

Changing Clusters of Indian States with respect to number of Cases of COVID-19 using incrementalKMN Method

Rabinder Kumar Prasad (✉ rkp@dibru.ac.in)

Dibrugarh University Institute of Engineering and Technology, Dibrugarh University, Assam ,786004, India

Rosy Sarmah

Tezpur University, Assam, 784028, India

Subrata Chakraborty

Dibrugarh University, Assam, 78004, India

Research Article

Keywords: COVID-19, clustering, growth rate, doubling time

Posted Date: October 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-91503/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Changing Clusters of Indian States with respect to number of Cases of COVID-19 using incrementalKMN Method

Rabinder Kumar Prasad · Rosy Sarmah ·
Subrata Chakraborty

Received: date / Accepted: date

Abstract The novel Coronavirus (COVID-19) incidence in India is currently experiencing exponential rise with apparent spatial variation in growth rate and doubling time. We classify the states into five clusters with low to high-risk category and identify how the different states moved from one cluster to the other since the onset of the first case on 30th January 2020 till the end of 15th September 2020. We cluster the Indian states into 5 groups using incrementalKMN clustering [20]. We observed and comment on the changing scenario of the formation of the clusters starting from before lockdown, through lockdown and the various unlock phases.

Keywords COVID-19 · clustering · growth rate · doubling time

1 Introduction

A severe acute respiratory disease, caused by a novel coronavirus, (COVID-19) started spreading in the month of November- December 2019 throughout China and received worldwide attention. The World Health Organization (WHO) officially declared the novel coronavirus (COVID-19) epidemic on 30th January 2020 as a public health emergency of international concern.

In India, the first case of novel coronavirus (COVID-19) was detected on 30th January in the State of Kerala [25]. As the number of confirmed novel coronavirus positive cases closed 500, the Govt. of India introduced “Janta Curfew” on 19th March 2020 and after that Govt. of India enforced 21 days lockdown Phase-I nationwide from 25th March – 14th April 2020 with nearly all services and factories were suspended [22].

As the number of confirmed cases of COVID-19 increased, on 14th April 2020, Govt. of India extended the lockdown Phase-I period till 3rd May 2020 i.e. lockdown Phase-II, with certain relaxations [6, 9]. In Phase-II, the lockdown area were categorised into *three* zone “red zone”, “orange zone”, and “green zone” based on

Department of CSE, Dibrugarh University, Assam, 786004, India,
E-mail: rkp@dibru.ac.in · Department of CSE, Tezpur University, Assam, 784028, India,
E-mail: rosy8@tezu.ernet.in · Department of Statistics, Dibrugarh University, Assam, 786004, India,
E-mail: subrata_stats@dibru.ac.in

number of confirmed cases [5].

On 1st May 2020, the Ministry of Home Affairs (MHA) and the Government of India (GoI) further extended the lockdown period i.e Phase-III to two weeks beyond 4th May based on the number of cases, with some relaxations [18,16].

Again, the National Disaster Management Authority (NDMA) and the Ministry of Home Affairs (MHA) extended the lockdown i.e. Phase-IV for a period of two weeks from 18th – 31st May 2020 on 17th May 2020 with additional relaxations. In this phase, the local bodies were given the authority to demarcate containment and buffer zones [4,8,17].

With a focus on the economy of India, the MHA issued fresh guidelines for the month of June 2020 based on containment zones in Phase-V. This phase is also known as “Unlock 1.0” [21].

The phase-VI of lockdown with limited restriction began on 1st July 2020 only in containment zones under the guidelines and instructions of the Ministry of Home Affairs (MHA) and the National Disaster Management Authority (NDMA). In all other areas of India till 31st July 2020, most of the activities were permitted except educational institutions, metros, recreational activities. This phase of lockdown is also known as “Unlock 2.0” [24].

Again, the National Disaster Management Authority (NDMA) and the Ministry of Home Affairs (MHA) extended the lockdown i.e. Phase-VII for a period for 1st August- 31st August 2020 with minimum restrictions. This phase of lockdown is known as “Unlock 3.0” [1,3].

On August 29, 2020, the Ministry of Home Affairs (MHA) has issued fresh guidelines for activities in lockdown Phase-VIII i.e “Unlock 4.0” [2]. In this phase, again the educational institutions have remained closed and metro rail, religious, entertainment, political, sports, academic functions and gatherings of up to 100 people were allowed [2].

The current surge in the number of COVID-19 cases has created a new social-demographic problem across India with a lot of fatalities. The machine learning community is continuously trying to predict the outcome of the disease in terms of number of cases, number of fatalities, etc. and has been quite successful at that. One can see [26,13,12,7,19,11] for different approaches in the literature. Cluster analysis can be effectively used to predict the number of cases of COVID-19. Cluster analysis is an important data analysis tool that searches patterns in a data set by grouping the data objects into clusters with respect to certain criteria. It plays a significant role in almost every area of science and engineering, like video processing, image processing, text analysis, bioinformatics, market research, privacy and security, wireless sensor networks, web social networks analysis, and document clustering, etc [10].

In this article, our main objective is to classify the states and UTs of India with respect to the daily incidence of positive cases of COVID-19 in five clusters designated as low to high-risk category. Then to study the dynamics of how the states and UTs changed clusters in phases over the period stretching from 1st case through the lockdowns, unlocks and up to 15th September 2020. For the process of clustering, we have used the Incremental k -means clustering (incrementalKMN) method [20] which is based on k -means [10] clustering method and provides improvement over the latter in the quality of clusters in term of minimizing the total Sum of Squared Error (SSE) [10].

2 Methodology

2.1 Clustering

Let us suppose, $\mathbf{D} = \{x_1, x_2, \dots, x_n\}$, be a data set with n number of data elements, and each data element characterized with m number of features : $x_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,m}\}$. The main objective of clustering is to group these data elements into homogeneous sub-groups such that the intra-cluster similarities are high while inter-cluster similarities are low. The data elements in each sub-group are called a cluster, and the union of all sub-groups is equal to the dataset \mathbf{D} . Clustering methods have been classified into five different categories, i.e., Partition-based, density-based, hierarchical-based, grid-based and model-Based [10]. In the last few decades, clustering algorithms have been extensively used to solve the problem of data-mining.

In this research, we have used incrementalKMN [20] clustering method on novel coronavirus (COVID-19) data set of India based on confirmed cases, which produced the k desired number of group of states of India i.e. the data set has been divided into k number of group of states. We apply the growth rate and doubling time from equation 1 and equation 2 on the k clusters produced by incrementalKMN [20] clustering method. The steps of incrementalKMN clustering is given below:

- Step 1: Select the value of k and dataset D
- Step 2: Set $i = 1$ and $C = \phi$, where C is the centroid list, which is initially empty.
- Step 3: Select first centroid i.e. c_i as mean of a given dataset D .
- Step 4: Update the centroid list $C = C \cup \{c_i\}$.
- Step 5: Assign each data objects to its nearest centroid.
- Step 6: Compute the SSE of each cluster.
- Step 7: Select i^{th} center i.e. ($i \in 2, 3, \dots, k$) is selected from maximum SSE of a cluster. The i^{th} center is a maximum distance from the data object and the centroid of maximum SSE cluster.
- Step 8: Repeat step 4 until it reaches k number of centroids and from the k number of clusters.

2.2 Compound Growth Rate [15]:

We have used compound growth rate over regular time intervals of confirmed case for each state phase wise. The growth rate of each state is computed as:

$$Growth_{rate} = \left(\frac{Present_Confirmed_Case}{Past_Confirmed_Case} \right)^{\left(\frac{1}{n}\right)} - 1 \quad (1)$$

where, $Present_Confirmed_Case$ is the number of confirmed positive cases on present date, $Past_Confirmed_Case$ is a number of confirmed positive cases on start date and n is a the number of days.

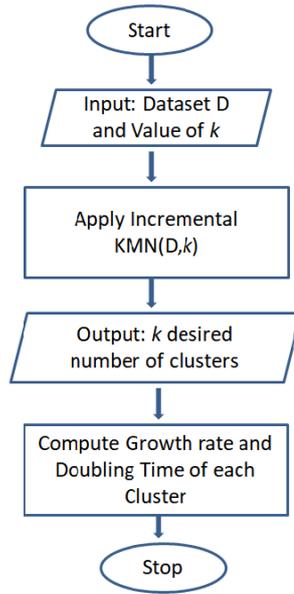


Fig. 1: The flowchart of Proposed Method

2.3 Doubling Time [14]:

Doubling time is the time it takes for a confirmed case to double in size. The doubling time of confirmed case for each cluster is computed with the help of equation 1, which is described as:

$$Doubling_{Time} = \frac{\ln(2)}{\ln(1 + Growth_{rate})} \quad (2)$$

where, \ln stand for natural logarithm.

The complete flowchart of the methodology adopted is given in Fig 1.

In Fig 1, the proposed method start with input data set D and value of k , which we have considered as 5. In the next step, we have applied the incrementalKMN clustering method to get the desired k number of clusters. In the next phase of the flowchart, it computes the growth rate and doubling time for each of the k clusters produced by the incrementalKMN clustering method.

3 Data Source

The data set of novel coronavirus (COVID-19) daily confirmed cases state-wise is collected from <https://api.covid19india.org/>. In this data set, the total number of confirmed cases in India was 50,14094 from 30th January to 15th September 2020. In novel coronavirus dataset, we have considered the following states and UTs, as AN (Andaman and Nicobar Islands), AP (Andhra Pradesh), AR (Arunachal Pradesh), AS (Assam), BR (Bihar), CH (Chandigarh), CT (Chhattisgarh), DN (Dadra and Nagar Haveli and Daman and Diu), DL(Delhi), GA (Goa),

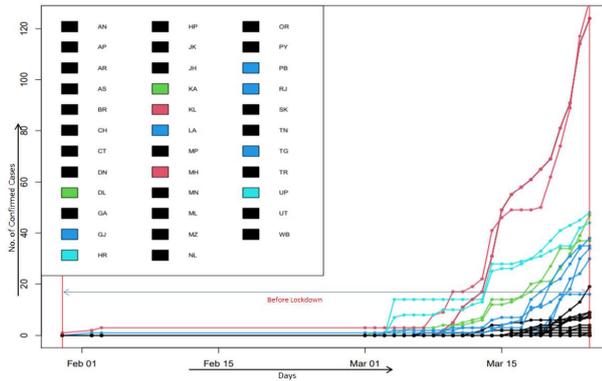


Fig. 2: Result of COVID-19 dataset before lockdown using IncrementalKMN clustering method($k=5$)

GJ (Gujarat), HR (Haryana), HP (Himachal Pradesh), JK (Jammu and Kashmir), JH (Jharkhand), KA (Karnataka), KL (Kerala), LA (Ladakh), MP (Madhya Pradesh), MH (Maharashtra), MN (Manipur), ML (Meghalaya), MZ (Mizoram), NL (Nagaland), OR (Odisha), PY (Puducherry), PB (Punjab), RJ (Rajasthan), SK (Sikkim), TN (Tamil Nadu), TG (Telangana), TR (Tripura), UP (Uttar Pradesh), UT (Uttarakhand), WB (West Bengal).

4 Result and Discussion

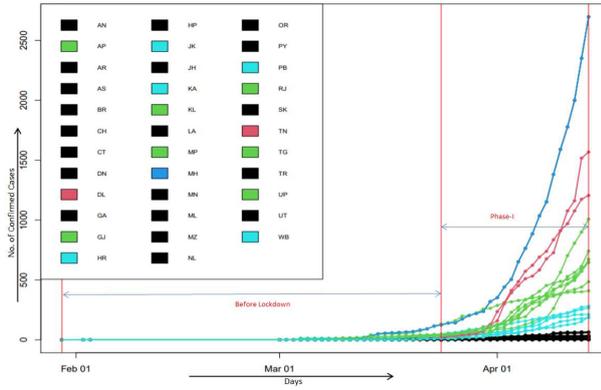
We have considered five different groups of states namely (i) high-risk, (ii) moderate-high, (iii) moderate, (iv) moderate-low, and (v) low-risk states with respect to the daily incidence of COVID-19 positive confirmed cases. The incrementalKMN clustering was used over the Indian COVID-19 dataset with $k = 5$. The first cluster with highest number of confirmed cases is therefore a high-risk group. The second cluster has the second highest confirmed cases and is a moderate-high group and so on. The fifth cluster has the lowest number of confirmed cases and is a low-risk group. The detailed phase-wise analysis is narrated in the next 6 subsections.

4.1 Scenario of novel coronavirus (COVID-19) in India Before Lockdown:

The first confirmed case of novel coronavirus in India was reported on 30th January 2020 in the state of Kerala [25]. The number of confirmed COVID-19 positive cases was approximately 500 on 19th March 2020 and by 24th March 2020 that is before lockdown the number of positive cases reached 658. In Fig 2 which depicts the situation prior to the start of the lockdown based on the number of confirmed cases, the states {MH and KL} were in the high-risk state, {HR and UP} were in moderate-high risk state, Union Territory DL and the state KA were in moderate risk state, {GJ, LA, PB, RJ, TG} were in moderate-low risk states and the remaining states and UTs were on the low-risk states.

