Application of Spectral Clustering for the Detection of High Priority Areas of Attention for COVID-19 in Mexico

Rodriguez-Aguilar Roman(✉)

Facultad de Ciencias Económicas y Empresariales, Universidad Panamericana, Augusto Rodin 498, 03920 Mexico City, Mexico
rrodrigueza@up.edu.mx

Abstract. The recent COVID-19 pandemic has represented a great challenge for health systems around the world. That is why it is necessary to propose strategies for prioritizing care and containing the pandemic. This work proposes the use of spectral clustering to characterize high-priority areas of care based on key information on the performance of the pandemic as well as health system variables. The result shows the generation of high priority areas not only due to the deaths observed but also due to the clinical, demographic and health system variables.

Keywords: Unsupervised models · Spectral cluster · COVID-19 · Priorization · Mexico

1 Introduction

The COVID-19 pandemic that we are currently experiencing has generated multiple effects at the health and economic level. In most countries, contingency plans were developed in the health systems to meet the growing demand for health services for pandemic care. The need for more resources in the health sector is evident and therefore also requires strategic planning of the allocation of these resources. The challenges have arisen in various dimensions, from being able to guarantee access to inputs and human resources to defining strategies to contain the wave of infections. At a technical level in various countries, various studies have emerged that allow evaluating both the impact and possible strategies for solving the contingencies observed in health systems. An essential point in these studies is the prioritization of the susceptible population and the infected population at the regional level in the countries. It is important to focus resources on those priority areas that make it possible to avoid an increase in infections as well as guarantee access to medical care for patients who require second and third level care.

The research carried out during the pandemic are grouped into strategies for the containment of the pandemic. The World Health Organization issued in June 2020 the document entitled “Maintaining essential health services: operational guidance for the COVID-19 context”, where it addresses strategic aspects and recommendations for

J. A. Marmolejo-Saucedo et al. (Eds.): COMPSE 2020, LNICST 359, pp. 130–142, 2021.
https://doi.org/10.1007/978-3-030-69839-3_9

nations to overcome the current health crisis, among the main aspects The maintenance of the essential operations of the health services is highlighted, through the coordination of the different actors, as well as the prioritization of essential services that allow adaptation to each context and needs. Guaranteeing essential drugs, equipment and supplies is a fundamental element, as well as establishing emerging financing mechanisms that guarantee the availability of resources. Strengthen monitoring, communication as well as the use of digital platforms to support essential health service delivery [1]. [2] proposes a strategy for the containment of the pandemic based on four main axes, stopping the international and national flow of passengers, creating administrative zones per million people, stopping unnecessary flow between non-emergency zones except for the transport of goods, as well as establishing an information-driven value chain of services to control the spread of the pandemic within an area. [3] analyze demand scenarios for hospital beds, as well as assisted ventilation equipment, according to different scenarios for the Brazilian health system. The results show a saturation of the Brazilian health system, due to the fact that various health microregions and macro regions would operate beyond their capacity. The construction of field hospitals is important, both in places where there are historically healthcare gaps, as well as in those where there is already pressure from demand. The third message refers to the regional organization of health services that, despite being adequate in situations of habitual demand, in times of pandemic, this design implies additional challenges, especially if the distance that the patient had to travel was very far. To inform Canada's response to COVID-19, a rapid-cycle priority identification process was conducted. Seven COVID-19 priorities were identified for health services and policy research: adaptation of the system and organization of care; decision making and ethics on resource allocation; rapid synthesis and comparative policy analysis of COVID-19 response and outcomes; sanitary workforce; virtual care; long-term consequences of the pandemic; and public and patient participation. Three additional cross-cutting themes were identified: supporting the health of indigenous peoples and vulnerable populations, digital and data infrastructure, and learning from health systems and knowledge platforms [4].

