ACHSM Asia-Pacific Journal of Health Management

RESEARCH ARTICLE

CLUSTERING COUNTRIES OF THE WORLD BASED ON THE TREND OF THE COVID-19 INCIDENCE: AN APPLICATION OF SHAPE-BASED K-MEANS ALGORITHM

Arefeh Dehghani Tafti[1], Yunes Jahani[1,2], Sara Jambarsang[3], Abbas Bahrampour [1,2,4]*

1. Department of Biostatistics and Epidemiology, Faculty of Public Health, Kerman University of Medical Sciences, Kerman, Iran
2. Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
3. Department of Biostatistics and Epidemiology, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
4. Griffith University, Brisbane, QLD, Australia

Correspondence: abahrampour@yahoo.com

ABSTRACT

BACKGROUND:

The urban Health Commission of Wuhan City, China, issued an emergency notice due to an incidence of viral pneumonia of unknown cause in December 2019. The World Health Organization officially named it the 2019 novel coronavirus. Since the course of the disease is not the same in the countries and regions of the world, and the study of this diversity is an important source of information for policymakers and researchers. The trend and progress of COVID-19 is more important than the time of its incidence, the current study aims to cluster selected countries of the world based on its incidence and it was done according to the trajectories shape.

METHODS:

The data set analyzed, included new cases of COVID-19(per million people) in 13 countries of the world, which were published on the World Health website from March 2020, to October 2021 monthly then analyzed by the k-means clustering method using Fréchet distance and R software V4.0.5. in addition, clustercrit and kmlshape packages were utilized for trajectory clustering (updated monthly in countries).

RESULT:

Research results show that, 13 countries of the world were classified into 2 clusters with high and low incidence. The cluster with high incidence included 8(62%) countries. The 2 cluster exhibits the highest outlier in COVID-19 incidence. In the analyzed nations, United States and Brazil exhibited the highest incidence rate in clusters 1 and 2, respectively.

CONCLUSION:

The present findings showed that there are 2 patterns in the epidemic of COVID-19. The first pattern includes severe fluctuations and the next pattern includes low fluctuations. The results revealed that the method used in this article has the potential to understand incidence trends regardless of the time of disease onset. Since the Covid-19 infection process experiences fluctuations over time that vary based on when the pandemic began in each country, it's essential to analyze the similarity and shape of infection trends collectively, independent of their starting points.

KEYWORDS
Clustering, COVID-19, K-means, Fréchet distance, Longitudinal data, incidence rate

INTRODUCTION

The COVID-19 sickness is a usual contagious and widespread infectious disease between humans and animals caused by a virus called severe acute respiratory syndrome coronavirus 2 (SARS-COV-2). This disease was first reported in December 2019 in the city of Wuhan, China, and gradually spread throughout the country and the world and It is generally considered to be under better control in many regions compared to earlier stages of the pandemic. The disease causes serious damage to the respiratory system and sometimes leads to death[1-3].

As of 4:14 pm CET, the World Health Organization (WHO) had recorded a cumulative total of 281,808,270 confirmed cases of COVID-19 globally on December 29, 2021. These cases have been reported from almost 216 countries, all of which have been grappling with the COVID-19 outbreak for the past three years. Accordingly, the statistics indicated a significant heterogeneity among different countries and regions. For instance, the confirmed cases in the USA and Africa by 2021 were 102,287,397 and 7,164,485, respectively. Moreover, Sweden has estimated an incidence rate of COVID-19 at 12,730.79 cases per 100,000 individuals, whereas Norway, its neighboring country, has reported 7,439.44 new cases per 100,000 people in its population [4].

The identification of the sources of these variations may contribute to the potential control and reduction of the burden of the disease.

The incidence rates of COVID-19 cases were different in various parts of the world. The reasons for this heterogeneity could be explained by various factors, including but not limited to different distributions of socio-demographic characteristics[5, 6], the prevalence of main risk factors and major comorbidities [7], environmental factors[8], governments policies[9, 10], health system infrastructures [1], social adherence to protocols[11], Etc. Identification of the origin of these heterogeneities can be important for introducing the best method to prevent the occurrence of future pandemics and the necessary measures to plan and implement the optimal preventive approaches[5, 6].