Table 1: Results on Novel coronavirus (COVID-19) before lockdown Dataset in India

<i>Cluster_ID</i>	<i>Before Lockdown</i>		<i>List of States and UTs</i>
	Growth Rate (Approx)	Doubling Time (days)	
I	23%	3.35	KL,MH
II	22%	3.48	DL,KA
III	26%	3.00	GJ,LA,PB,RJ,TG
IV	8%	9.01	HR,UP
V	7%	10.25	AN,AP,AR,AS,BR, CH,CT,DN,GA,HP, JK,JH,MP,MN,ML, MZ,NL,OR,PY,SK, TN,TR,UT,WB

Fig. 3: Result of COVID-19 Lockdown Phase-I dataset using IncrementalKMN clustering method ($k=5$)

In Table 1, the growth rate and doubling time (in days) of each cluster of states are shown. The growth rate of clusters I, II, and III states i.e. {GJ, LA, PB, RJ, TG}, {KL, MH}, {DL, KA} were approximately same and doubling time was 3-4 days (approximately). Whereas, the growth rate of cluster IV i.e. the states {HR and UP} were much less than the top 3 clusters but close to that of cluster V, and doubling time of 9 days (approximately) was nearly 3 times that of the top clusters. For the rest of the states and UTs before lockdown, the growth rate was low, and the doubling time was approximately 10-11 days.

4.2 Scenario of Novel coronavirus (COVID-19) in India Lockdown Phase-I:

The first lockdown in India started on 25th March 2020, when nearly all services and factories were suspended and continued till 14th April 2020 [22]. In Fig 3 based on the number of confirmed cases, the state MH was still in the high-risk category state, the Union Territory DL and the state TN were in moderate-high risk state. Whereas the states {AP, GJ, KL, MP, RJ, TG, UP} were in moderate

and {HR, JK, KA, PB, WB} were in moderate-low risk category and followed by the following states and {UTs AN, AR, AS, BR, CH, CT, DN, GA, HP, JK, JH, LA, MN, ML, MZ, NL, OR, PY, SK, TR, UT} in the low-risk category of states.

Table 2: Results on Novel coronavirus (COVID-19) Phase-I Dataset in India

Cluster Id	Before Lockdown		Phase-I		States and UTs
	Growth Rate (Approx)	Doubling Time	Growth Rate (Approx)	Doubling Time	
I	29%	2.72	15%	4.96	MH
II	17%	4.41	19%	3.98	DL, TN
III	22%	3.48	15%	4.96	AP, GJ, KL, MP, RJ, TG, UP
IV	20%	3.80	11%	6.64	HR, JK, KA, PB, WB
V	4%	17.67	8%	9.01	AN, AR, AS, BR, CH, CT, DN, GA, HP, JH, LA, MN, ML, MZ, NL, OR, PY, SK, TR, UT

The growth rate and doubling time for all the states in different clusters are shown in Table 2. Accordingly, the growth rate of the state of MH before the lockdown was high and in lockdown phase-I, the growth rate decreased considerably from 29% to 15%. Similarly, the 3 days doubling time before the lock down climbed up to 5 days after lockdown phase-I. On the contrary, the average growth rate of cluster II i.e. {DL, TN} has increased, and doubling time in lockdown phase-I was decreased as compared to before lockdown. Similarly, the average growth rate and doubling time of cluster III and IV i.e. states {AP, GJ, KL, MP, RJ, TG, UP} and states {HR, JK, KA, PB, WB} have improved in lockdown phase-I. But in cluster V i.e. states and UTs {AN, AR, AS, BR, CH, CT, DN, GA, HP, JH, LA, MN, ML, MZ, NL, OR, PY, SK, TR, UT}, the average growth rate has increased and doubling time also decreased in lockdown phase-I as compared to before lockdown.

4.3 Scenario of Novel coronavirus (COVID-19) in India Lock down Phase-II:

In phase-II, the lockdown was extended nationwide up to 3rd May 2020 with certain relaxations [6,9]. In Fig 4 based on the number of confirmed cases, the state MH was again sitting in the high-risk category of states and the Union Territory DL and state GJ were in moderate-high risk state. Similarly, the states {MP, RJ, TN, UP} were in moderate risk state and {AP, JK, KA, KL, TG, WB} were in moderate-low risk state, and the rest of the states and UTs were in the low-risk state.

Table 3 shows the growth of the rate and doubling time of each cluster of states and UTs. The growth rate for both the cluster I that is {MH} and cluster II i.e. {DL and GJ} reported as 8% is a substantial decrease from their 15% and 17% in lockdown phase-I respectively. Similarly, the doubling time of approximately 9 days for these two clusters have nearly doubled as compared to lock down phase-I. Similarly, the growth rate of cluster III comprising {MP, RJ, TN, UP} has decreased from 17% in lockdown phase -I to 6% (approximately) in lockdown

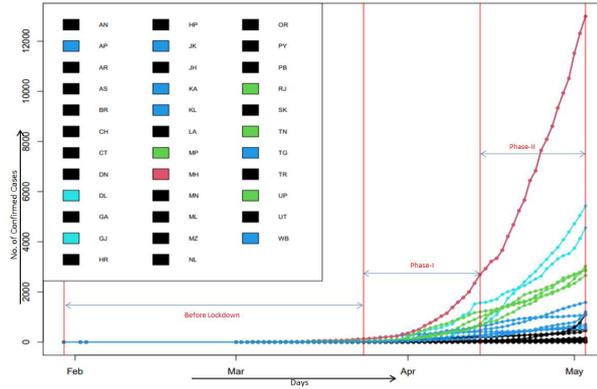


Fig. 4: Result of COVID-19 Lockdown Phase-II data set using IncrementalKMN clustering method ($k=5$)

Table 3: Results on Novel coronavirus (COVID-19) Phase-II Dataset in India

<i>Cluster Id</i>	<i>Before Lockdown</i>		<i>Phase-I</i>		<i>Phase-II</i>		<i>States and UTs</i>
	Growth Rate (Approx)	Doubling Time	Growth Rate (Approx)	Doubling Time	Growth Rate (Approx)	Doubling Time	
I	29%	2.72	15%	4.96	8%	9.01	MH
II	39%	2.10	17%	4.41	8%	9.01	DL,GJ
III	14%	5.29	17%	4.41	6%	11.89	MP,RJ,TN,UP
IV	21%	3.64	13%	5.67	5%	14.21	AP,JK,KA,KL,TG,WB
V	5%	14.21	8%	9.01	4%	17.67	AN,AR,AS,BR,CH,CT,DN,GA,HR,HP,JH,LA,MN,ML,MZ,NL,OR,PY,PB,SK,TR,UT

phase-II, and doubling time has gone from 5 days to 12 days (approximately). The growth rate of cluster IV i.e., {AP, JK, KA, KL, TG, WB} has decreased from 13% (approximately) in lockdown phase-I to 5% in lockdown phase-II, and doubling time has increased from 6 days to 14 days (approximately). Finally, the growth rate of cluster V i.e. {AN, AR, AS, BR, CH, CT, DN, GA, HR, HP, JH, LA, MN, ML, MZ, NL, OR, PY, PB, SK, TR, UT} has improved in fact nearly half as compared to lock down phase-I and the doubling time of 18 days (approximately) was nearly double as compared to 9 days lockdown phase-I.

4.4 Scenario of Novel coronavirus (COVID-19) in India Lockdown Phase-III:

Third phase of lockdown started from 4th–17th May 2020 with some more relaxations [18, 16]. The country was divided into 3 zones: red zones, orange zones, and green zones [23]. In Fig 5, again the state MH was in a high-risk category, and DL, GJ, and TN formed the moderate-high risk category. The states {MP, RJ,

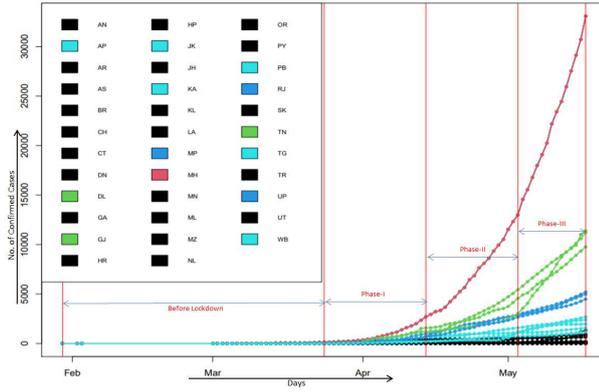


Fig. 5: Result of COVID-19 Lockdown Phase-III dataset using IncrementalKMN clustering method ($k=5$)

UP} were in moderate risk, {AP, JK, KA, PB, TG, WB} were in moderate-low risk state, and the rest of the states and UTs were in the low-risk category.

Table 4: Results on Novel coronavirus (COVID-19) Phase-III Dataset in India

Cluster Id	Before Lockdown		Phase-I		Phase-II		Phase-III		States and UTs
	Growth Rate (Approx)	Doubling Time							
I	29%	2.72	15%	4.96	8%	9.01	6%	11.89	MH
II	32%	2.50	18%	4.19	7%	10.24	6%	11.89	DL,GJ,TN
III	13%	5.67	17%	4.41	6%	11.89	4%	17.67	MP,RJ,UP
IV	22%	3.48	14%	5.29	6%	11.89	4%	17.67	AP,JK,KA, PB,TG,WB
V	5%	14.21	8%	9.01	3%	23.45	4%	17.67	AN,AR,AS, BR,CH,CT, DN,GA,HR, HP,JH,KL, LA,MN,ML, MZ,NL,OR, PY,SK,TR, UT

A significant observation from Table 4 is that in phase-III the growth rate of confirmed cases of all clusters except cluster V due to lockdown have decreased and the doubling time of all clusters increased except for cluster V. The doubling time of cluster I and II i.e. {MH} and {DL, GJ, TN} have increased to 12 days (approximately). Similarly, the doubling time of cluster III, IV and V i.e. the states and UTs {MP, RJ, UP}, {AP, JK, KA, PB, TG, WB} and {AN, AR, AS, BR, CH, CT, DN, GA, HR, HP, JH, KL, LA, MN, ML, MZ, NL, OR, PY, SK, TR, UT} have gone down up to 18 days approximately.

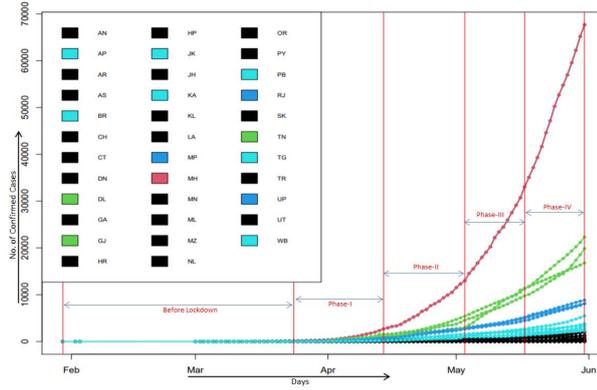


Fig. 6: Result of COVID-19 Lockdown Phase-IV dataset using IncrementalKMN clustering method ($k=5$)

4.5 Scenario of Novel coronavirus (COVID-19) in India Lockdown Phase-IV:

In phase-IV, the lockdown was extended for another two weeks from 17th – 31th May 2020 with some additional relaxations. Here, red zones were further divided into containment and buffer zones [4,8,17]. In Fig 6, the state MH was remained in the high-risk state based on the number of confirmed cases, and {DL, GJ, TN} were in moderate-high risk state. The states {MP, RJ} were in a moderate-risk group, {AP, BR, JK, KA, PB, TG, WB} were in moderate-low risk, and the rest of the states were in the low-risk state.

The growth rate of clusters I and II decreased marginally while that of II and IV remained almost the same as in the previous phases of lockdown. Now the doubling time of cluster I i.e. the state MH is 14 days (approximately) and cluster II, III, and IV i.e. the states and UTs {DL, GJ, TN}, {MP, RJ, UP}, and {AP, BR, JK, KA, PB, TG, WB} are all 18 days (approximately). But for cluster V the growth rate more than doubled from phase II. This may be attributed to the home coming of migrants from the red zone. Similarly, the situation is reflected with the doubling time as well. Except for the cluster V, all clusters in lockdown phase-IV has increased their doubling time. For cluster V, i.e. the states and UTs {AN, AR, AS, CH, CT, DN, GA, HR, HP, JH, KL, LA, MN, ML, MZ, NL, OR, PY, SK, TR, UT} doubling time deteriorated from 18 days to 8 days in Table 5.

