Analysis of Spatial Spread Relationships of Coronavirus (COVID-19) Pandemic in the World using Self Organizing Maps

Patricia Melin*, Julio Cesar Monica, Daniela Sanchez, Oscar Castillo

Tijuana Institute of Technology, Tijuana, Mexico

A R T I C L E   I N F O

A B S T R A C T

We describe in this paper an analysis of the spatial evolution of coronavirus pandemic around the world by using a particular type of unsupervised neural network, which is called self-organizing maps. Based on the clustering abilities of self-organizing maps we are able to spatially group together countries that are similar according to their coronavirus cases, in this way being able to analyze which countries are behaving similarly and thus can benefit by using similar strategies in dealing with the spread of the virus. Publicly available datasets of coronavirus cases around the globe from the last months have been used in the analysis. Interesting conclusions have been obtained, that could be helpful in deciding the best strategies in dealing with this virus. Most of the previous papers dealing with data of the Coronavirus have viewed the problem on temporal aspect, which is also important, but this is mainly concerned with the forecast of the numeric information. However, we believe that the spatial aspect is also important, so in this view the main contribution of this paper is the use of unsupervised self-organizing maps for grouping together similar countries in their fight against the Coronavirus pandemic, and thus proposing that strategies for similar countries could be established accordingly.

1. Introduction

Recently we have witnessed the rapid spread of the Coronavirus around the globe, beginning originally in China and then spreading to Korea and Japan, and after that to Europe and America. In particular, in the case of Europe, Italy and Spain have been hit very hard with the spread of the virus, having many confirmed cases and deaths. After that, in the American continent, the United States has also been hit very hard with the spread of the virus. So it is very critical understanding all the facets of this problem, for being able to cope with its complexity and at the same limit its negative impact on the health of the population around the world and also the economic implications for the countries.

Due to the importance of finding ways to control the propagation of the virus, many papers have been put forward on these last months related to different aspects of this problem, and in particular several authors have attempted to apply computational intelligence techniques in this area. As a sample of these works we can mention the ones below.

The coronavirus disease (COVID-19) is a viral infection highly transmittable caused by severe acute respiratory syndrome coron-

avirus 2 (SARS-CoV-2), which originally appeared in Wuhan, China, and it has sequentially propagated around the world. The intermediate source of origin and transfer to humans is not known, but the quick human to human transfer has been confirmed in many experiments. Nowadays there is not yet a clinically approved antiviral drug or vaccine that can be used against COVID-19. Recently at the end of 2019, the city of Wuhan, China, the epicenter of the current COVID-19 experienced an outbreak of a novel coronavirus that killed more than eighteen hundred and infected thousands of individuals within the first two months of the epidemic [9]. More recently, the epicenter has moved to other cities in Europe and then in America.

The patients' most notable found symptoms (according to the collected experimental data) are dry cough, dyspnea, fiver and bilateral lung infiltrates on imaging. Initially all the cases were associated to Wuhan's Huanan Seafood Wholesale Market, which trades in seafood and a wide variety of live animal species. Due to the many reported cases up to January 30th 2020, the World Health Organization (WHO) declared the Chinese outbreak of COVID-19 to be a Public Health Emergency of International Concern posing a high risk to countries with vulnerable health systems around the world [10].

There have recently been several studies with the goal of understanding the patterns of COVID-19, and one of this is: using a dataset of X-ray medical images from patients with common bac-

---

* Corresponding author.
E-mail addresses: pmelin@tectijuana.mx (P. Melin), ocastillo@tectijuana.mx (O. Castillo).

teria pneumonia confirmed with COVID-19 disease to identify possible patterns that may lead to the automatic diagnosis disease using convolutional neural networks, and the results demonstrate that the used method has significant effects on the automatic detection and diagnosis of COVID-19 [11]. Another interesting study is the investigation of the cases of COVID-19 in China using dynamic statistical techniques [12]. Other cases are: predicting commercially available antiviral drugs that may act on the novel coronavirus using a deep learning model [13] and early prediction of the 2019 novel coronavirus outbreak in mainland China based on simple mathematical model [14]. Also, the paper in [15] offers pointers to, and describes, a range of practical online/mobile GIS and mapping dashboards and applications for tracking the 2019/2020 coronavirus epidemic and associated events as they unfold around the world. In addition, in [16] the authors proposed applying the concept of cartograms to visualize both the expansion and spread of COVID-19. Finally, we have to mention that some research has been done using Artificial Intelligence (AI), for example the study in [17] in which the authors proposed the use of machine learning algorithms for improving possible case identifications of COVID-19 more quickly when using a mobile phone-based web survey. Also several AI techniques are applied in analyzing data and decision-making processes in healthcare. This means that AI-driven tools can help in identifying COVID-19 outbreaks, as well as forecast their nature of spread rate across the world [18].