Given that health systems possess rich datasets; data mining is an efficient method to discover the hidden patterns of large raw data in the field of medicine. Furthermore, prediction and diagnosis systems based on this approach can reduce disease costs, waiting time, and human errors[8, 12]. Additionally, access to valid and timely health information is the cornerstone of public health activities. Therefore, having systems for collecting, analyzing, and disseminating data for experts and policymakers to identify problems and needs, track progress, evaluate the impact of interventions, and make evidence-based decisions about health policies and programs are of utmost importance[13, 14].

The high and continuous prevalence of the COVID-19 disease in various countries around the world and continuation of patients suffering from this disease to this day highlight the importance of researchers and experts paying attention to the different dimensions of this global problem. Therefore, the need to use low-cost and quick-yield solutions to manage resources and facilities in the health and treatment system raises the need to conduct a study with a data mining approach that focuses on macro and multicenter data.

In this regard, the application of new and efficient statistical approaches such as K-means clustering algorithms creates the opportunity to find homogeneous subgroups in the data. In this way, the data points in each cluster are as similar as possible according to the similarity measure such as Euclidean-based distance or correlation-based distance[15].

Therefore, the distance measure selection is an important step in this method. Certain research overlooked the temporal pattern and promptly utilized the cumulative number of COVID-19 cases at a specific moment[9, 16, 17]. Of course, in k-means clustering based on common distances such as Euclidean, the distances are calculated at each fixed point in time, so the paths (trends) with time delay and advance are not placed in the same cluster.

Therefore, considering that the pattern of the outbreak of COVID-19 and the starting point of its epidemic are different in the world and the progress of the phenomenon of COVID-19 is more important than the moment of its occurrence, therefore the clustering of the countries of the world according to the similarity and shape of pattern of the COVID-19 incidence in the global level with k-means method using Fréchet distance is important. Therefore, in the current research, we considered the approach of clustering 13 countries of the world globally by using the k-means clustering method based on the shape of the trend of the incidence of COVID-19 using the Frechet distance.

METHODS

DATA SOURCE

In this study cross-sectional, we utilized COVID-19 data sourced from the WHO COVID-19 Online Dashboard (https://github.com/owid-GitHub), spanning from March 2020 to October 2021, at 20 monthly intervals. The dataset consisted of new daily infection cases (per 1,000,000 people) for each country, confirmed by positive PCR tests. Taking into account the varying trends and the distribution of the coronavirus across different nations, we chose countries from five continents that exhibited diverse infection rates. As a result, we analyzed 13 countries with distinct incidence rates.

STATISTICAL METHODS

K-means Clustering Method

The k-means method is a method for cluster analysis with the purpose of dividing *n* observations into *k* clusters *(k ≤ n)* where each observation belongs to the cluster with the closest mean [18].

The k-means clustering method is a hill climbing algorithm that is introduced as a special case of the EM algorithm for iterative convergence [19, 20]. To initialize this method, the number of clusters *(k)* is usually required as input in advance. To reach an optimal partition, the algorithm is successively repeated between the two stages of waiting(E) and maximization(M). These two steps are repeated until stabilization is achieved in cluster assignment. To formulate the k-means algorithm, it is assumed that there is a data set with *n* observations $\boldsymbol{X} = \{\boldsymbol{x_1}.\cdots.\boldsymbol{x_n}\}$ **clustered in** $\boldsymbol{k}$ **groups** $\boldsymbol{C} = \{\boldsymbol{c_k}.\boldsymbol{K} = \mathbf{1}.\cdots.\boldsymbol{k}\}$[21, 22] · In longitudinal data, each observation x_i *(i=1, …,n)* shows a trajectory which is created by values of *ith* observer in different times $\{\boldsymbol{j} = \mathbf{1}.\cdots.\boldsymbol{t}\}$ and is shown by $\boldsymbol{X_i} = \{\boldsymbol{x_{i1}}.\boldsymbol{x_{i2}}.\cdots.\boldsymbol{x_{it}}\}$. x_{it} is the measured value of *ith* subject in the time t. k-means tries to find a partition where the set of distances between each observation and its cluster center is minimal. Suppose that Z_k, which is the average of the observations belonging to the corresponding cluster, represents the center of the cluster C_k. The square of the distance between Z_k and all observations X_i in the cluster C_k can be defined as the follows:

$$(1)\quad \boldsymbol{SD}(\boldsymbol{C_K}) = \sum_{\boldsymbol{x_i}\in\boldsymbol{C_k}}\|\boldsymbol{x_i} - \boldsymbol{z_k}\|^2$$