4.6 Scenario of Novel coronavirus (COVID-19) in India Lockdown Phase-V:

The phase V (or Unlock-I) of lockdown started from 1st – 30th June 2020 with only limited restrictions [21]. In our study, we have considered the novel coronavirus (COVID-19) data set till 20th June 2020. The state MH was in the high-risk state based on the number of confirmed cases and {DL, TN} was in moderate-high risk state which is shown in Fig 7. Similarly, the state GJ was in moderate risk state, and {MP, RJ, UP, WB} were in moderate-low risk state, and the rest of the states and UTs were in the low-risk state.

Table 6 shows the growth rate of all clusters has decreased or remained the same

Table 5: Results on Results on Novel coronavirus (COVID-19) Phase-IV Dataset in India

Cluster Id	Before Lockdown		Phase-I		Phase-II		Phase-III		Phase-IV		States and UTs
	Growth Rate (Approx)	Doub-ling Time									
I	29%	2.72	15%	4.95	8%	9.01	6%	11.89	5%	14.21	MH
II	32%	2.50	18%	4.19	7%	10.24	6%	11.89	4%	17.67	DL,GJ,TN
III	13%	5.67	17%	4.414	6%	11.89	4%	17.67	4%	17.67	MP,RJ,UP
IV	21%	3.63	14%	5.29	7%	10.24	4%	17.67	4%	17.67	AP,BR,JK,KA, PB, TG, WB
V	5%	14.21	8%	9.01	3%	23.45	4%	17.67	9%	8.04	AN,AR,AS,CH, CT, DN, GA, HR, HP, JH, KL, LA, MN, ML, MZ, NL, OR, PY, SK, TR, UT

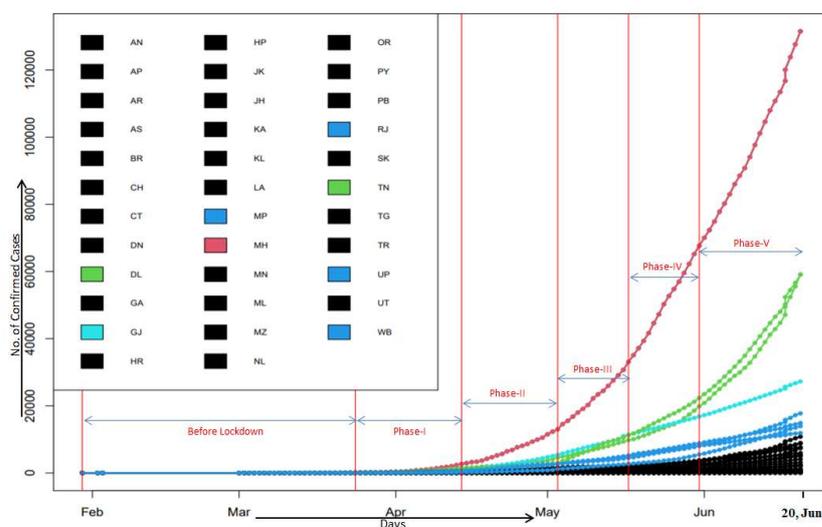


Fig. 7: Result of COVID-19 Lockdown Phase-V dataset till 20th June 2020 using IncrementalKMN clustering method ($k=5$)

in lockdown phase-V as compared to previous phases. Similarly, the doubling time of all clusters has increased or remained the same. The doubling time of cluster I i.e. the state MH is in 24 days (approximately). Similarly, cluster II i.e. the states {DL, TN} required 14-15 days (approximately) to double and cluster III i.e. the state GJ is required 35 days to double. The cluster IV and V have 24 days (approximately) and 9 days (approximately) of doubling time.

From what we have found it was expected that some of the states/UTs lying in the Category V will soon move to the Category IV in the next week or so starting 21th June 2020.

In order to verify this, we have then extended our study covering data up to 30th June 2020 to cover the full Unlock I period to see how the cluster changed in the last 10 days of this phase. The result is shown in Fig 9. As expected some of the states {AP, AS, BR, HR, JK, KA, OR, PB, TG} which were in Category V moved to category IV. In fact from Fig 8 where we have demarcated 6 instead of

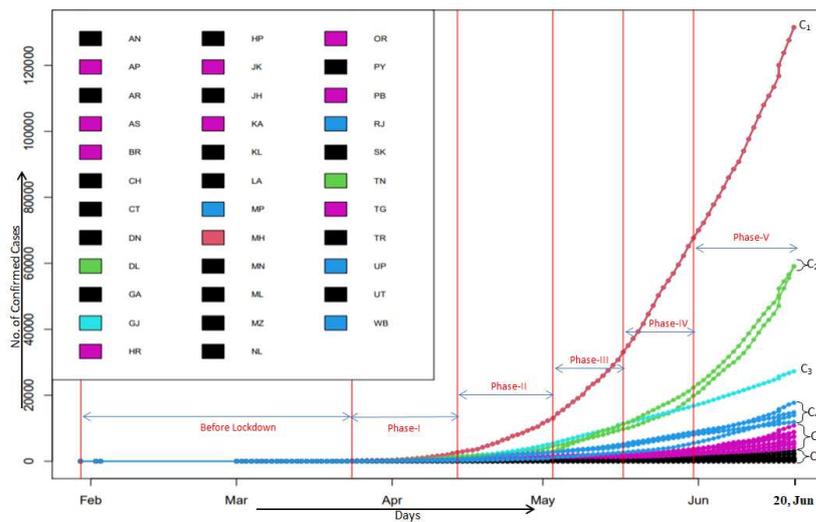


Fig. 8: Result of COVID-19 Lockdown Phase-V dataset till 20th June 2020 using IncrementalKMN clustering method ($k=6$)

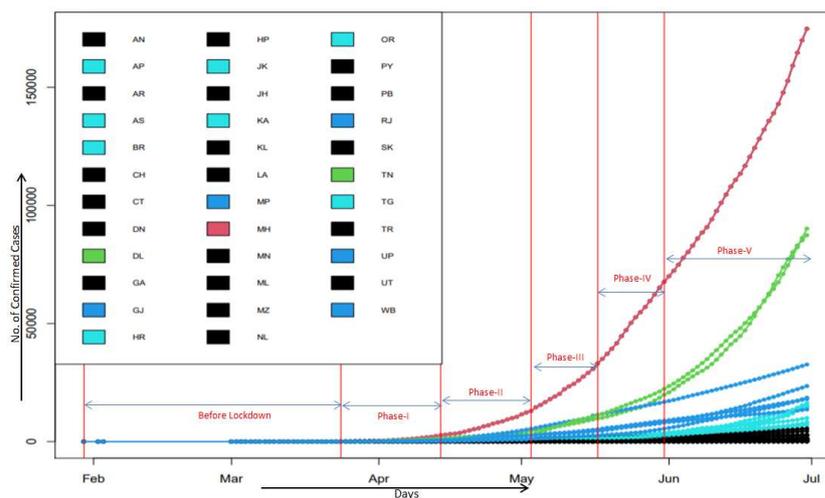


Fig. 9: Result of COVID-19 Lockdown Phase-V dataset using IncrementalKMN clustering method ($k=5$)

5 clusters to reveal hidden groups with the cluster V gave a clear indication of the tendency of the above states to move towards the next higher risk category (see also Table 7).

Table 6: Results on Novel coronavirus (COVID-19) Phase-V Dataset till 20, June 2020 in India

Cluster Id	Before Lockdown		Phase-I		Phase-II		Phase-III		Phase-IV		Phase-V till 20, June		States and UTs
	Growth Rate (Approx)	Doubling Time	Growth Rate (Approx)	Doubling Time									
I	29%	2.72	15%	4.96	8%	9.01	6%	11.89	5%	14.21	3%	23.45	MH
II	17%	4.41	19%	3.98	5%	14.21	7%	10.24	5%	14.21	5%	14.21	DL,TN
III	60%	1.47	14%	5.29	11%	6.64	5%	14.21	3%	23.45	2%	35.00	GJ
IV	18%	4.19	16%	4.67	7%	10.24	4%	17.67	4%	17.67	3%	23.45	MP,RJ,UP,WB
V	8%	9.01	9%	8.04	4%	17.67	4%	17.67	8%	9.01	8%	9.01	AN,AP,AR,AS,BR,CH,CT,DN,GA,HR,HP,JK,JH,KA,KL,LA,MN,ML,MZ,NL,OR,PB,PY,SK,TG,TR,UT

Table 7: Results on Novel coronavirus (COVID-19) Phase-V Dataset in India

Cluster Id	Before Lockdown		Phase-I		Phase-II		Phase-III		Phase-IV		Phase-V		States and UTs
	Growth Rate (Approx)	Doubling Time											
I	29%	2.72	15%	4.96	8%	9.01	6%	11.89	5%	14.21	3%	23.45	MH
II	17%	4.41	19%	3.98	5%	14.21	7%	10.24	5%	14.21	5%	14.21	DL,TN
III	26%	3.00	16%	4.67	8%	9.01	4%	17.67	4%	17.67	3%	23.45	GJ,MP,RJ,UP,WB
IV	12%	6.12	15%	4.96	5%	14.21	5%	14.21	7%	10.24	5%	14.21	AP,AS,BR,HR,JK,KA,OR,TG
V	6%	11.89	6%	11.89	3%	23.45	3%	23.45	9%	8.04	7%	10.24	AN,AR,CH,CT,DN,GA,HP,JH,KL,LA,MN,ML,MZ,NL,PY,PB,SK,TR,UT

4.7 Summary of Findings of data till June 2020

- i. From the high-risk category in the pre lockdown period KL (Kerala) moved to the low-risk category while MH (Maharashtra) remained in the high-risk category throughout the period. Whether this can be interpreted as successful implementing of lockdown rules in KL is arguable.
- ii. Lockdown seemingly had its impact on two other states HR (Haryana) and UP (Uttar Pradesh) as these states improved and went down to low risk and moderate low-risk category.
- iii. We observed that some of the states like AP (Andhra Pradesh), AS (Assam), BR (Bihar), HR (Haryana), JK (Jammu and Kashmir), KA (Karnataka), OR (Odisha), PB (Punjab), and TG (Telangana) has been moving up the ladder towards higher risk level in next phases of lockdown.
- iv. Also it appears that some of the states/UTs where the onset was late might show surge in coming days and weeks to move to a higher category of risk.

In next section, we extended our study to covering data upto 15th September 2020 to examine the above observations to justify our approach.

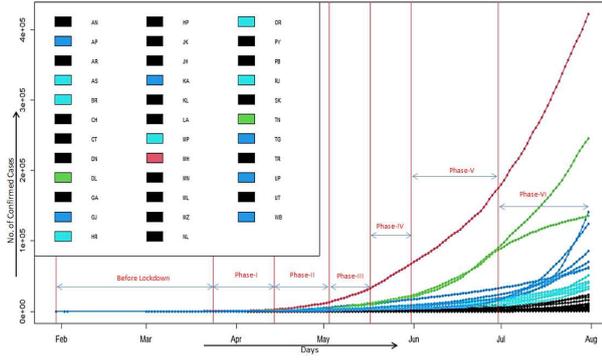


Fig. 10: Result of COVID-19 Lockdown Phase-VI dataset using IncrementalKMN clustering method ($k=5$)

5 Changing Clusters Scenario of Novel coronavirus (COVID-19) in India from 1st July to 15th September 2020

5.1 Scenario of Novel coronavirus (COVID-19) in India Lockdown Phase-VI:

Sixth phase of lockdown (or Unlock-2.0) started from 1st–31st July 2020 with some more relaxations as compared to Unlock-1.0. In this phase of lockdown, most of the activities were permitted except educational institutions, metros, recreational activities. In Fig 10, the state MH was in a high-risk category, and {DL, TN} formed the moderate-high risk category and DL will soon move to the next lower risk category in next week or starting 1st August 2020. The states {AP, GJ, KA, TG, UP, WB} were in moderate risk, {AS, BR, HR, MP, OR, RJ} were in moderate-low risk state, and the rest of the states and UTs were in the low-risk category.

Table 8 shows the growth rate and doubling time of each cluster of states and UTs. As compared to Phase-V, The growth rate and doubling time of cluster I and cluster IV that is {MH} and {AS, BR, HR, MP, OR, RJ} remained same in lockdown phase-VI. Cluster II i.e. {DL and TN} reported as in approximately 2% is a substantial decrease from their 5% (approximately) phase-V respectively. Similarly, the doubling time of approximately 31 days nearly doubled as compared to lock down phase-V. Similarly, the growth rate of cluster III comprising {AP, GJ, KA, TG, UP, WB} has increased from 4% (approximately) in lockdown phase-V to 5% (approximately) in lockdown phase-VI, and doubling time has gone from 17 days to 15 days (approximately). Finally, the growth rate of cluster V i.e. {AN, AR, CH, CT, DN, GA, HP, JK, JH, KL, LA, MN, ML, MZ, NL, PY, PB, SK, TR, UT} has improved as compared to lockdown phase-V and the doubling time required 16-17 days (approximately).