However, most of the previous works deal with the temporal aspect of the problem, which means that these works are attempting to predict or forecast in different ways the coronavirus numeric data. Of course, this facet of the problem is also important, as governments want to be able to know the estimated future values of the coronavirus cases to make the right decisions regarding funds to be assigned to solving the problem. On the other hand, it is our firm believe that the spatial aspect is also very important, so in this regard the main contribution of this paper is the use of unsupervised self-organizing Kohonen maps for grouping together similar countries in their fight against the Coronavirus pandemic, and thus in this way be able to propose that strategies for similar countries could be established accordingly. In our opinion, this contribution is very important as it could complement the temporal perspective that has been developed by most of the previous papers by providing the spatial component to achieve a complete solution to the Coronavirus problem.

The remaining contents of the paper are structured in the following form. Section 2 outlines the fundamental concepts of self-organizing maps, which are a particular form of unsupervised neural networks. Section 3 describes the problem at hand and the proposed methodology in this work. Section 4 summarizes the simulation results achieved with the proposed approach. Finally, Section 5 offers the conclusions and possible future works.

2. Self-organizing maps

The Self-organizing maps (SOM), also called the Kohonen map, is a model being used to explore and visualize patterns in high-dimensional datasets. This model was first introduced by Teuvo Kalevi Kohonen in 1982. SOM is a clustering technique that identifies groups in a dataset without having to use traditional statistical techniques. The SOM consists of only two layers: the input layer and the output layer [1]. The goal of this neural network is to transfer all input data objects with *n* attributes (*n* dimensions) to the output in a way that the objects are related to each other. The SOM is based on an unsupervised training where there is no given output target, the objective of the algorithm is to find the set of centroids (neurons) to represent the cluster, but with topological restrictions. Topology refers to a centroid arrangement on the output grid, the most common used topology grids are the hexagonal
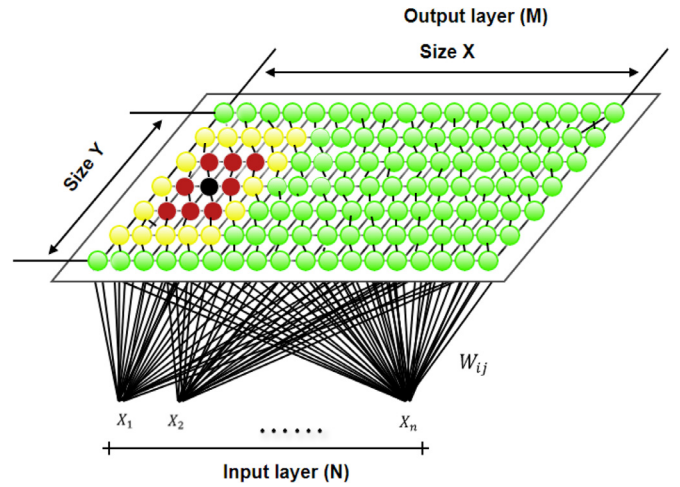

Fig. 1. Example of the SOM neural network general architecture.

and rectangular. Each of all data objects in the dataset is assigned to each centroid. Each neuron in the SOM grid is closely related to each other and each of the inputs are connected to each of the output nodes by means of a connection weight. Weights from N input nodes to M output nodes are initialized in small values randomly [2]. The activation of the output units according to Kohonen's is shown in the Eq. 1. The modification of the weights is shown in Eq. 2:

$$O_j = F_{min} d_j = F_{min} \left(\sum_i \left(X_i - W_{ij} \right)^2 \right) \tag{1}$$

$$\Delta W_i j = O_j \eta \left(X_i - W_{ij} \right) \tag{2}$$

where O_j =activation of output unit j, X_i =activation value from input unit, W_{ji} = lateral weights connecting to output unit, d_j = neurons in neighborhood, F_{min} = unity function returning 1 or 0, η = gain term decreasing over time.