Clinical studies focused on the analysis of the pandemic as well as possible treatments and the subsequent development of a vaccine. [5] analyze the main clinical treatments used so far to treat the disease through a systematic literature review. [6] raises the need to establish a set of public health interventions to mitigate the spread of the epidemic, taking the Italian case as a reference, establishing possible scenarios of the pandemic for Canada. On the other hand, issues related to the economic and social impacts due to the pandemic are addressed, a part of these studies focus on the analysis of supply chains for supplies and medicines, as well as different sectors economically affected by the pandemic. [7] Analyze at drug supply networks facing shortage challenges in many situations, such as current outbreaks like COVID-19. Drug shortages can occur due to manufacturing problems, lack of infrastructure, and immediate reaction mechanisms. A hypothetical case study is presented using optimization and simulation algorithms to observe the impact of COVID-19 on a regional supply network. [8] evaluates the main implications of COVID-19 on drug shortages, emphasizing the need for international and regional collaboration, as well as communication and flexibility within health systems. The risks that the current pandemic represents for timely access to medicines not only related to the pandemic are of great relevance for the correct management of the pandemic as well as for the correct management of future health crises. [9] addresses the potential benefits of using artificial intelligence and internet of things methods to prevent the spread of COVID-19 through a systematic review of the main applications and scope that artificial intelligence could have in the prevention and control of the pandemic. [10] propose the planning and operation of alternative care centers in disaster response situations should include the participation of pharmacists in key decision-making processes in the early planning stages. [11] show the potential application of simulation models to reduce the impact of COVID-19, highlighting the management of resources within specific regions, considering in an integral way the clinical care process as well as the timely access to the necessary resources to face the pandemic.

In the case of Mexico, there is a fragmented health system, linked to the patient's employment situation. There are three major public health systems: the IMSS (Mexican Institute of Social Security) linked to the population that works in the private sector, the ISSSTE (Institute of Social Security and Services of State Workers, for its acronym in Spanish) and INSABI (Institute of Health for Wellbeing, for its acronym in Spanish) that serves the rest of the open population and that replaced the Seguro Popular. In addition to having a fragmented health system, Mexico has an extensive and dispersed territory, which complicates logistics operations to guarantee timely access to medicines and supplies in a health context like the current one. The Mexican territory is characterized by its diversity and population dispersion, there are remote communities in the mountains of difficult access. Historically, health strategies have been defined to be able to access these remote communities, but in the framework of the COVID-19 pandemic, these operations are seen as an important challenge for the Mexican health system, which is why a prioritization mechanism is proposed region of high priority areas for pandemic care.

The health ministry has defined health jurisdictions that allow it to plan resources as well as access to health services. The Sanitary Jurisdiction is a regional administrative unit constituted based on demographic, epidemiological, geographical, political and social criteria. It brings together rural and urban municipalities and has health centers, specialized medical units, hospitals and other organizations to develop prevention and health protection programs and services. In the context of the current pandemic, it is necessary to structure the organization and allocation of resources based on the characteristics of each region affected by the pandemic, which requires the use of care areas that may not be organized within the framework of a jurisdiction sanitary. However, it is necessary to define these priority areas, albeit temporarily, in order to focus efforts and resources to contain the pandemic and guarantee access to medical care in a timely manner for the affected population, as well as establish strategies to contain the pandemic. The objective of this research is to propose the use of an unsupervised machine learning model for the prioritization of areas of attention for COVID-19, considering fundamental elements such as the availability of human resources and infrastructure, as well as the behavior of the pandemic throughout the Mexican territory. The work is structured as follows, in the first section the status of the pandemic in Mexico is addressed using statistics issued by the Ministry of Health, about tests carried out, number of infections and deaths, as well as mortality rates by groups old. The next section addresses the methodology to be used and the main sources of information. Finally, the results are presented as well as the main conclusions and recommendations of the study.

2 COVID-19 Pandemic in Mexico

The impact of the pandemic in Mexico has been differentiated according to the health conditions of the patients as well as regional aspects. The information available as of October 2020 published by the Ministry of Health shows a total of one million possible cases, of which 52% were confirmed as positive. The proportion of men with respect to women has a higher incidence, in the same way two categories are integrated in addition to positive and non-positive cases, which are inappropriate results (referring to problems with the tests) as well as cases outside the sample (Fig. 1). The increase in the number of cases and deaths has been consistent since the beginning of the pandemic, in recent months no evidence has been identified that indicates that the pandemic has happened in Mexico.

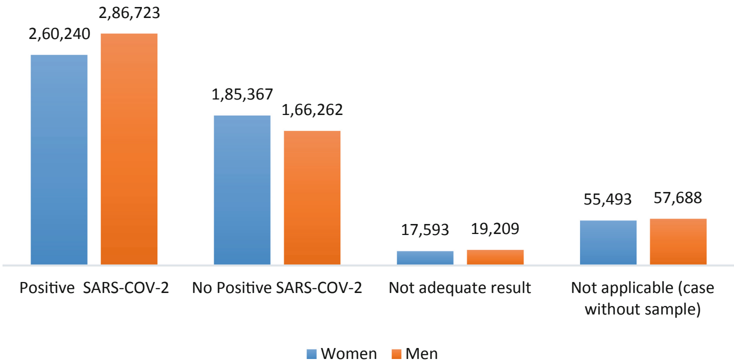

Fig. 1. Total cases registered as of October 2020. Source: Ministry of Health of Mexico.