5.2 Scenario of Novel coronavirus (COVID-19) in India Lockdown Phase-VII:

In Phase-VII or (Unloack 3.0), the lockdown was extended for another one month from 1st – 31st August 2020 with limited restrictions. The state MH was in a high-

Table 8: Results on Novel coronavirus(COVID-19) Phase-VI (Unlock 2.0) Dataset in India

Cluster Id	Phase-V		Phase-VI		States and UTs
	Growth Rate (Approx)	Doubling Time	Growth Rate (Approx)	Doubling Time	
I	3.10%	22.74	2.78%	25.26	MH
II	4.74%	14.97	2.24%	31.24	DL,TN
III	4.25%	16.65	4.80%	14.77	AP,GJ,KA,TG, UP,WB.
IV	3.90%	18.14	3.90%	18.12	AS,BR,HR,MP, OR,RJ.
V	7.03%	10.20	4.40%	16.11	AN,AR,CH,CT, DN,GA,HP,JK, JH,KL,LA,MN, ML,MZ,NL,PY, PB,SK,TR,UT.

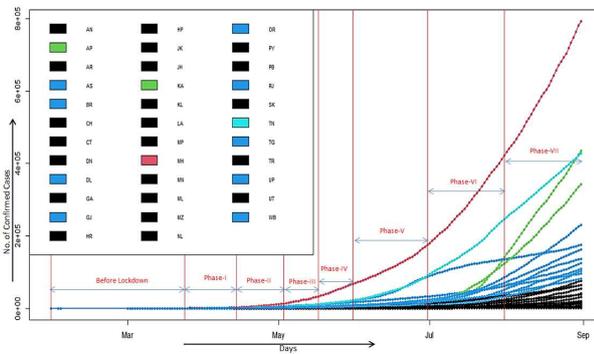


Fig. 11: Result of COVID-19 Lockdown Phase-VII dataset using IncrementalKMN clustering method ($k=5$)

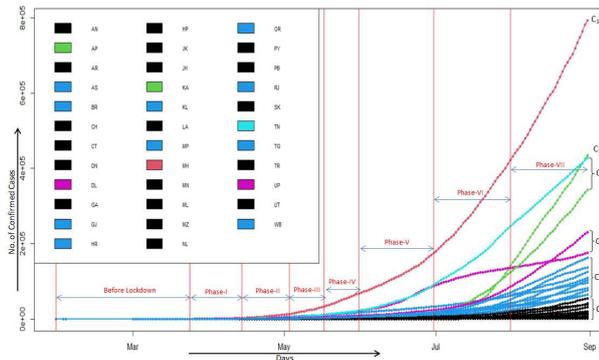


Fig. 12: Result of COVID-19 Lockdown Phase-VII dataset using IncrementalKMN clustering method ($k=6$)

Table 9: Results on Novel coronavirus (COVID-19) Phase-VIII (Unlock 3.0) Dataset in India

Cluster Id	Phase-V		Phase-VI		Phase-VII		States and UTs
	Growth Rate (Approx)	Doubling Time	Growth Rate (Approx)	Doubling Time	Growth Rate (Approx)	Doubling Time	
I	3.10%	22.74	2.78%	25.26	1.98%	35.37	MH
II	4.58%	15.46	3.15%	22.36	1.73%	40.48	TN
III	4.91%	14.46	7.08%	10.13	3.34%	21.1	AP,KA
IV	4.01%	17.65	3.77%	18.74	2.45%	28.69	AS,BR,DL,GJ,OR, RJ,TG,UP,WB.
V	6.75%	10.61	4.25%	16.67	3.16 %	22.27	AN,AR,CH,CT,DN, GA,HR,HP,JK,JH, KL,LA,MP,MN,ML, MZ,NL,PY,PB,SK, TR,UT.

risk category, and TN formed the moderate-high risk category and {AP, KA} was in moderate risk state which is shown in Fig 11. The states and UTs {AS, BR, DL, GJ, OR, RJ, TG, UP, WB} were in moderate-low risk state, and the rest of the states and UTs were in the low-risk category.

From Table 9, the growth rate and doubling time of confirmed cases of all clusters in Phase-VII has been improved as compared to Phase-VI i.e. "Unlock-2.0".

We have observed from above that some of the states/UTs lying in the Category III will soon merge to the Category II and some of the states/UTs lying in the Category IV will soon move to Category III in the next week or so starting 1st September 2020.

Again, in order to examine this, we have extended our study covering data up to 15th September 2020 to cover the Unlock 4.0 period i.e. Phase-VIII to inspect how the cluster changed in the last 15 days of this phase. The result is shown in Fig 13. As we have expected that the some of the states and UT { DL, UP} which were in Category IV moved up to category III. In fact from Fig 12 where we have 6 instead of 5 clusters to demonstrate hidden groups with the cluster IV gave a clear signal of the tendency of the above state towards the next higher risk category (see also Table 10).

6 Conclusion

In this study, we used incrementalKMN [20] clustering method to classify the Indian states and UTs in *five* different stages of risk on the basis of the number of confirmed cases of novel coronavirus (COVID-19). We then evaluated the growth rate and doubling time of confirmed cases of each cluster. As on 30th June 2020, the state MH (Maharashtra) was on high-risk category with the doubling time of confirmed cases as 23-24 days (approximately). Similarly, the union territory of DL (Delhi) and the state TN (Tamil Nadu) are in moderate-high risk state and doubling time of these set of states is 14-15 days (approximate). The state GJ (Gujarat) is in moderate risk state and has decreased their growth rate during

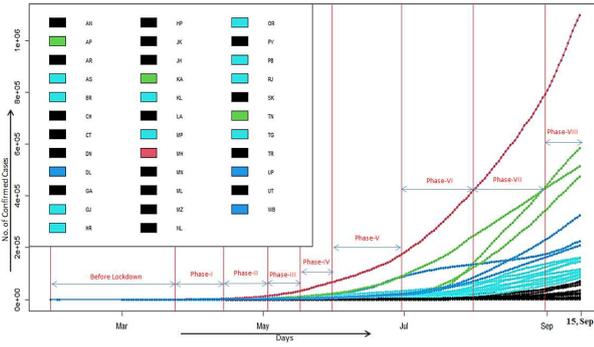


Fig. 13: Result of COVID-19 Lockdown Phase-VIII dataset till 15th September 2020 using IncrementalKMN clustering method ($k=5$)

Table 10: Results on Novel coronavirus (COVID-19) Phase-VIII (Unlock 4.0) Dataset in India, till 15th September 2020

Cluster Id	Phase-V		Phase-VI		Phase-VII		Phase-VIII		States and UTs
	Growth Rate (Approx)	Doubling Time							
I	3.10%	22.74	2.78%	25.26	1.98%	35.37	2.06%	33.98	MH
II	4.80%	14.78	5.77%	12.36	2.80%	25.08	1.67%	41.98	AP,KA,TN
III	4.12%	17.16	3.26%	21.59	2.18%	32.14	1.78%	39.28	DL,UP,WB
IV	3.64%	19.39	3.72%	18.99	2.65%	26.46	2.05%	34.19	BR,GJ,HR,KL,MP,OR,PB,RJ,TG
V	7.34%	9.78	4.43%	16.00	3.22%	21.89	2.57%	27.30	AN,AR,AS,CH,CT, DN,GA,HP,JK,JH, LA,MN,ML,MZ,NL, PY,SK,TR,UT

the lockdown phases and the doubling time of this state is 35 days (approximate). Moreover, the states MP (Madhya Pradesh), RJ (Rajasthan), UP (Uttar Pradesh), and WB (West Bengal) are in moderate-low risk state and the growth rate and doubling time are same as the state of MH. Rest of the states are in the low-risk cluster. Based on our observation, we expected some of the states namely AP (Andhra Pradesh), AS (Assam), BR (Bihar), HR (Haryana), JK (Jammu and Kashmir), KA (Karnataka), OR (Odisha), PB (Punjab), and the TG (Telangana) to move next level of risk and it shows in later phases of lockdown.

Based on our findings are extended our study up to 15th September 2020 to cover the Unlock 4.0 period and found the changing clusters as expected. As on 15th September, 2020 i.e. Unlock 4.0, again the state MH (Maharashtra) is on high-risk category and the growth rate and doubling time are improved. Similarly, the state TN (Tamil Nadu), AP (Andhra Pradesh) and KA (Karnataka) are in moderate-high risk state. Based on our observation, the state/UTs DL (Delhi), UP (Uttar Pradesh) and WB (West Bengal) are in moderate risk sate and has decreased their growth rate and increased their doubling time during the unlock phase. Moreover, the states AS (Assam), BR (Bihar), GJ (Gujarat), HR (Haryana), KL (Kerala),

MP (Madhya Pradesh), OR (Odisha), PB (Punjab), RJ (Rajasthan) and the TG (Telangana) which were in the low-risk group now are in moderate-low risk state. Rest of the states are in the low-risk cluster.

The current study is based on the number of confirmed cases and subject to reporting biases if any in the data source. We have not considered other relevant factors, co-variate and non-pharmaceutical interventions which might completely alter the picture favourably. Number of confirmed cases alone can not give a true picture of the prevalence of the disease as it is proportional to the number of tests conducted. Our aim in this work is not to prove or disprove anything but present the pattern for everyone to see and realize that the situation we are in still remains critical. There is no room to rejoice now by saying we are in low risk compared to MH. Yes! MH is ahead of us but not in terms of numbers but we are behind them in time scale and it is a matter of time before we reach and experience that stage.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. India unlock 3.0 guidelines: interstate travel rules in the country, <https://www.timesnownews.com/auto/features/article/india-unlock-3-guidelines-interstate-travel-rules-in-the-country/630430>
2. Guidelines for phased re-opening (unlock 4) (Aug 2020), https://www.mha.gov.in/sites/default/files/MHAOrder_Unlock4_29082020.pdf
3. Unlock 3.0 guidelines: here is what's allowed, what's not.2020 (Jul 2020), <https://indianexpress.com/article/india/unlock-3-0-guidelines-rules-whats-allowed-whats-not-6529596/>
4. Banerjea, W.B.A.: Coronavirus lockdown extended till 31 may, says ndma (May 2020), <https://www.livemint.com/news/india/covid-19-lockdown-4-0-coronavirus-lockdown-extended-till-31-may-says-ndma-11589715203633.html>
5. BBC: India coronavirus: All major cities named covid-19 'red zone' hotspots (Apr 2020), <https://www.bbc.com/news/world-asia-india-52306225>
6. Bhaskar, U.: India to remain closed till 3 may, economy to open up gradually in lockdown 2.0 (Apr 2020), <https://www.livemint.com/news/india/pm-modi-announces-extension-of-lockdown-till-3-may-11586839412073.html>
7. Blumenstock, J.: Machine learning can help get covid-19 aid to those who need it most. Nature (2020)
8. Desk, T.H.N.: India lockdown 4.0 guidelines: What's allowed and what's not? (May 2020), <https://www.thehindu.com/news/national/lockdown-4-0-guidelines-whats-allowed-and-whats-not/article31609394.ece>
9. Dutta, P.K.: In coronavirus lockdown extension, modi wields stick, offers carrot on exit route (Apr 2020), <https://www.indiatoday.in/coronavirus-outbreak/story/in-coronavirus-lockdown-extension-modi-wields-stick-offers-carrot-on-exit-route-1666741-2020-04-14>
10. Han, J., Pei, J., Kamber, M.: Data mining: concepts and techniques. Elsevier (2011)
11. Javaid, M., Haleem, A., Vaishya, R., Bahl, S., Suman, R., Vaish, A.: Industry 4.0 technologies and their applications in fighting covid-19 pandemic. Diabetes Metabolic Syndrome: Clinical Research Reviews 14(4), 419–422 (2020)
12. Kumar, A., Gupta, P.K., Srivastava, A.: A review of modern technologies for tackling covid-19 pandemic. Diabetes Metabolic Syndrome: Clinical Research Reviews 14(4), 569–573 (2020)