The lateral connections enable the SOM to learn "competitively", meaning that the output neurons in the output layers compete for the classification of the input patterns. At the beginning of the training, the input patterns are presented to the SOM and the output object with the nearest weight vector will be the winner to represent that cluster. Equation 1 shows how the Euclidean distance is used to select the winning neuron [3].

In Figure 1, the SOM neural network structure is illustrated with its neighborhoods around the winning neuron.

Artificial neural networks, such as the SOM have widely been used in many applications, such as for identification of groundwater salinity sources [4], Determination of plant communities based on bryophytes [5], Prediction of arthritis [6]. However, here the SOM is applied to classify 199 countries of the world and the 32 states of Mexico with confirmed cases of the COVID-19 to identify if there is a pattern within the: too high, high, medium and low clusters being used. The world dataset was obtained from the Humanitarian Data Exchange (HDX) [7], and the Mexican dataset from the Mexico's Government website [8].

3. Proposed Method

The Data base used for the experiments was obtained from the Humanitarian Data Exchange (HDX) [7], which includes data from the countries where COVID-19 cases have occurred from January 22, 2020 to May 13, 2020. The consulted datasets were the following: time_series_covid19_confirmed_global, time_series_covid19_recovered_global, and
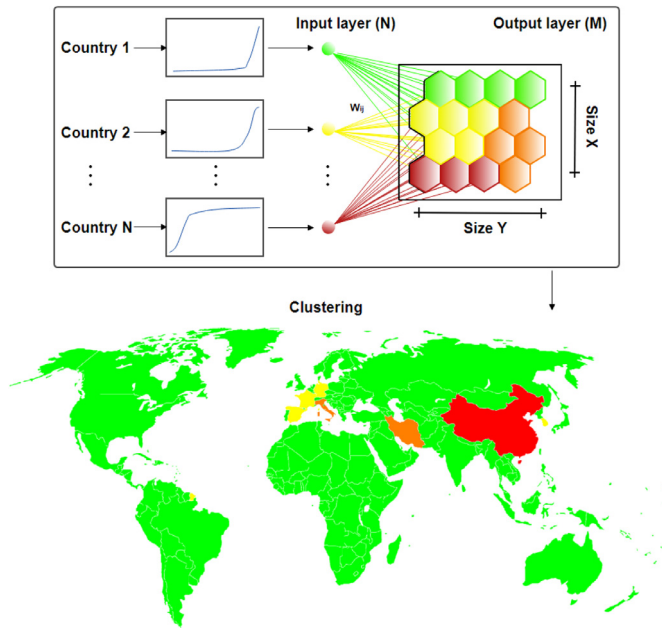
Fig. 2. An example SOM neural network used for clustering and classification of countries.
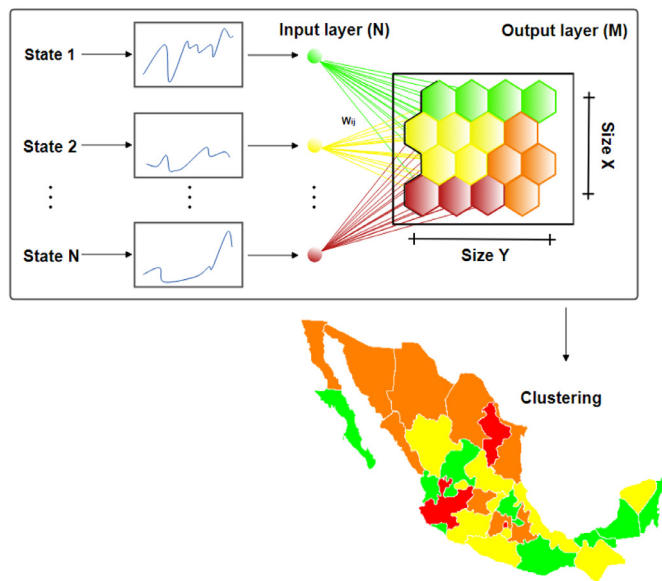

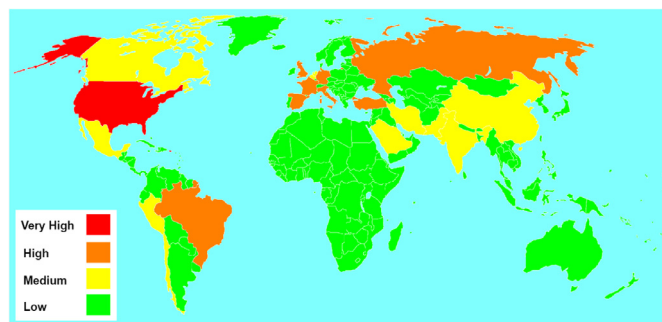**Fig. 3.** Structure of SOM neural network used for clustering the 32 states of Mexico.