When analyzing the distribution by federal entity, a higher incidence of confirmed cases is observed in Mexico City, the State of Mexico, Guanajuato, Nuevo León and Veracruz as the states with the highest number of confirmed cases. These five states account for 44% of the total confirmed cases throughout the country. For their part, the states with the lowest number of cases are Colima, Nayarit, Morelos, Zacatecas and Campeche with only 4% of the total confirmed cases (Fig. 2).

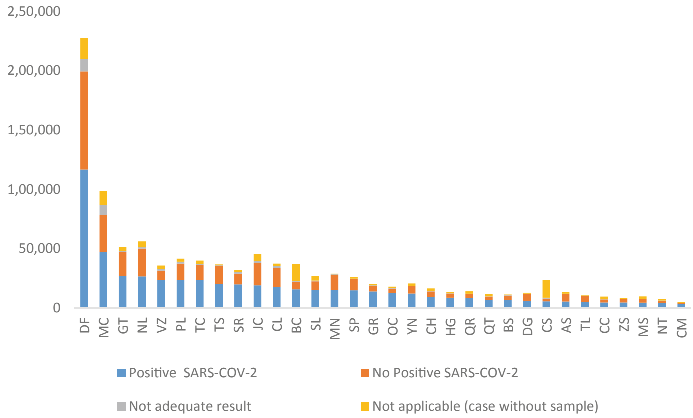

Fig. 2. Total cases registered as of October 2020 by state. Source: Ministry of Health of Mexico.

The distribution of cases by age group shows a concentration of a greater number of cases in people between 25 and 54 years of age, concentrating 63% of the total confirmed cases. The lowest incidence of cases is observed in ages 0 to 19 years with less with a participation of 4% of all confirmed cases (Fig. 3).

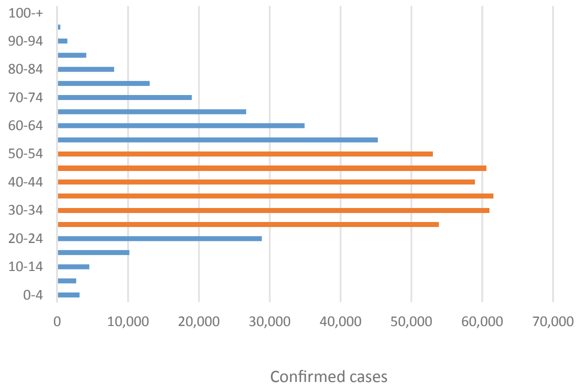

Fig. 3. Total confirmed as of October 2020, cases per five-year age. Source: Ministry of Health of Mexico.

Mortality by age groups is more vulnerable to age groups over 60 years of age, with an average rate of 407 deaths per thousand inhabitants. The lowest mortality rates are found in young population groups aged 0 to 34 years with an average of 14 deaths per thousand inhabitants (Fig. 4). It is important to identify the populations at risk by age and based on this information to be able to focus efforts by the health system.

It is clear that the pandemic in Mexico has not reached its peak and is far from having been overcome. For this reason, it is important to analyze possible scenarios and action strategies using technically robust tools to be able to solve in the best possible way this great challenge for the Mexican health system. It is important to emphasize that