13. Kumar, S.: Monitoring novel corona virus (covid-19) infections in india by cluster analysis. *Annals of Data Science* (2020)
14. Manias, M.: Doubling time calculator (Jan 2020), https://www.omnicalculator.com/math/doubling_time
15. Murphy, C.B.: Understanding the compound annual growth rate – cagr (May 2020), <https://www.investopedia.com/terms/c/cagr.asp>
16. newsworld24: Lockdown extension till may 17: Read mha guidelines (May 2020), <https://www.newsworld24.in/2020/05/lockdown-extension-till-may-17-read-mha-guidelines.html>
17. Online, E.: Lockdown 4.0 guidelines: Nationwide lockdown extended till may 31, with considerable relaxations (May 2020), <https://economictimes.indiatimes.com/news/politics-and-nation/centre-extends-nationwide-lockdown-till-may-31-with-considerable-relaxations/articleshow/75790821.cms>
18. Online, E.: Lockdown extended by 2 weeks, india split into red, green and orange zones (May 2020), <https://economictimes.indiatimes.com/news/politics-and-nation/govt-extends-lockdown-by-two-weeks-permits-considerable-relaxations-in-green-and-orange-zones/articleshow/75491935.cms>
19. Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Acharya, U.R.: Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in Biology and Medicine* 121, 103792 (2020)
20. Prasad, R.K., Sarmah, R., Chakraborty, S.: Incremental k-means method. *Lecture Notes in Computer Science Pattern Recognition and Machine Intelligence* p. 38–46 (2019)
21. Sharma, N., Ghosh, D.: "unlock1": Malls, restaurants, places of worship to reopen june 8 (May 2020), <https://www.ndtv.com/india-news/lockdown-extended-till-june-30-malls-restaurants-can-reopen-from-june-8-except-in-containment-zones-2237910>
22. Singh, K.D., Goel, V., Kumar, H., Gettleman, J.: India, day 1: World's largest coronavirus lockdown begins (Mar 2020), <https://www.nytimes.com/2020/03/25/world/asia/india-lockdown-coronavirus.html>
23. Thacker, T.: Centre issues state-wise division of covid-19 red, orange green zones (May 2020), <https://economictimes.indiatimes.com/news/politics-and-nation/centre-issues-state-wise-division-of-covid-19-red-orange-green-zones/articleshow/75486277.cms>
24. Tiwary, D.: Unlock 2: More flights, trains, but no schools and colleges till july 31 (Jun 2020), <https://indianexpress.com/article/coronavirus/unlock-2-guidelines-july-coronavirus-6482179/>
25. Ward, A.: India's coronavirus lockdown and its looming crisis, explained (Mar 2020), <https://www.vox.com/2020/3/24/21190868/coronavirus-india-modi-lockdown-kashmir>
26. Zarikas, V., Pouloupoulos, S.G., Gareiou, Z., Zervas, E.: Clustering analysis of countries using the covid-19 cases dataset. *Data in Brief* 31, 105787 (2020)

Figures

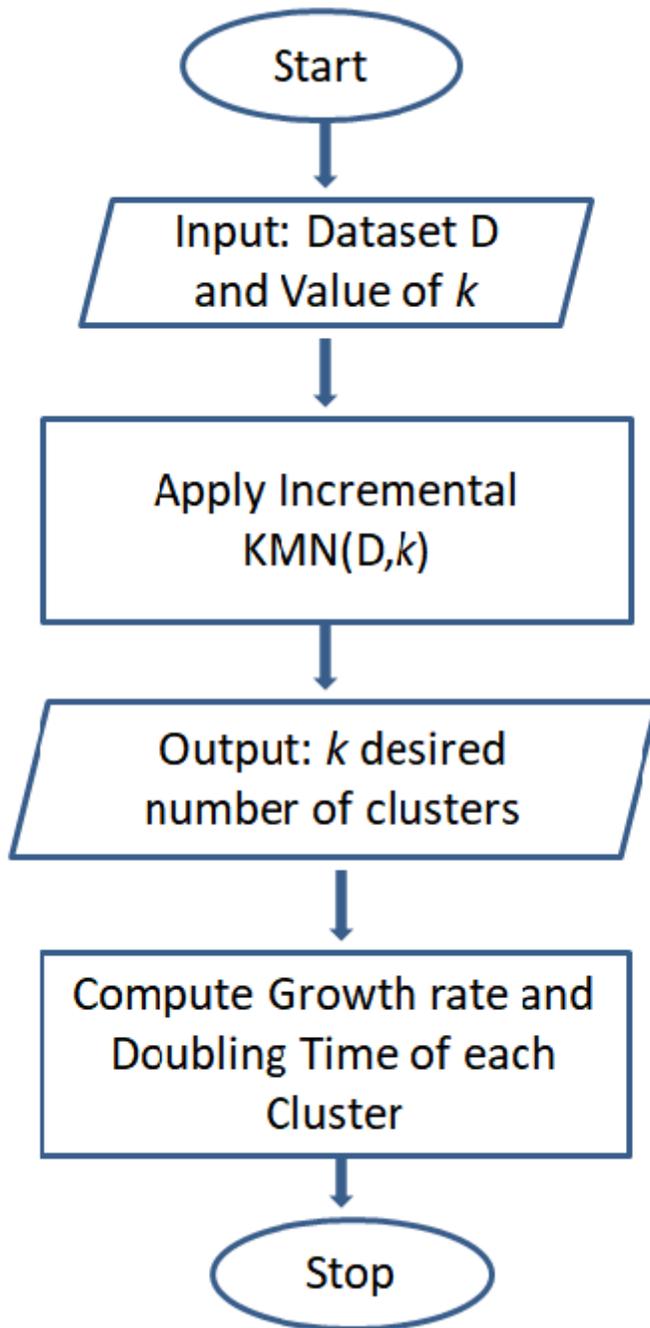


Figure 1

The flowchart of the Proposed Method

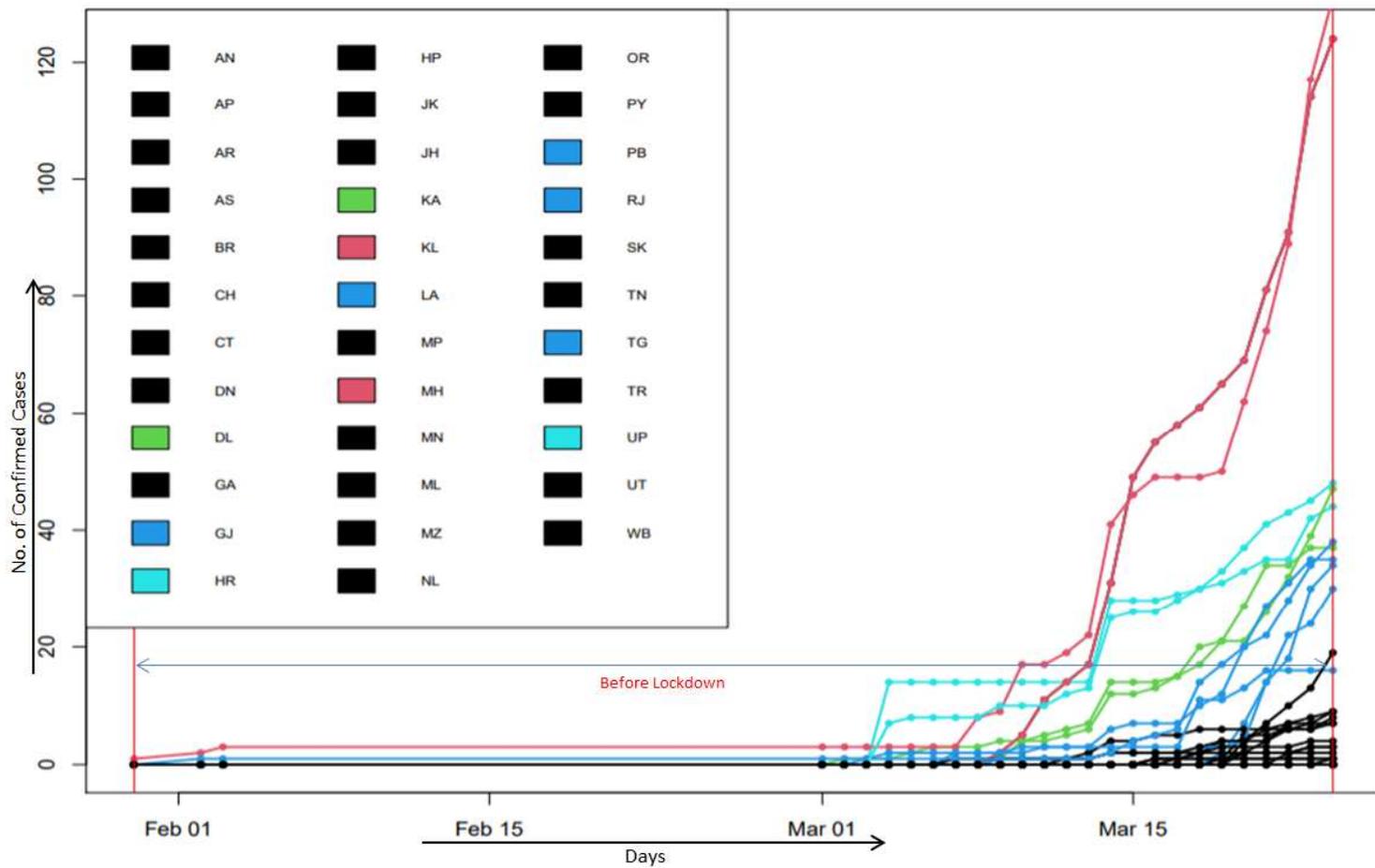


Figure 2

Result of COVID-19 dataset before lockdown using IncrementalKMN clustering method

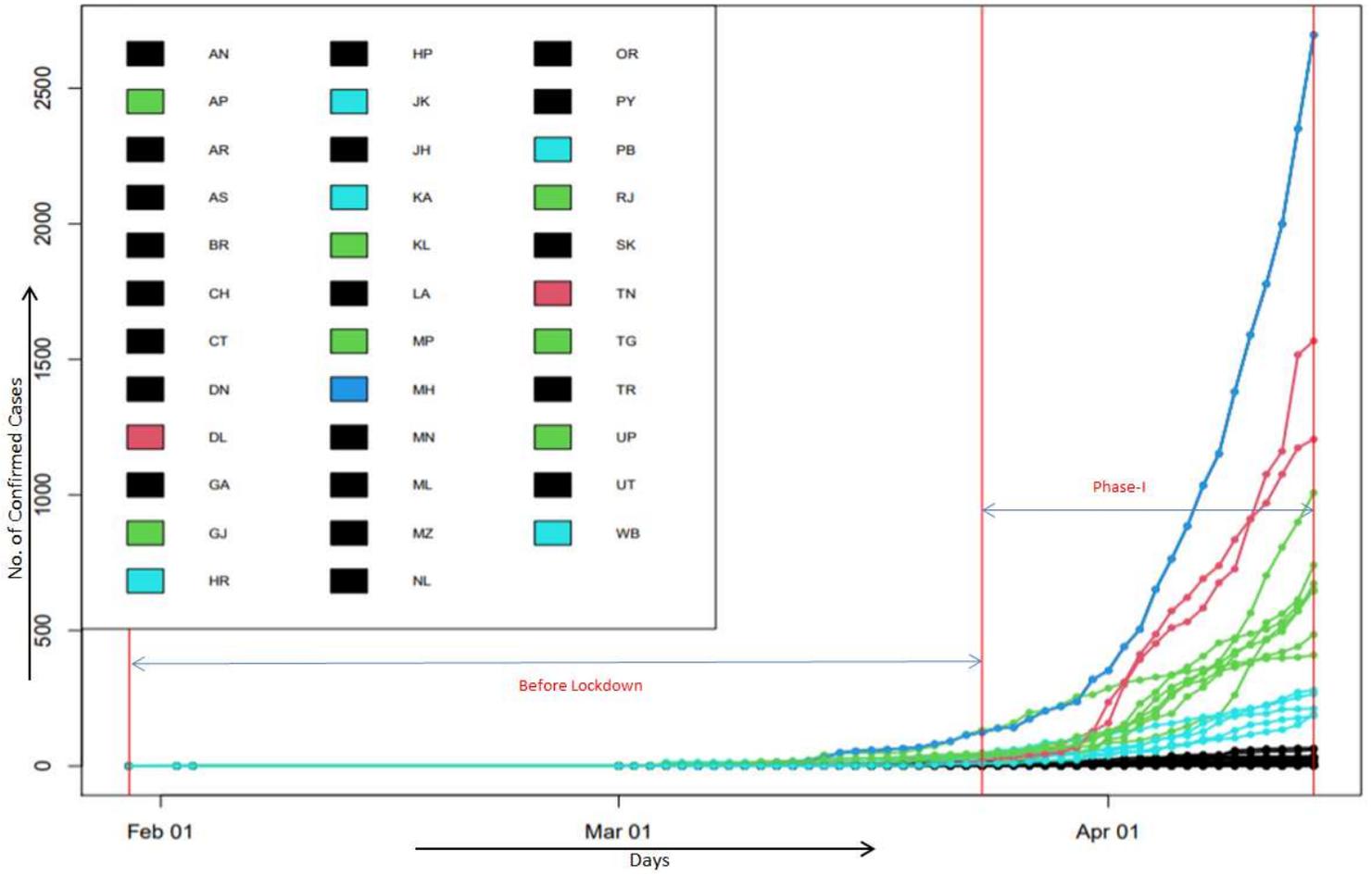


Figure 3

Result of COVID-19 Lockdown Phase-I dataset using IncrementalKMN clustering method