Fig. 4. Classification of countries according to confirmed Coronavirus cases.

time_series_covid19_deaths_global. The data includes the confirmed, recovered and deaths cases for countries, respectively.

In Figure 2 a sample of a SOM neural network used for clustering and is classification for the countries is shown.

Also the data set of the 32 states of Mexico was used for illustrating the clustering depending in the similarity patterns the data base was obtained from the Mexican dataset in the Mexico's Government website [8]. Figure 3 shows the structure of the SOM used for clustering the 32 states of Mexico.

In the case of the 32 states of Mexico, two of the most prevalent diseases in the population were also studied, which are hypertension and diabetes. This is in order to find similarities and form groupings by states between the diseases and Covid-19. The database of these diseases was obtained from the open data web page of the Mexican Institute of Social Security (IMSS) [9].

4. Simulation Results

The proposed method based on the Kohonen self-organizing maps was used to form groupings or clusters of countries in the world, and after that their classification was done by considering 4 classes according to the severity of the number of Coronavirus cases: Very High, High, Medium and Low (indicated by red, orange, yellow and green, respectively, in the maps). In Table 1, countries are ordered by number of cases occurring in the clusters, and then alphabetically inside the clusters. In the following Figures we show the obtained results with the proposed method using the publicly available data sets of confirmed, recovered and death cases.

In Figure 4 we show a plot of the clusters formed with the SOM method, clearly indicating the classes for the Covid-19 Confirmed cases for the 22-01-2020 to 13-05-2020 period of time.

In Figure 5 we show a plot of the clusters formed with the SOM method, clearly indicating the classes for the Covid-19 recovered cases for the January 22 of 2020 to May 13 of 2020 period of time.

In addition, the same analysis can be done for the spatial distribution of deaths due to Coronavirus around the globe. In Figure 6 we show a plot of the clusters formed with the SOM
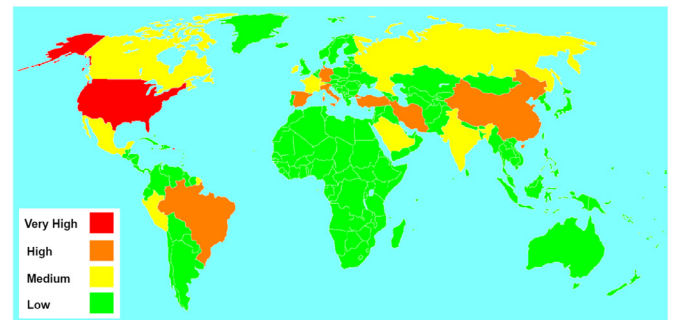

Fig. 5. Classification of countries according to recovered Coronavirus cases.
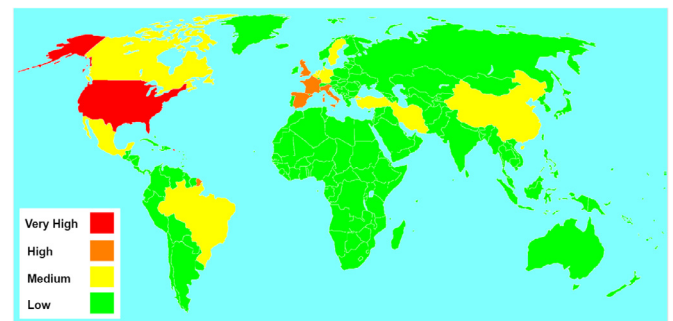