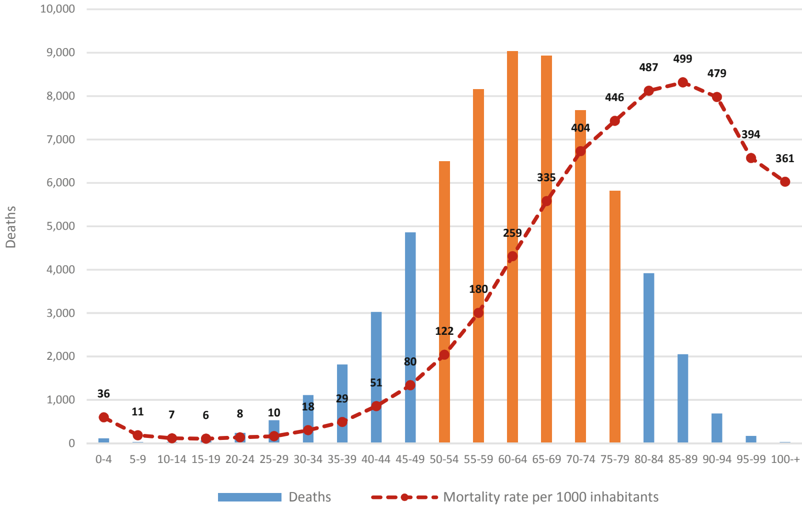

Fig. 4. Mortality rate per thousand inhabitants per five-year age. Source: Ministry of Health of Mexico.

the information used is the available information published by the Ministry of Health, which could be modified due to the constant registration and validation activities. In addition to the incidence of cases, it is important to take into account the severity of these and their demand for specialized health services, according to the information from the Ministry of Health of the total of confirmed patients as of October 2020, 74% correspond to outpatient cases and 26% to cases that required hospitalization. Of the total number of hospitalized patients, 9% required the use of intensive care units and 19% required to be intubated.

3 Materials and Methods

The objective of the work is to identify high priority areas of attention for the COVID-19 pandemic throughout the Mexican territory, it seeks to integrate information regarding the evolution of the pandemic as well as variables related to the state of the health system of each federative entity, taking into account the availability of infrastructure, human and financial resources, as well as epidemiological aspects of the populations served.

3.1 Description of the Data

The information used corresponds to the last available cut-off regarding resources and infrastructure, there is a lag in the availability of information on the health sector, so the latest information available on these items corresponds to the year 2019. Additionally, corresponding variables were integrated to the evolution of the pandemic, for which there are records from January to October 2020. Table 1 shows the variables considered according to each item of analysis, four essential items are considered: 1) health resources, 2) services provided, 3) health spending and 4) evolution of the pandemic.

Table 1. Variables considered in five dimensions of analysis.

Variable	Description
TPS	Total public spending on health
DN	Doctors and nurses in contact with the patient
MU	Medical units
BDS	Beds in the hospitalization area
ICU	Intensive care units
HD	Hospital discharges
NDD	Noncommunicable disease discharges
CPC	COVID-19 positive cases
CDTH	COVID-19 deaths

The information considered is key for estimating the groups to be served with higher priority according to their structural characteristics, in addition to the incidence and deaths of patients from COVID-19. The disaggregation of information is only available at the state level for all variables, it would be ideal to have more disaggregated information at the municipality level, but this proposal can be a reference for future studies that allow a higher level of disaggregation.

3.2 Spectral Cluster Analysis

The use of an unsupervised machine learning model is proposed, because there is no previous antecedent of a classification of this nature. The objective is to identify specific patterns according to the profiles of the states in relation to the severity of the pandemic in those states. Within the available cluster analysis methodologies, the spectral cluster was considered as a classification method given its methodological characteristics that allow the identification of a data structure that does not necessarily present groupings with no conventional boundaries. One of the advantages of the spectral cluster is that it uses a similarity measure which is less restrictive than a distance measure, this allows generating more flexible clusters.

The spectral grouping method uses the information from the data similarity matrix to perform the dimensionality reduction. Using the similarity matrix as input, each pair of points in the data set is evaluated. This methodology considers that the points to be grouped are not part of a space vector. The variables to consider or the attributes are incorporated in a single dimension, the similarity, which takes a numerical value for each pair of points (i, j). Given a set of data, the similarity matrix can be defined as a symmetric matrix A, where $A_{ij} \geq 0$ represents a measure of similarity between data points (i, j). The general approach for spectral clustering is to use a standard clustering method (usually K-means or K-medoids) on relevant eigenvectors of a Laplacian matrix of A. There are several ways to define a Laplacian with different mathematical interpretations, so spectral groupings will also have different interpretations. Relevant eigenvectors are those that correspond to several smaller Laplacian eigenvalues. The main goal of spectral clustering is to find the global clustering of the data set that emerges from the pairwise interactions of its points. Namely, we want to put points that are similar to each other in the same cluster, dissimilar points in different clusters [12].