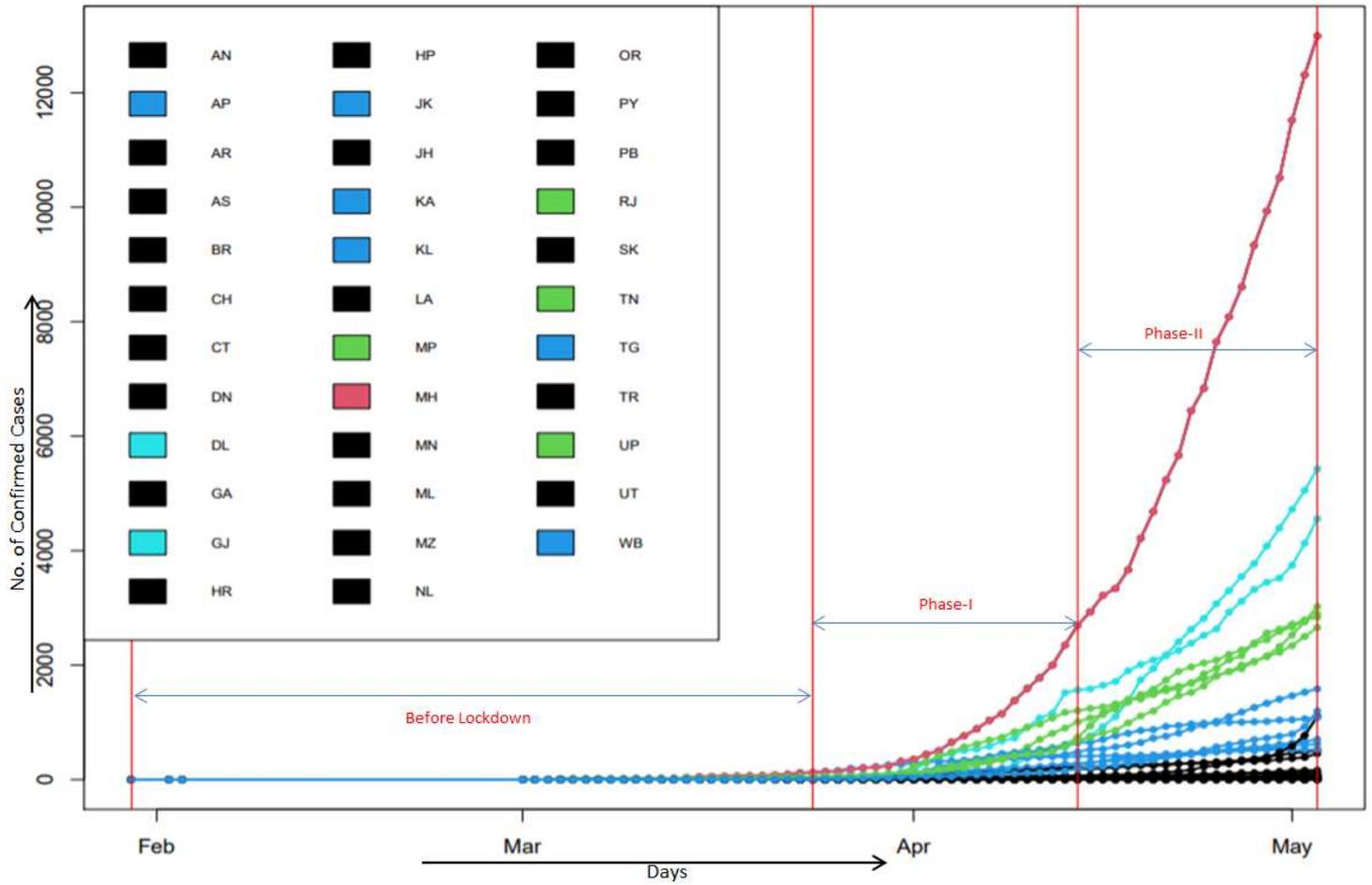


Figure 4

Result of COVID-19 Lockdown Phase-II dataset using IncrementalKMN clustering method

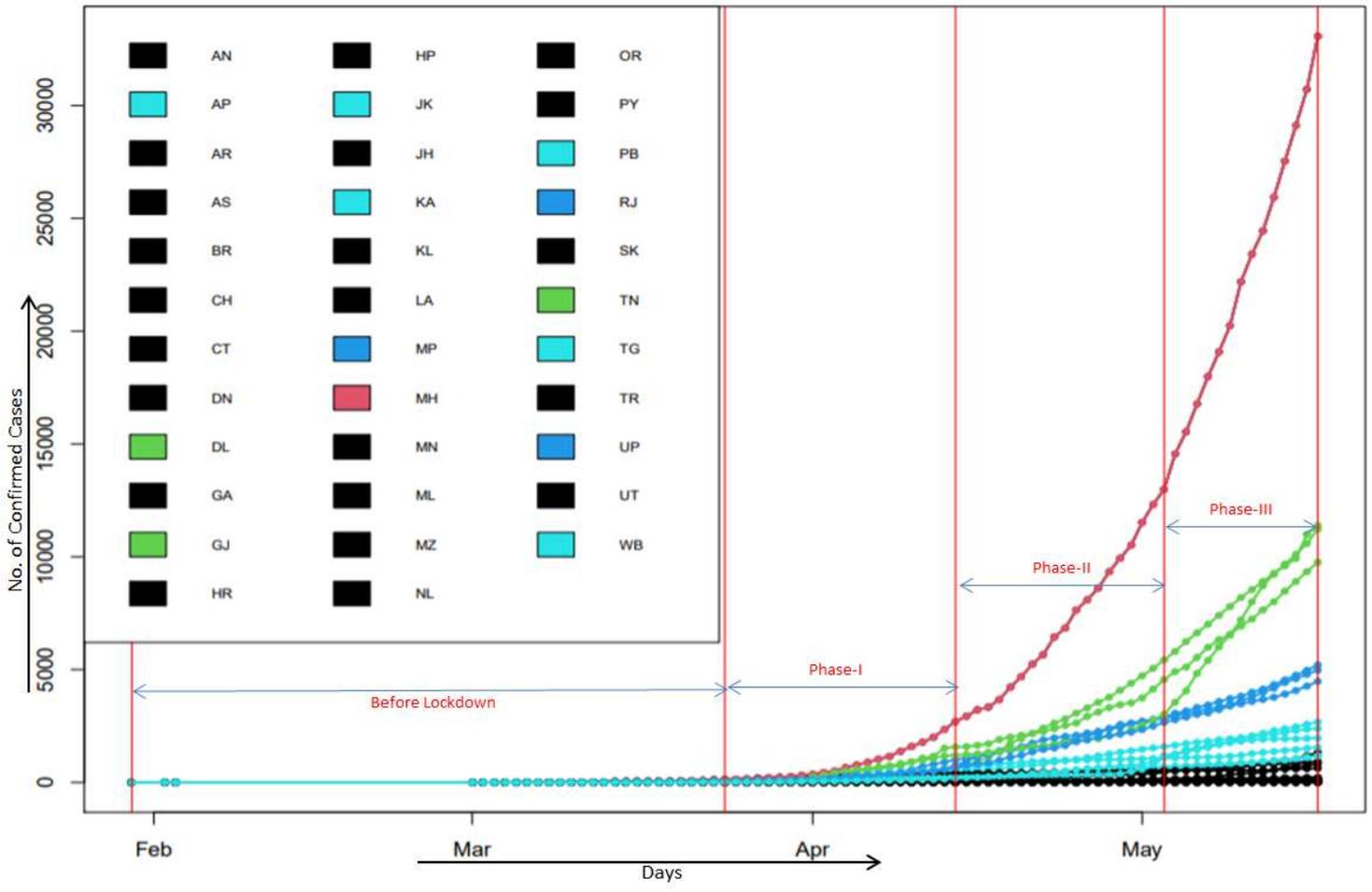


Figure 5

Result of COVID-19 Lockdown Phase-III dataset using IncrementalKMN clustering method

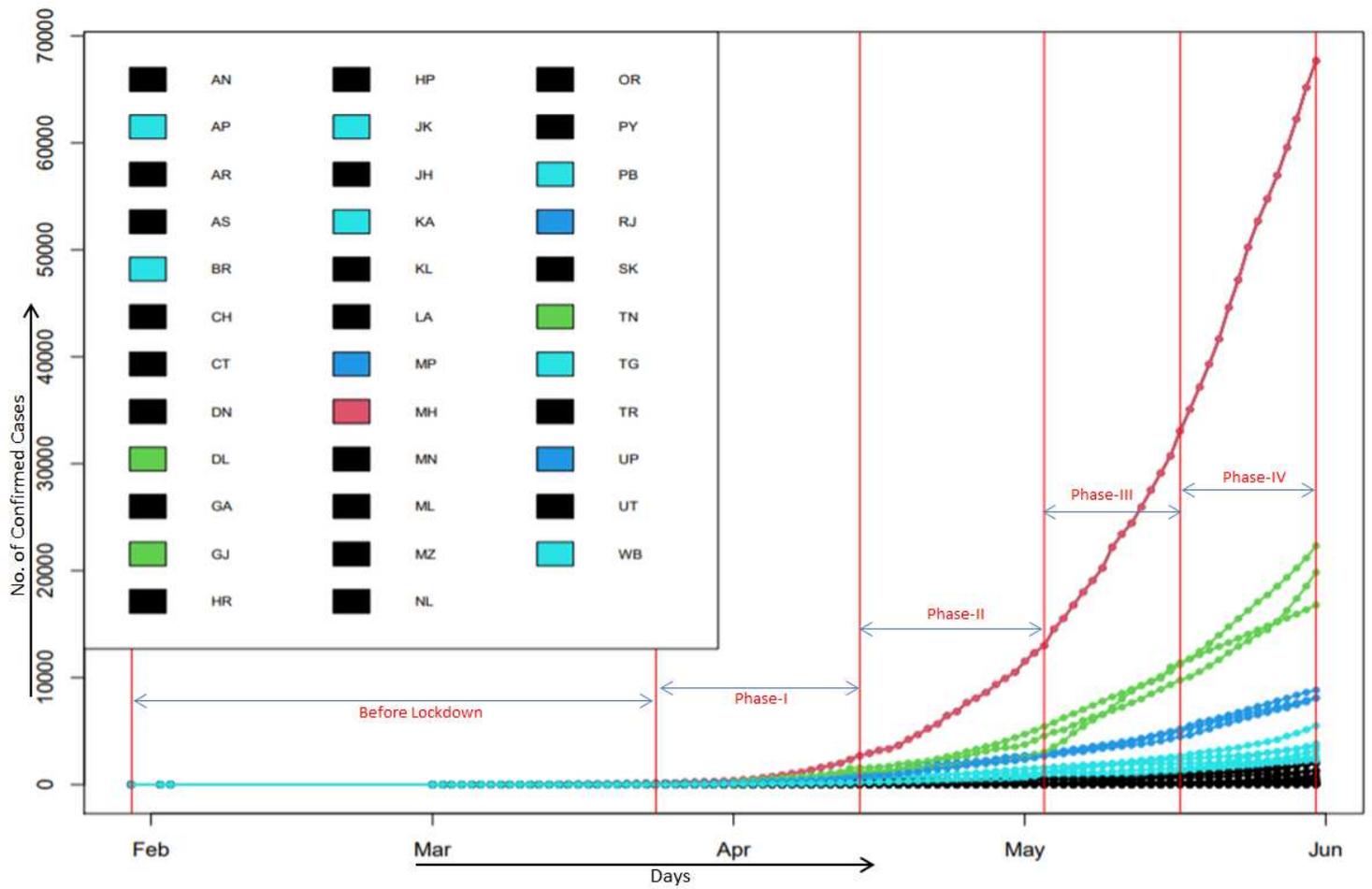


Figure 6

Result of COVID-19 Lockdown Phase-IV dataset using IncrementalKMN clustering method