The aim of the k-means algorithm is to minimize the sum of squared distances between each observation and the corresponding center in all *k* clusters [23], which is The objective function is expressed as per relation (2).

$$(2)\quad \boldsymbol{argmin}\sum_{\boldsymbol{k}=\mathbf{1}}^{\boldsymbol{K}}\sum_{\boldsymbol{x_i}\in\boldsymbol{C_k}}\|\boldsymbol{x_i} - \boldsymbol{z_k}\|^2$$

k-means Clustering under Fréchet Distance (according to shapes of the trajectories)

Fréchet distance was introduced by Morris in 1906 [24]. This distance is a similarity measure for geometric shapes and unlike the Euclidean distance, it considers each trajectory as a curve and is able to identify the clusters according to shape of the trajectories and not their classical distance. The first algorithm for calculating this distance was presented by Alt and Godau in 1995 [25].

Formally, according to two definitions: (1) a reparameterization α of [0,1], a non-decreasing continuous function spanning α: [0,1] → [0,1] with the condition α (0) = 0 and α (1) = 1. (2) Consider a metric space S where a curve f in S is a continuous mapping from the unit interval [0,1] to S.

Consider two given curves f and g located in S. Fréchet distance between two curves f and g in mathematical writing is defined as follows [25-28]:

$$(3)\quad \boldsymbol{\delta_F}(\mathbf{f}.\mathbf{g}) = \underset{\boldsymbol{\alpha}.\boldsymbol{\beta}\ \ \mathbf{t}\in[\mathbf{0}.\mathbf{1}]}{\mathbf{inf}\mathbf{max}}\{\|\mathbf{f}(\boldsymbol{\alpha}(\mathbf{t})) - \mathbf{g}(\boldsymbol{\beta}(\mathbf{t}))\|\}$$

where ||.|| is the corresponding norm and is usually the Euclidean norm, and α and β are re-parameterized [0,1]. Considering that the progress of the COVID-19 epidemic in the selected countries has been more important than the moment of its occurrence, we employ the k-means method using the Fréchet distance to clustering selected countries according to the incidence trends similarity[28].

Data preparation, and analysis were done by *R* software (4.0.5), in addition, clustercrit and kmlshape packages were utilized for trajectory clustering (updated monthly in countries).

We specified the number of desirable clusters according to the Cubic Clustering Criterion (CCC) index and bootstrap value. The CCC and the bootstrapped value (test statistics) can be utilized to determine the optimal number of clusters through the Ward minimum variance method, k-means, or other approaches that focus on minimizing the within-cluster sum of squares. The effectiveness of the CCC is assessed using Monte Carlo methods. Therefore, according toCCC index and bootstrap value=-1.9181 we chose to put the trajectories in 2 clusters.

The Calinski-Harabasz index was employed to assess the model's fit and the optimal number of clusters, yielding a Calinski-Harabasz score of 250.

RESULTS

Figure 1.a shows the trends of the monthly incidence of COVID-19 in 13 countries during 20 months (March, 2020 to October, 2021) by the trends in each cluster, along with the average trends in each cluster in red and green. According to diagram 1.a, 8 countries (62%) belong to the first cluster and 5 countries (38%) belong to the second cluster. Both clusters seem to trend downward over the study period, just there are two different templates.

Figure 1.b and Figure 1.c show the trends of the monthly update of COVID-19 belonging to the first and second clusters, respectively. According to diagram 1.b, in the first cluster, the trend of the incidence of COVID-19 is almost constant until September 2020, and after that the upward trend begins with a moderate slope until we encounter the peak of incidence in this cluster in March 2021. After a short increasing period, the downward trend starts with a relatively steep slope. But in this cluster, after the peak of the incidence of COVID-19 in March 2021, we are facing a downward trend with a steep slope, so that after a sharp decrease in the incidence again in July 2021, the incidence will increase. However, the peak incidence in July 2021 is less intense than in March 2021, and at the end of the period, the incidence in this cluster continues its downward trend.

According to Figure 1.c, the development trend in the second cluster has less fluctuations than the first cluster. So that from the beginning of the study until August 2020, the trend of incidence is almost constant, and from August 2020 onwards, the upward trend starts with a moderate slope, which continues until May and June 2021, and then the incidence is controlled and the decreasing trend begins.