The concepts based on graph theory is generally used for the spectral cluster. Let the points to be grouped be $V = (1, \ldots, n)$ nodes of a graph G, and the graph edges be represented by the pairs (i, j) with $A_{ij} \geq 0$. Similarity is the weight of edge ij.

$$G = (V, E), E = \left\{(i, j), A_{ij} > 0\right\} \subseteq V\,x\,V \tag{1}$$

G is an undirected and weighted graph. A partition of nodes of the graph in K participles or clusters was known as (K-way) graph cut. Therefore, the spectral clustering can be seen as the search for slices in the graph G. In addition to the similarity matrix, there is a set of derived matrices that are a fundamental part of the spectrum cluster methodology. One of the main matrices used is the Normalized Laplacian matrix, which is defined as follows.

$$L_{(n\times n)} = I - D^{-1/2} A D^{-1/2} \tag{2}$$

Where, A is a similarity matrix, $D = diag(d_1, \ldots, d_n)$ is a diagonal matrix for the node degrees and I is the unit matrix and $P = D^{-1}A$ is a random walk matrix. Based on the information captured by the similarity matrix and the Laplacian matrix, the objective of the spectral cluster is to project the information in a space of less dimension where the segmentation of the groups is natural and allows groups to be adjusted to unconventional forms. A generalized version of the spectral clustering algorithm is shown below [13, 14].

Algorithm Spectral Clustering
Input Similarity matrix A, number of clusters K

1. Transform A
   Calculate $di \leftarrow \sum_{j=1}^{n} A_{ij}, j = 1{:}\ n$ *the node degrees.*
   From the transition matrix P with $P_{ij} \leftarrow \frac{A_{ij}}{d_i}, for\ i, j = 1{:}\ n$
2. Eigen-decomposition
   Compute the largest K eigenvalues $\lambda_1 \geq \cdots \geq \lambda_K$ and eigenvectors $V_1, \ldots V_K$ of P.
3. Embed the data in $K - th$ principal subspace
   Let $x_i = [V_{i2}\, V_{i3} \ldots V_{iK}] \in R^{n\times(K-1)}, for\ i = 1, \ldots\ n.$
4. Run the K-means algorithm of the data $x_{1:n}$
   Output the clustering C obtained in step 4.

Among the main applications of the spectral cluster methodology are image recognition [15], image segmentation [16], urban land use [17], network community detection [18], speech separation [19], and demand response applications [20].

4 Results

4.1 Priorization of States for COVID-19 Attention

When estimating the similarity matrix and the Laplacian matrix of these, it is possible to extract the eigenvectors that represent the projection of the data in a smaller dimension, in this case it is observed that with the first two eigenvectors the greater amount of information of the dimension is represented original, it is decided to keep the first two dimensions of the spectral decomposition. Figure 5 shows the eigenvalues of each eigenvector of the Laplacian graph. In this case, the largest gap is between the eigenvectors 2 and 3, so a K = 4 would correspond. But nevertheless, in this case, three partitions were selected to be able to consider a pandemic monitoring semaphore in terms of selecting those states that represent the highest priority, entities on alert and entities without significant risks.

The first three dimensional eigenvectors of the spectral decomposition will be used to run the K-means algorithm with three partitions. With the information generated from the spectral cluster, three partitions were generated using a K-means algorithm, and the t-SNE 3D algorithm (t-Distributed Stochastic Neighbor Embedding) was applied to observe the separation in the projected space through spectral decomposition, we can observe that the three groups sought are clearly segmented in a three-dimensional space (Fig. 6).

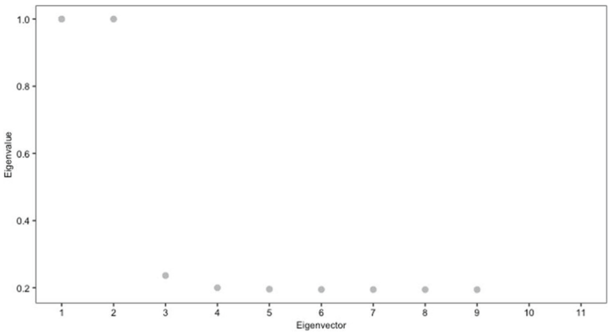

Fig. 5. Eigenvalues of each eigenvector of the Laplacian graph.

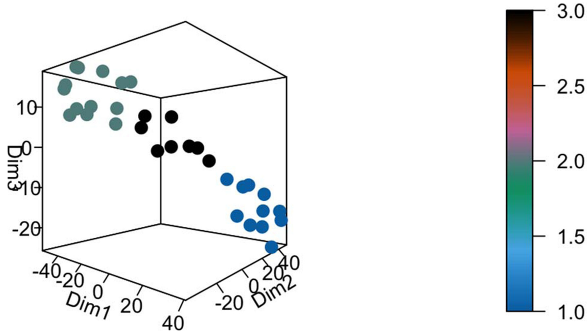

Fig. 6. t-SNE 3D graph for the three partitions.

