVID-19 were directly proportional; the districts with a high percentage of confirmed or active cases had a high number of cured as well as death cases.

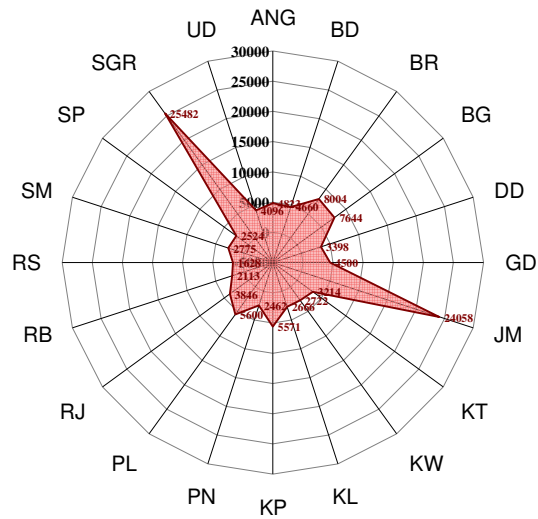


Figure 5. Radar chart showing the confirmed case of coronavirus disease (COVID-19).

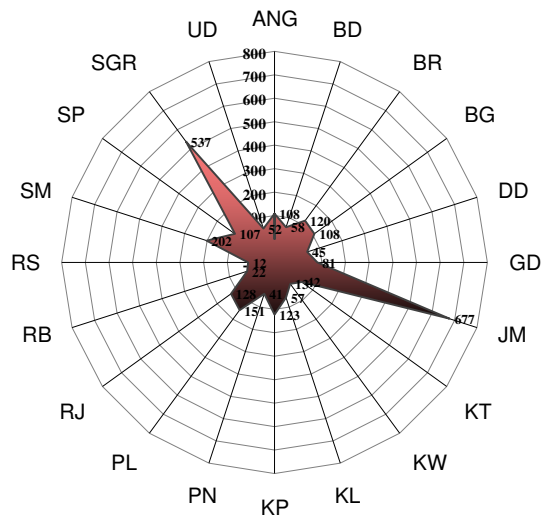


Figure 6. Radar chart showing active cases of coronavirus disease (COVID-19).

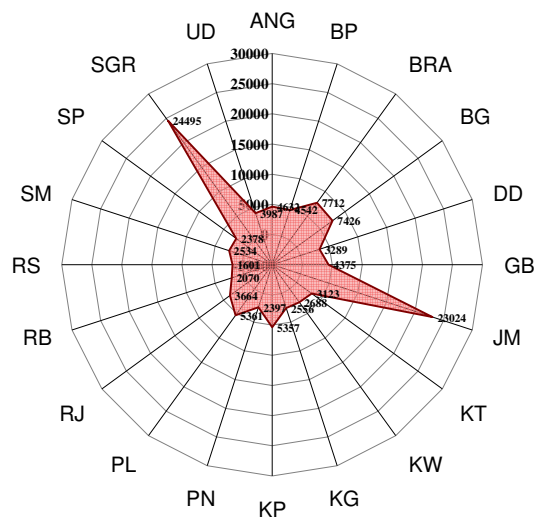


Figure 7. Radar chart showing cured cases of coronavirus disease (COVID-19).

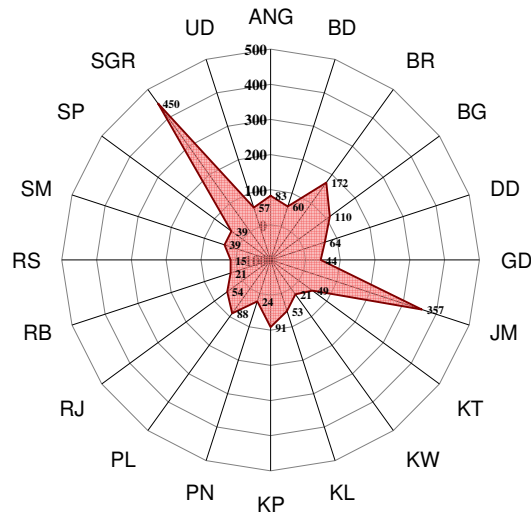


Figure 8. Radar chart showing death cases of coronavirus disease (COVID-19).

The radar chart (Figure 5) showed that all the districts had a good percentage of confirmed COVID-19 cases, but districts like SGR, JM, BR and BG were in the severe zone. Similarly, in other radar charts (Figures 6-8), districts like SGR, JM, BR, BG, SM, PL and RJ showed a high number of active cases. Similarly, districts like SGR, JM, RJ, PL and KP showed severity in death cases. Few studies are based unequivocally on Indian COVID-19 data. Kumar [25] has used cluster analyzing to monitor the novel COVID-19 cases in Maharashtra, India. Das [26] has used the epidemiological model to estimate the basic reproduction number at national and some state levels. Ray et al. [27] used a predictive model for case counts in India. Considering the great diversity in every aspect of India and its vast population, it would be a much better idea to monitor the COVID-19 cases at each of the states individually. It would help to decide further plans and actions to contain the spread of the disease, which can be crucial for the COVID-19 affected states

4. CONCLUSIONS

In this study, hierarchical (Brey-Curtis) cluster analysis was carried out to classify districts of Jammu and Kashmir based on similarity among COVID-19 cases to visually understand the impact of COVID-19. This technique grouped 20 different affected districts into four cluster and radar charts for each of the cases (variables). All the districts of J&K under clusters (Figure 1 & 5) were affected severely with COVID-19. The radar charts (Figures 5-8) showed the number of confirmed, active, cured and death cases, respectively. The trend in radar charts depicted a good percentage of cured cases in some districts: ANG, BP, BRA, BD, JM, SGR, PL and KP. It was also observed that the districts like SGR, JM, SM and PL have higher cases may need optimization of monitoring techniques which could help the government in making better policies and actions.