In addition, by taking a closer look at the trend of the incidence of COVID-19 within each cluster, it is possible to obtain multiple fluctuations in the incidence of COVID-19 in countries. For example, during the study period, the United States had strong incidence peaks compared to Iran and Malaysia, in other words, in the first 8 months of the beginning of the study (from March 2020 to October 2020), the Malaysia and Iran had a relatively stable incidence trend and continued the upward trend of the onset but the upward trend of COVID-19 in America started from the beginning of the study period. In Malaysia and Iran, the first peak was in August 2021, and in America, it was in December 2020 and August 2021, and the intensity of the peak in December 2020 and August 2021 in America was higher than in the two countries of Iran and Malaysia. In the second cluster, Brazil had more peaks compared to India, in other words, Brazil had 3 peaks in July 2020, March 2021 and July 2021, but India had two peaks in September 2020 and May 2021. Of course, the severity of the prevalence of COVID-19 in Brazil was higher than in India.

All clusters seem to show an increased rate in the study period, but there are several different patterns. Based on the fluctuation pattern of the monthly incidence of COVID-19 and the magnitude of the occurrence (increased rate), the first and second clusters can be introduced as clusters with high and low incidence respectively, and in terms of the slope of increase, it can be introduced as clusters with severe and mild increase respectively.

Table 1 contains the clustering list of countries based on the trend of monthly incidence of COVID-19 with the k-means method based on the Figure.

FIGURE 1. (A) OBSERVED TRAJECTORIES OF MONTHLY COVID-19 INCIDENCE (PER 1,000,000 POPULATION) BY CLUSTER. (B) OBSERVED TRAJECTORIES OF THE MONTHLY INCIDENCE OF COVID-19 (PER 1,000,000 POPULATION) BELONGING TO THE FIRST CLUSTER. (C) OBSERVED TRAJECTORIES OF MONTHLY INCIDENCE OF COVID-19 (PER 1,000,000 POPULATION) BELONGING TO THE SECOND CLUSTER