Table 2. Profile of the clusters generated (average values for each variable)

Cluster	States	TPS*	DN	MU	BDS	ICU	HD	NDD	CPC	CDTH
I	AS, BS, CC, CM, DG, MS, NT, QT, QR, TL, ZS	7,117,448	6,466	298	912	5	69,879	27,985	5,336	5,356
II	BC, CS, CH, DF, GT, JC, MC, MN, NL, PL, SR, TS, VZ	34,386,594	28,307	981	4,849	23	288,100	134,068	28,428	28,744
III	CL, GR, HG, OC, SP, SL, TC, YN	14,060,391	13,413	810	2,062	11	143,021	57,138	14,838	14,869

* Thousands of pesos.

According to the three groups generated, it can be seen that they meet the objective, states with the highest incidence and mortality in COVID-19 are grouped, which present problems of resource availability as well as high values in associated diseases as a risk factor. Table 2 shows the profiles identified for each generated group, group I contains 11 states, group I is made up of 13 states, and group III contains 8 states. Group I corresponds to the states with lower risk, these are states with low health resources but also with low incidence and deaths from COVID-19, an interesting aspect of this group is the low incidence in discharges related to chronic diseases non-transmissible, in addition to being relatively small states in size and population, so the risk of contagion due to high population density is lower. Group II corresponds to the largest entities that in turn concentrate greater resources in health and greater risks related to chronic non-communicable diseases, in addition to presenting the highest incidence and deaths from

COVID-19, in this case a particular characteristic in Mexico is since the flow of patients is generally concentrated in the largest cities, the data corresponding to confirmed cases and deaths from COVID-19 correspond to the entity where the medical unit where the patient was treated is located and not to their state of origin. Group III corresponds to entities with lower risk with moderate health resources as well as confirmed patients and deaths from COVID-19.

It is important to note that the results generated are based on the information available on the pandemic so far, so they are subject to the quality of the information available. However, the methodological structure is applicable for detailed and reliable information that may not be public at the moment.

5 Conclusions

The pandemic that we are currently experiencing has generated the need to propose and implement innovative ideas for its management and attention to the problems related to the attention of COVID-19. Mexico is no exception since the rate of infections and deaths has been increasing since the first months. A fundamental element for the correct management and overcoming of this crisis is the efficient and timely allocation of health resources that allow to face this crisis.

Various studies have been developed worldwide in relation to the care of the health crisis, from different perspectives, most of them focused on the clinical care of the pandemic, but a significant proportion seeks to solve the problems directly and indirectly linked to the health crisis, such as streamlining logistics processes to ensure timely access to health resources, the development of simulation models and artificial intelligence that allow defining the most efficient and effective strategies to contain contagions as well as to avoid saturation of health services. It is precisely in this sense that the present work seeks to provide technical evidence for decision-making, integrating an unsupervised support model based on spectral clustering.

The application of a spectral clustering model makes it possible to address the grouping of states in the Mexican republic, allowing the integration of structures with unconventional borders, by projecting an n-dimensional space in a low-dimensional space through the use of the similarity matrix and the Laplace matrix. The results generated show congruence, three groups are identified based on the availability of health resources as well as the risk due to the growing demand for medical care and deaths from COVID-19, identifying the priority states will allow a more efficient and targeted allocation of resources to be able to act strategically in the face of the current contingency. In the same way, the results generated can serve as antecedent for a much more disaggregated grouping and with a greater number of variables that allow the allocation of resources to be focused even more. However, it is necessary to have detailed information in a timely manner to achieve this goal.

Annex

See Table 3.

Table 3. Catalog of federative entities of the Mexican Republic.