**Authors' Contributions:** This work was carried out in collaboration between the authors. DP reviewed and edited the first draft of manuscript. TI conceptualized, designed and managed the analysis of the study. JAM wrote the first draft of manuscript. AG edited the final manuscript. SAZ managed the literature searches. All authors approved the final version of the manuscript.

**Conflict of Interest:** The authors have no conflict of interest to declare.

**Acknowledgment:** We are thankful to all the colleagues who made valuable comments about the manuscript.

## REFERENCES

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020; 395(10223): 497-506.
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020; 382(8): 727-733.
3. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L. et al. Presumed asymptomatic carrier transmission of COVID-19. *Jama*. 2020; 323(14): 1406-1407.
4. Phan LT, Nguyen TV, Luong QC, Nguyen TV, Nguyen HT, Le HQ, et al. Importation and Human-to-Human Transmission of a Novel Coronavirus in Vietnam. *N Engl J Med*. 2020; 382(13): 1199-1207.
5. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *New Engl J Med*. 2020; 382(18): 1708-1720.
6. Giovanetti M, Benvenuto D, Angeletti S, Ciccozzi M. The first two cases of 2019-nCoV in Italy: Where they come from? *J Med Virol*. 2020; 92: 518-521.
7. World Health Organization. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). [https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)).
8. Rather RA, Islam T, Rehman IU, Pandey D. Development of vaccine against coronavirus disease-2019 (COVID-19) in India. *Asian J Adv Med Sci*. 2021; 3(2): 13-21.
9. Zumla A, Chan JF, Azhar EI, Hui DSC, Yuen KY. Coronaviruses - drug discovery and therapeutic options. *Nat Rev Drug Discov*. 2016; 15(5): 327-347.
10. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan. *China Lancet* 2020; 395(10223): 497-506.
11. Chan JF, Lau SK, Woo PC. The emerging novel Middle East respiratory syndrome coronavirus: the “knowns” and “unknowns”. *J Formos Med Assoc*. 2013; 112(7): 372-381.
12. Channappanavar R, Zhao J, Perlman S. T cell-mediated immune response to respiratory coronaviruses. *Immunol Res*. 2014; 59(1-3): 118-128.
13. Cheng VC, Lau SK, Woo PC, Yuen KY. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clin Microbiol Rev*. 2007; 20(4): 660-694.
14. Chan JF, Lau SK, To KK, Cheng VC, Woo PC, Yuen KY. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clin Microbiol Rev*. 2015; 28(2): 465-522.
15. Gretebeck LM, Subbarao K. Animal models for SARS and MERS coronaviruses. *Curr Opin Virol*. 2015; 13: 123-129.
16. Gralinski LE, Baric RS. Molecular pathology of emerging coronavirus infections. *J Pathol*. 2015; 235(2): 185-195.
17. Drosten C, Günther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med*. 2003; 348(20): 1967-1976.
18. Liu J, Zheng X, Tong Q, Li W, Wang B, Sutter K, et al. Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV. *J Med Virol*. 2020; 92(5): 491-494.



19. Zargar SA, Islam T, Rehman IU, Pandey D. Use of Cluster Analysis To Monitor Novel Corona Virus (Covid-19) Infections In India. *Asian J Adv Med Sci.* 2021; 3(2): 1-7.
20. COVID-19 Monitoring Dashboard maintained by Ministry of Health and Family Welfare Government of India (GoI). Available from <https://www.mohfw.gov.in/>. Accessed on 5 January 2021.
21. Dilts D, Khamalah J, Plotkin A. Using cluster analysis for medical resource decision making. *Med Decis Mak.* 1995; 15: 333-347.
22. McLachlan GJ. Cluster analysis and related techniques in medical research. *Stat Methods Med Res.* 1992; 1: 27-48.
23. Borgen FH, Barnett DC. Applying cluster analysis in counseling psychology research. *J Counsel Psychol.* 1987; 34(4): 456.
24. Johnson RA, Wichern DW. 2002. *Applied multivariate statistical analysis* (Vol. 5, No. 8). Upper Saddle River, NJ: Prentice hall.
25. Kumar S. Use of cluster analysis to monitor novel coronavirus-19 infections in Maharashtra, India. *Indian J Med Sci.* 2020; 72(2): 44.
26. Das S. Prediction of COVID-19 disease progression in India: Under the effect of national lockdown. 2020; <http://arxiv.org/abs/2004.03147>.
27. Ray D, Salvatore M, Bhattacharyya R, Wang L, Du J, Mohammed S, et al. Predictions, Role of Interventions and Effects of a Historic National Lockdown in India's Response to the the COVID-19 Pandemic: Data Science Call to Arms. *Harvard Data Sci Rev.* 2020; <https://doi.org/10.1162/99608f92.60e08ed5>.