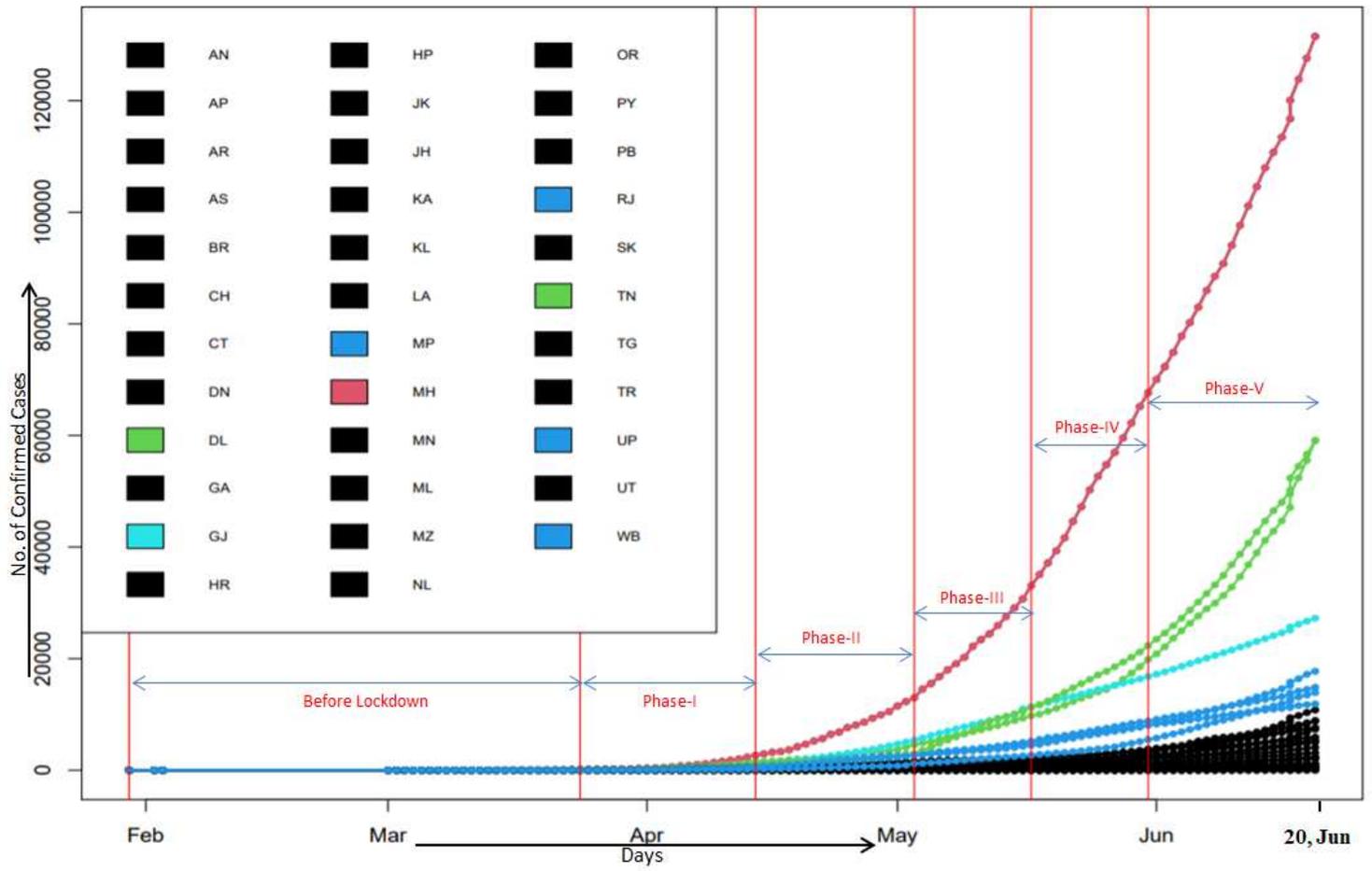


Figure 7

Result of COVID-19 Lockdown Phase-V dataset till 20th June 2020 using IncrementalKMN clustering method(k=5)

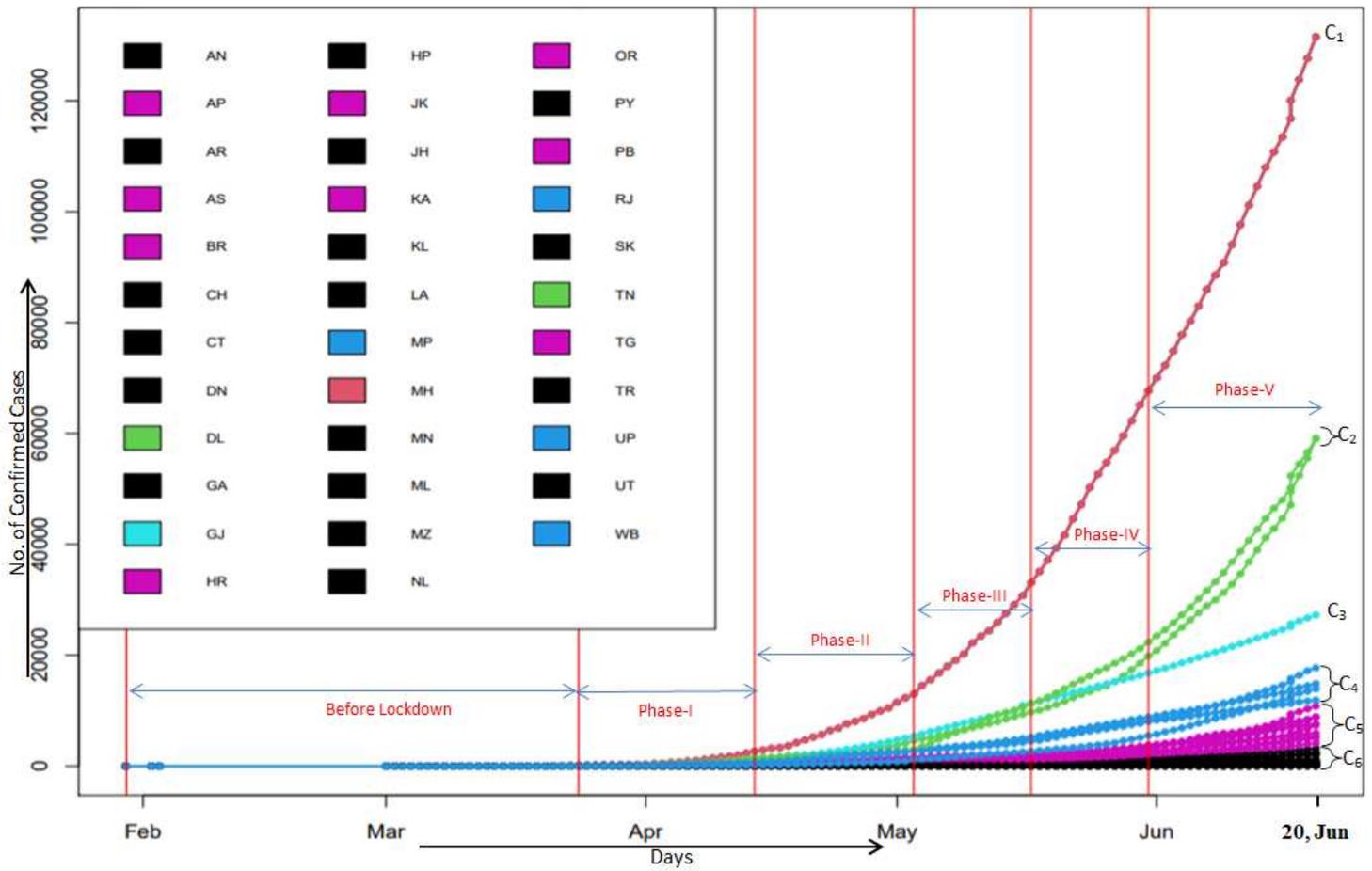


Figure 8

Result of COVID-19 Lockdown Phase-V dataset till 20th June 2020 using IncrementalKMN clustering method($k=6$)

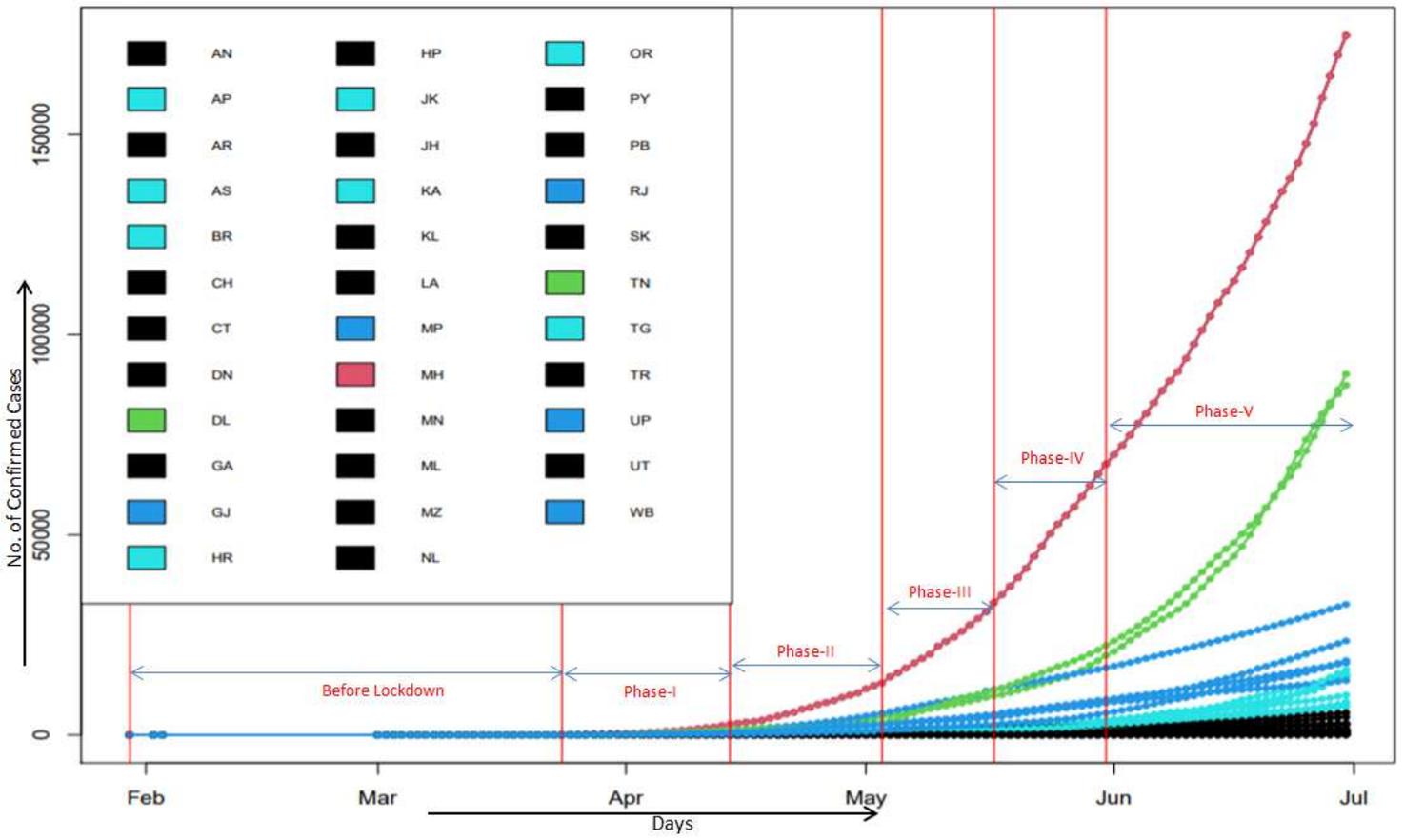


Figure 9

Result of COVID-19 Lockdown Phase-V dataset using IncrementalKMN clustering method

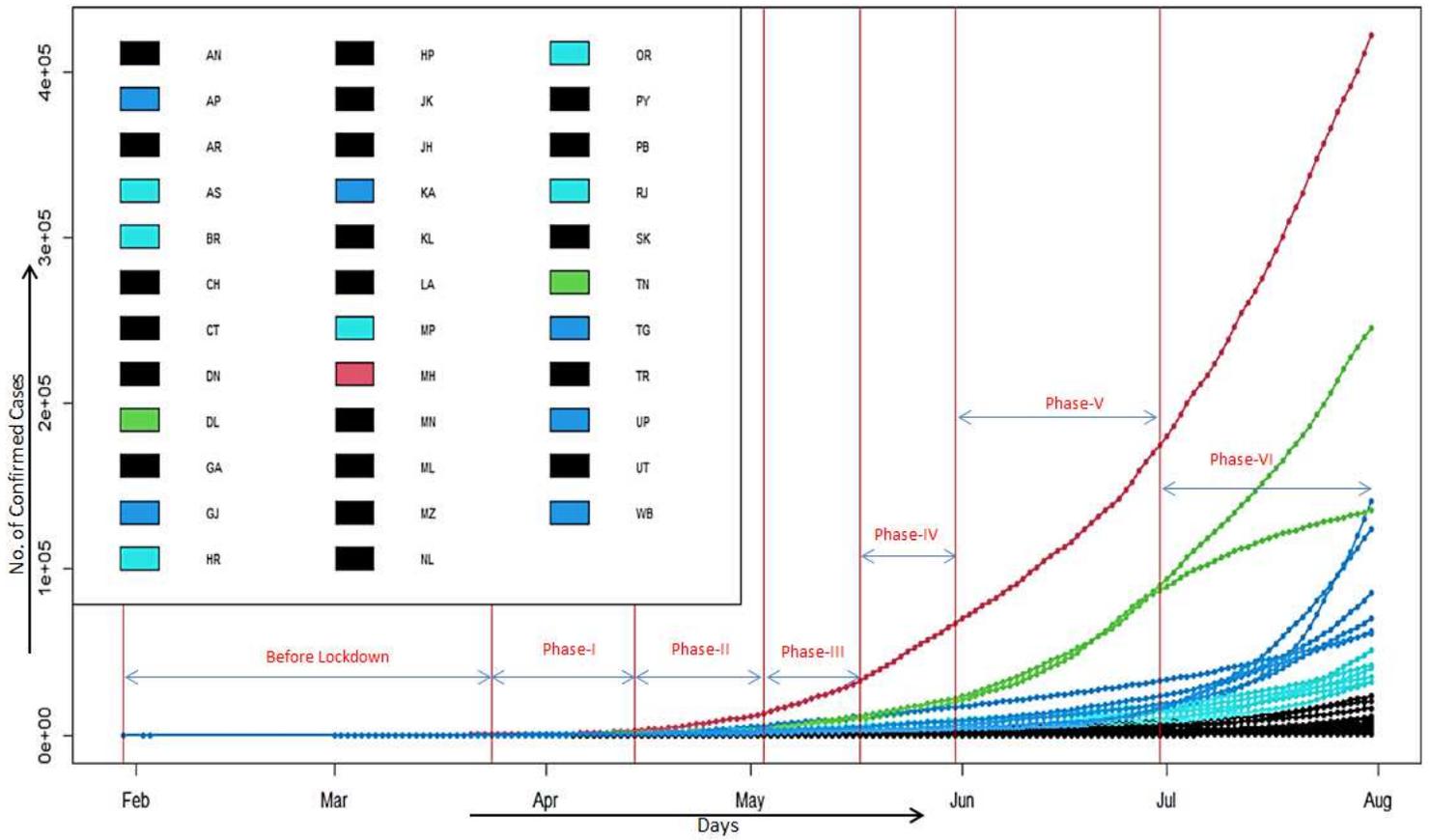


Figure 10

Result of COVID-19 Lockdown Phase-VI dataset using IncrementalKMN clustering method