State	Abbreviation
AGUASCALIENTES	AS
BAJA CALIFORNIA	BC
BAJA CALIFORNIA SUR	BS
CAMPECHE	CC
COAHUILA DE ZARAGOZA	CL
COLIMA	CM
CHIAPAS	CS
CHIHUAHUA	CH
CIUDAD DE MÉXICO	DF
DURANGO	DG
GUANAJUATO	GT
GUERRERO	GR
HIDALGO	HG
JALISCO	JC
MÉXICO	MC
MICHOACÁN DE OCAMPO	MN
MORELOS	MS
NAYARIT	NT
NUEVO LEÓN	NL
OAXACA	OC
PUEBLA	PL
QUERÉTARO	QT
QUINTANA ROO	QR
SAN LUIS POTOSÍ	SP
SINALOA	SL
SONORA	SR
TABASCO	TC
TAMAULIPAS	TS
TLAXCALA	TL
VERACRUZ	VZ
YUCATÁN	YN
ZACATECAS	ZS

References

1. World Health Organization. (2020). Maintaining essential health services: operational guidance for the COVID-19 context: interim guidance, 1 June 2020. World Health Organization
2. Baveja, A., Kapoor, A., Melamed, B.: Stopping Covid-19: a pandemic-management service value chain approach. Ann. Oper. Res. **289**, 173–184 (2020)
3. Noronha, K.V, et al.: Pandemia por COVID-19 no Brasil: análise da demanda e da oferta de leitos hospitalares e equipamentos de ventilação assistida segundo diferentes cenários. Cadernos de Saúde Pública, **36**(6), e00115320. Epub 17 June 2020
4. McMahon, M., Nadigel, J., Thompson, E., Glazier, R.H.: Informing Canada's Health System Response to COVID-19: Priorities for Health Services and Policy Research. Healthcare policy = Politiques de sante, **17**(1), 112–124 (2020)
5. Hussain, S., Xie, Y., Li, D., et al.: Current strategies against COVID-19. Chin. Med. **15**, 70 (2020)
6. Scarabel, F., Pellis, L., Bragazzi, N.L., Wu, J.: Canada needs to rapidly escalate public health interventions for its COVID-19 mitigation strategies. Infect. Dis. Model. **5**, 316–322 (2020)
7. Lozano-Diez, J.A., MS, J.A., Rodriguez-Aguilar, R.: Designing a resilient supply chain: an approach to reduce drug shortages in epidemic outbreaks. PHAT, EAI (2020)
8. Shuman, A.G., Fox, E., Unguru, Y.: Preparing for COVID-19-related drug shortages. Ann. Am. Thorac. Soc. **17**(8), 928–931 (2020)
9. Adly, A.S., Adly, A.S., Adly, M.S.: Approaches based on artificial intelligence and the internet of intelligent things to prevent the spread of COVID-19: scoping review. J. Med. Internet Res. **22**(8), e19104 (2020)
10. Nelsen, G., et al.: Considerations for development of pharmacy support models for COVID-19 alternate care sites. Am. J. Health-Syst. Pharm. **77**(19), 1592–1597 (2020)
11. Currie, C.S.M., et al.: How simulation modelling can help reduce the impact of COVID-19. J. Simul. **14**(2), 83–97 (2020)
12. Meila, M.: Spectral Clustering: A Tutorial for the 2010's, pp. 1–23. CRC Press, Boca Raton (2016)
13. Meila, M., Shi, J.: Learning segmentation by random walks. Adv. Neural. Inf. Process. Syst. **13**, 873–879 (2001)
14. Meila, M., Shi, J.: A random walks view of spectral segmentation. In: Jaakkola, T., Richardson, T., (eds.), Artificial Intelligence and Statistics AISTATS (2001b)
15. Tolliver, D.A., Miller, G..: Graph partitioning by spectral rounding: applications in image segmentation and clustering. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), New York, NY, USA, 2006, pp. 1053–1060 (2006)
16. Schultz, T., Kindlmann, G.L.: Open-box spectral clustering: applications to medical image analysis. IEEE Trans. Vis. Comput. Graph. **19**(12), 2100–2108 (2013)
17. Frias-Martinez, V., Frias-Martinez, E.: Spectral clustering for sensing urban land use using twitter activity. Eng. Appl. Artif. Intell. **35**, 237–245 (2014)
18. Shen, G., Ye, D.: A distance-based spectral clustering approach with applications to network community detection. J. Ind. Inf. Integr. **6**, 22–32 (2017)
19. Bach, F.R., Jordan, M.: Learning spectral clustering, with application to speech separation. J. Mach. Learn. Res. **7**(71), 1963–2001 (2006)
20. Lin, S., Li, F., Tian, E., Fu, Y., Li, D.: Clustering load profiles for demand response applications. IEEE Trans. Smart Grid **10**(2), 1599–1607 (2019)