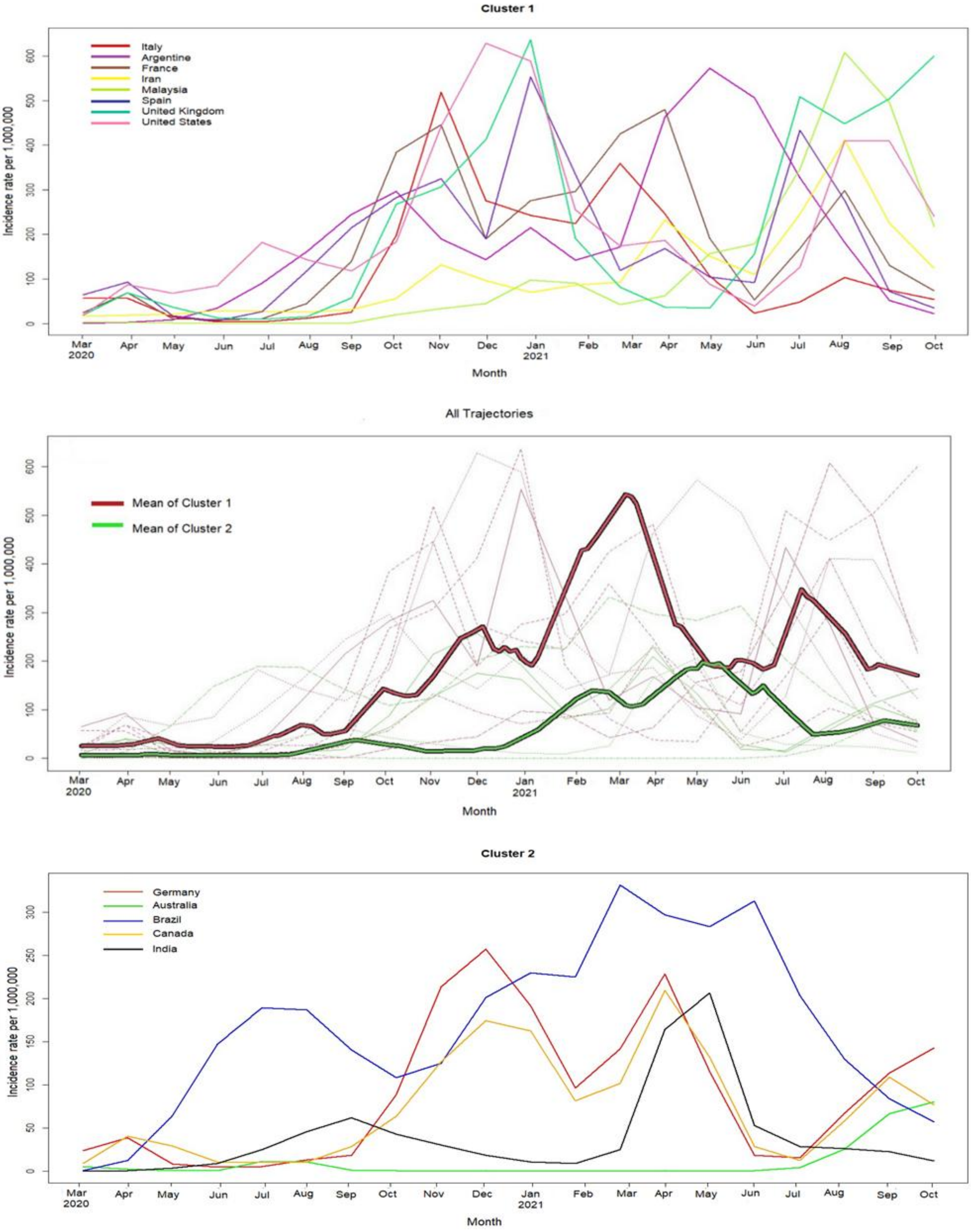


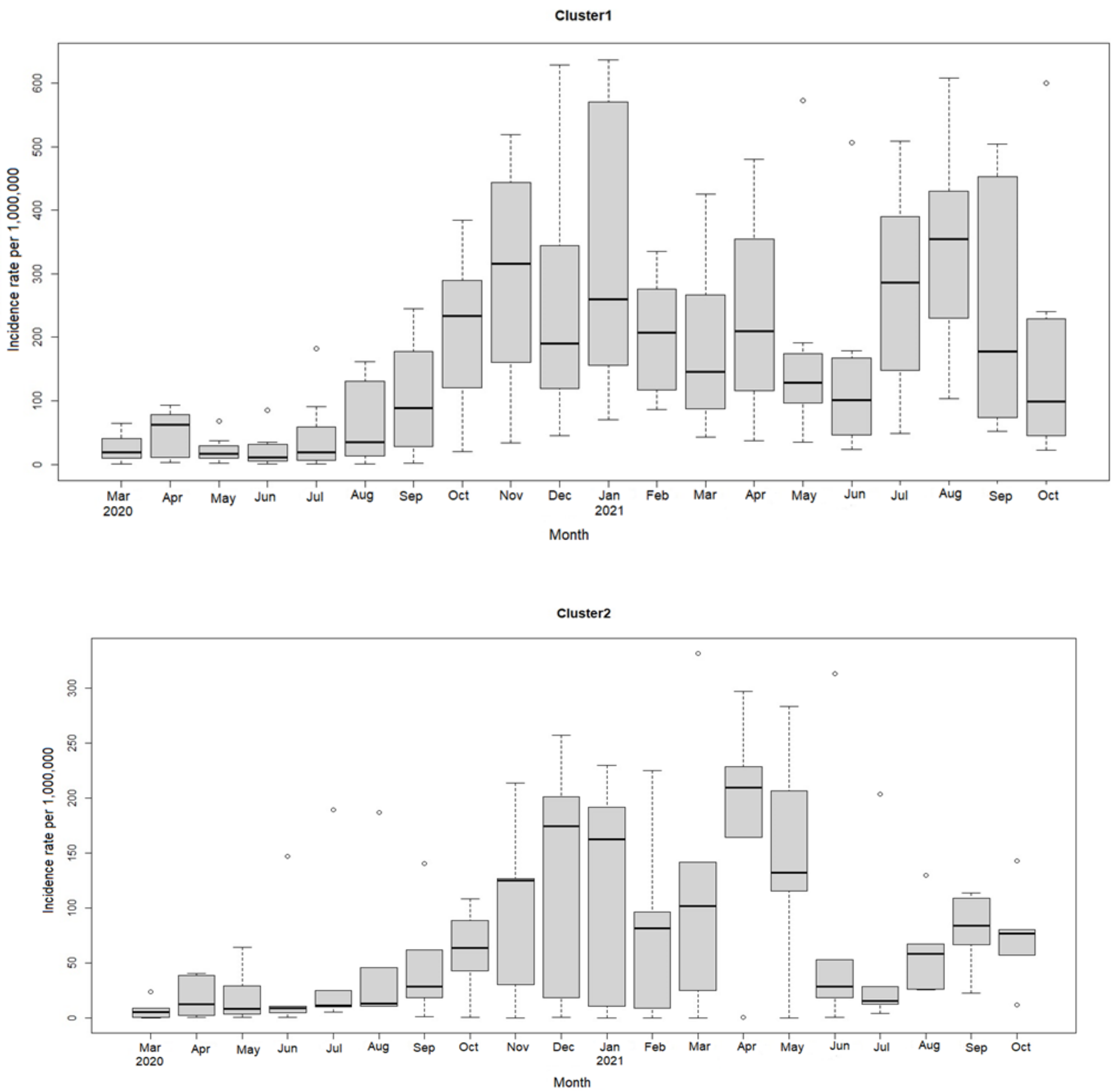
TABLE 1 CONTAINS THE CLUSTERING LIST OF COUNTRIES BASED ON THE TREND OF MONTHLY INCIDENCE OF COVID-19 WITH THE K-MEANS METHOD BASED ON THE FIGURE.