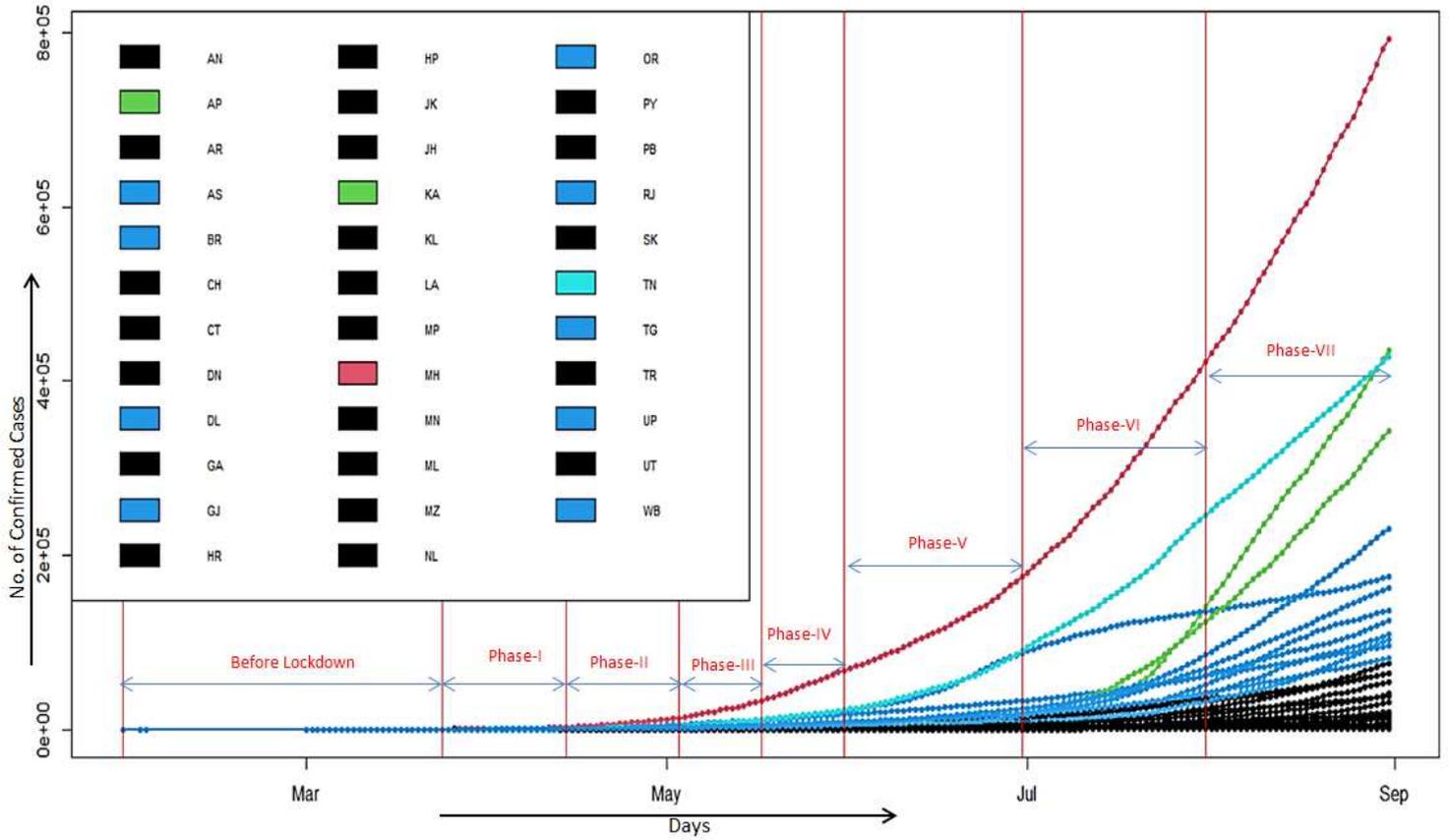


Figure 11

Result of COVID-19 Lockdown Phase-VII dataset using IncrementalKMN clustering method

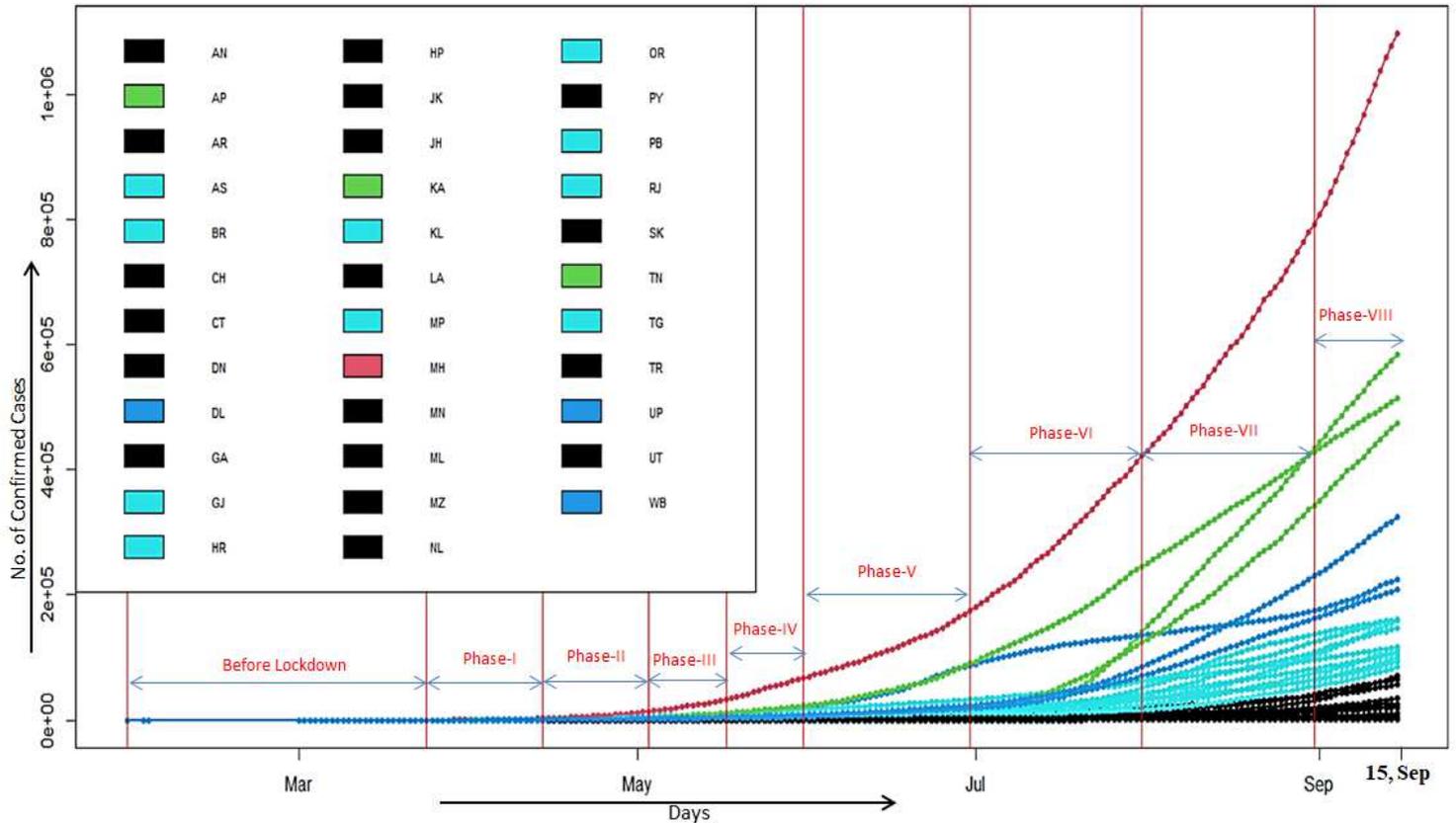


Figure 12

Result of COVID-19 Lockdown Phase-VIII dataset till 15th September 2020 using IncrementalKMN clustering method

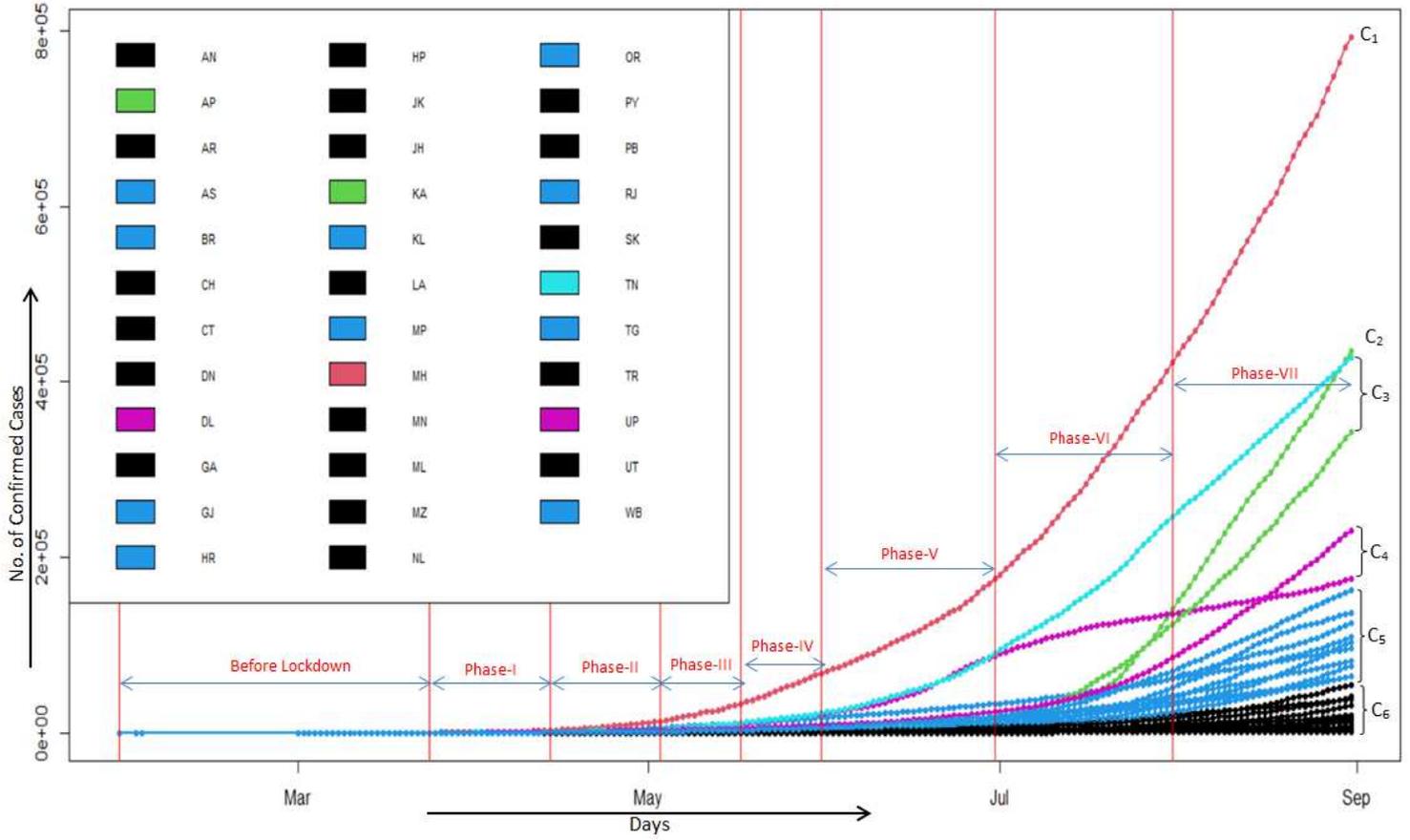


Figure 13

Result of COVID-19 Lockdown Phase-VII dataset using IncrementalKMN clustering method