Cluster 1(sever incidence)
Italy, France, United Kingdom, United States, Malaysia, Argentina, Spain, Iran
Cluster 2(mild incidence)
Germany, India, Brazil, Australia, Canada

Figure 2 shows the distribution of the trend of COVID-19 in each cluster using box plots. According to Figure 2.a and Figure 2.b, the highest outlier of the incidence of COVID-19 belongs to cluster 2 (Figure 2.b) and among the countries in cluster 2, it belongs to Brazil. On average, the fluctuations of the incidence of COVID-19 between countries have increased after 7 months from the beginning of the study (September 2020).

In the analyzed nations, United States and Brazil exhibited the highest incidence rate in clusters 1 and 2, respectively, while Iran and Australia demonstrated the lowest incidence rate in those same clusters.

FIGURE 2. THE BOX PLOTS OF MONTHLY INCIDENCE RATE OF COVID-19 IN THE (A) FIRST CLUSTER AND (B) SECOND CLUSTER COUNTRIES DURING THE STUDY PERIOD



DISCUSSION

In this paper, an efficient clustering method was used, which is based on a modified k-means algorithm, a clustering method that clusters trajectories according to their shape. A prominent feature of this clustering method is that it allows clustering of cases whose trajectories have similar shapes, but their positions have changed in time. In fact, in this method, the development process is more important than the moment of its occurrence in each country. As a result, the clusters were classified according to shape of the trajectories in each country, and the countries with the same trend shape were placed in one cluster.

In this study, investigations were carried out on 13 countries of the world, belonging to the continents of Asia, Europe, America and Asia Pacific.

In this research, it was found that in addition to all the changes in the conditions of 13 countries of the world, the pattern of the spread of the epidemic of Covid-19 follows 2 different paths based on the shape of the paths. However, these patterns are almost like, there are several special features for each that make each one different from the other.

Based on the results, the trend pattern with severe fluctuations in the first cluster including 8 countries (62%) was more than the second cluster including 5 countries (38%). Most of the countries in this cluster are characterized as high- or average-income ones. In fact, the monthly incidence of COVID-19 in some countries of the first cluster in the period from September 2020 to March 2021 has increased sharply after a sharp decrease in the period from June 2021 to July 2021. In addition, the changes within the trajectories in the first cluster are higher than the second cluster. The change in the trend of COVID-19 incidence probably shows the concern in controlling the prevalence of the disease.

In order to clustering the patterns of COVID-19, Zarikas et al used hierarchical clustering of time series studies belonging to 30 countries during the epidemic onset and the next 80 days. Four separate clusters were obtained for the incidence rate. The first and second clusters followed a similar trajectory pattern, strongly increased and stable. The only difference that we observe was in the incidence level. Accordingly, the countries of Italy, France, America, England, and Spain belonged to the cluster with a severe incidence which was similar to the present study. Zarikas et al failed to take trends and patterns into consideration, and therefore their results failed to apply to policymakers who must regulate their conclusions as a result of novel information [29].

Gohri et al. used the k-means algorithm to cluster the countries of the world. According to the clustering of the incidence of COVID-19, the countries of the world were in three different patterns. The results of the present study were relatively similar to those of Gohri et al. In this regard, the countries of Germany, Australia, Brazil, Canada, and India showed a similar pattern in both studies [6].

Another study conducted by Mahmoudi et al, selected seven different countries and categorized them into three different patterns using a complex fuzzy clustering method. The clustering results of their study showed that the distribution of dispersion in Spain and Italy was almost similar and different from other countries [16]. However, in the present study, the Central European countries had the same pattern as the United States.

Melin et al used self-organizing maps to cluster recoveries, confirmed cases, and deaths associated with COVID-19 and arbitrarily selected four clusters. If several clusters had been collected, our results would have been relatively similar. Take, for instance, the countries of America, France, Italy, England, and Spain, which exhibit a similar pattern in both investigations. In addition, the incidence trend of COVID-19 in Canada, India, Iran, Australia, Argentina, and Malaysia was lower than in these countries [17].

Several studies used a greater number of prospective variables, resulting in different clusters. In Rizvi et al study the data from 89 countries and cross-referencing 18 indicators of environmental conditions and health systems were thoroughly examined. Their K-means algorithm resulted in four clusters. The K-means algorithm they used produced a total of four clusters. Therefore, all high-income countries were included in one cluster regardless of the actual trend of COVID-19[1].

As another attempt to cluster countries based on COVID-19, Pesin et al obtained the optimal clusters from the k-means clustering method and the elbow method. The optimal clusters in these methods were four and three, respectively. Also, the two-stage clustering method was used for the clustering process. Similar to the results of the present study, in a study using the Elbow method, the countries of England, Brazil, Italy, the USA, and Spain were placed in one cluster, and in the k-means method, the above-mentioned countries belonged to two separate clusters. Obviously, in this study the course of the disease was not considered [30].

Another study by Mahmoudi et al., selected and categorized seven countries. Their ability to choose additional countries was impeded by the complex fuzzy clustering method. Consequently, they were able to identify three distinct patterns through the utilization of this intricate technique. Due to the greater diversity in the selection of countries in our study and taking into account the course of infection, the countries were divided into two clusters [16]. According to our Findings, Central European countries had almost the same pattern as the United States, and the incidence trends of Eastern European countries are close to African, South American, and Asian countries.

CONCLUSION

The results of this research showed that by using the k-means clustering method using Fréchet distance, there are two significant patterns in the epidemic of COVID-19 disease in the countries of the world according to the form of the disease occurrence. The first pattern included severe fluctuations in the incidence. In other words, the incidence of the disease in different months was faced with an upward trend with a steep slope in the next phase and a rapid control and reduction of the incidence of the disease. This process was repeated 2-3 times in the countries belonging to this cluster. The second pattern included low fluctuations in the incidence of the disease with a gentle slope. In other words, in the first stage, the incidence of the disease increases with a gentle slope, and in the next stage, it decreases again with a gentle slope.

While we could have concentrated on cumulative incidence and overall trends, this study emphasizes the significance of examining incidence rate patterns among countries and the subtle distinctions within clusters. Understanding these characteristics and variations is crucial for predicting future trends in monthly COVID-19 incidence and for shaping health and prevention policies. The pandemic's impact in contemporary society has generated a vast amount of data and information. Nonetheless, certain data quality issues persist, yet they still offer valuable insights. This research addresses various data types from low- and middle-income countries. Despite significant efforts, challenges remain in the registration systems, which is a critical limitation. However, we assure readers that our methodology is sound and minimizes the effects of these challenges.

The objective of this study was to categorize countries globally based on COVID-19 incidence trends using a univariate k-means analysis. However, the observed incidence trajectories for each country during the study period were notably influenced by factors such as vaccination rates, quarantine measures, and national policies. Therefore, to obtain more precise results, we recommend that future research on COVID-19 clustering employing the k-means method consider their shapes and the effects of time-dependent covariates.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The current research project with the code IR.KMU.REC.1400.602 has been approved by the National Ethics Committee in Biomedical Research and the Modeling in Health Research Center of Kerman University of Medical Sciences.

References

1. Rizvi SA, Umair M, Cheema MA. Clustering of countries for COVID-19 cases based on disease prevalence, health systems and environmental indicators. Chaos, Solitons & Fractals 2021; 151:111240.
2. Shanbehzadeh M, Kazemi-Arpanahi H, Orooji A, Mobarak S, Jelvay S. Performance evaluation of selected machine learning algorithms for COVID-19 prediction using routine clinical data: With versus Without CT scan features. Journal of education and health promotion 2021; 10:285.
3. Sheikhi K, Shirzadfar H, Sheikhi M. A Review on Novel Coronavirus (Covid-19): Symptoms, Transmission and Diagnosis Tests. 2020.
4. WHO. Covid19 Dashboard. Geneva: World Healh Organization. 2021; Available from: https://covid19.who.int/.
5. Fountoulakis KN, Fountoulakis NK, Koupidis SA, Prezerakos PE. Factors determining different death rates because of the COVID-19 outbreak among countries. Journal of public health (Oxford, England) 2020; 42:681-7.

6. Gohari K, Kazemnejad A, Sheidaei A, Hajari S. Clustering of countries according to the COVID-19 incidence and mortality rates. BMC Public Health 2022; 22:632.
7. Wolff D, Nee S, Hickey NS, Marschollek M. Risk factors for Covid-19 severity and fatality: a structured literature review. Infection 2021; 49:15-28.
8. Albahri AS, Hamid RA, Alwan JK, Al-Qays ZT, Zaidan AA, Zaidan BB, et al. Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review. Journal of medical systems 2020; 44:122.
9. Kim S, Castro MC. Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). International journal of infectious diseases: IJID : official publication of the International Society for Infectious Diseases 2020; 98:328-33.
10. Kinnunen J, Georgescu I, Hosseini Z, Androniceanu AM. Dynamic indexing and clustering of government strategies to mitigate Covid-19. 2021.
11. Behera VNJ, Ranjan A, Reza M. Graph based Clustering Algorithm for Social Community Transmission Prediction of COVID-19. arXiv preprint arXiv:201100414 2020.
12. Ahmed A, Hannan SA. Data mining techniques to find out heart diseases: an overview. International Journal of Innovative Technology and Exploring Engineering (IJITEE) 2012; 1:18-23.
13. Odero W, Rotich J, Yiannoutsos CT, Ouna T, Tierney WM. Innovative approaches to application of information technology in disease surveillance and prevention in Western Kenya. Journal of biomedical informatics 2007; 40:390-7.
14. Silwattananusarn T, Tuamsuk K. Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012. International Journal of Data Mining & Knowledge Management Process 2012; 2.
15. Abdullah D, Susilo S, Ahmar AS, Rusli R, Hidayat R. The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. Quality & quantity 2022; 56:1283-91.
16. Mahmoudi MR, Baleanu D, Mansor Z, Tuan BA, Pho K-H. Fuzzy clustering method to compare the spread rate of Covid-19 in the high risks countries. Chaos, Solitons & Fractals 2020; 140:110230.
17. Melin P, Monica JC, Sanchez D, Castillo O. Analysis of Spatial Spread Relationships of Coronavirus (COVID-19) Pandemic in the World using Self Organizing Maps. Chaos, solitons, and fractals 2020; 138:109917.
18. MacQueen J, editor. Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability; 1967 1967.
19. Genolini C, Falissard B. KmL: a package to cluster longitudinal data. Computer methods and programs in biomedicine 2011; 104:e112-21.
20. Celeux G, Govaert G. A classification EM algorithm for clustering and two stochastic versions. Computational Statistics & Data Analysis 1992; 14:315-32.
21. Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. Biometrics 1982; 38:963-74.
22. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annual review of clinical psychology 2010; 6:109-38.
23. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognition Letters 2010; 31:651-66.
24. Fréchet MM. Sur quelques points du calcul fonctionnel. Rendiconti del Circolo Matematico di Palermo (1884-1940) 1906; 22:1-72.
25. Alt H, Godau M. Computing the Fréchet Distance between Two Polygonal Curves. Int J Comput Geometry Appl 1995; 5:75-91.
26. Genolini C, Ecochard R, Benghezal M, Driss T, Andrieu S, Subtil F. kmlShape: An Efficient Method to Cluster Longitudinal Data (Time-Series) According to Their Shapes. PloS one 2016; 11:e0150738.
27. Agarwal PK, Avraham RB, Kaplan H, Sharir M. Computing the Discrete Fréchet Distance in Subquadratic Time. SIAM Journal on Computing 2014; 43:429-49.
28. DehghaniTafti A, Jahani Y, Jambarsang S, Bahrampour A. Longitudinal Data Clustering Methods: A Systematic Review. J Biostat Epidemiol 2023; 9:396-411.
29. Zarikas V, Poulopoulos SG, Gareiou Z, Zervas E. Clustering analysis of countries using the COVID-19 cases dataset. Data in brief 2020; 31:105787.
30. PASIN O, PASIN T. Clustering of countries in terms of deaths and cases of COVID-19. J Health Soc Sci 2020; 5:587-94.