Fig. 6. Classification of countries according to death related Coronavirus cases.
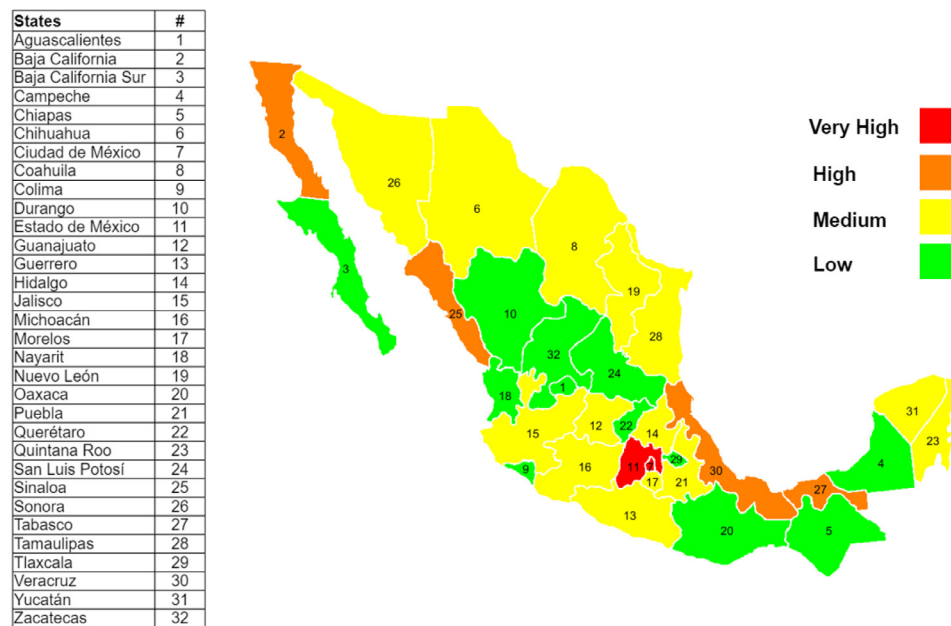
States	#
Aguascalientes	1
Baja California	2
Baja California Sur	3
Campeche	4
Chiapas	5
Chihuahua	6
Ciudad de México	7
Coahuila	8
Colima	9
Durango	10
Estado de México	11
Guanajuato	12
Guerrero	13
Hidalgo	14
Jalisco	15
Michoacán	16
Morelos	17
Nayarit	18
Nuevo León	19
Oaxaca	20
Puebla	21
Querétaro	22
Quintana Roo	23
San Luis Potosí	24
Sinaloa	25
Sonora	26
Tabasco	27
Tamaulipas	28
Tlaxcala	29
Veracruz	30
Yucatán	31
Zacatecas	32

Fig. 7. Classification of states in Mexico according to confirmed Coronavirus cases.

Table 1
The results of confirmed cases of Covid-19 around the world (up to May 13, 2020).

Cluster	Country	Value
Very High	United States (US)	1390361
High	Brazil	189157
	France	178184
	Germany	174098
	Italy	222104
	Russia	242271
	Spain	228691
	Turkey	143114
	United Kingdom	230986
Medium	Belgium	53981
	Canada	73568
	Chile	34381
	China	84024
	India	78055
	Iran	112725
	Mexico	40186
	Netherlands	43410
	Pakistan	35298
	Peru	76306
	Saudi Arabia	44830
Low	Afghanistan	5226
	Albania	880
	Algeria	6253
	…	…

Table 2
The results of confirmed cases of Covid-19 in the states of Mexico (up to May 13, 2020).

Cluster	State	Value
Very High	Ciudad de México	10946
	Estado de México	6813
High	Baja California	2764
	Sinaloa	1620
	Tabasco	1976
	Veracruz	1574
Medium	Chihuahua	768
	Coahuila	616
	Guanajuato	580
	Guerrero	670
	Hidalgo	637
	Jalisco	699
	Michoacán	678
	Morelos	915
	Nuevo León	717
	Puebla	1213
	Quintana Roo	1177
	Sonora	642
	Tamaulipas	799
	Yucatán	924
Low	Aguascalientes	398
	Baja California Sur	409
	Campeche	226
	Chiapas	450
	Colima	46
	Durango	127
	Nayarit	252
	Oaxaca	291
	Querétaro	315
	San Luis Potosí	338
	Tlaxcala	438

method, clearly indicating the classes for the Covid-19 death cases for the January 22 of 2020 to May 13 of 2020 period of time.

We were also interested in taking down this spatial analysis to the country level, and for this we applied it to the country of Mexico. In this case, we have to consider 32 states in Mexico and the SOM method clusters states according to their similarities to other states, finding in this way a colored map similar to the world map. In Figure 7 we can find the clustering of states in Mexico according to the confirmed Coronavirus cases during the period of time from February 27 of 2020 to May 13 of 2020. In Table 2, states of Mexico are ordered by number of cases in the clusters, and then alphabetically inside the clusters.

In addition, the same analysis can be done for the spatial distribution of deaths due to Coronavirus in the states of Mexico. In Figure 8 we show a plot of the clusters formed with the SOM
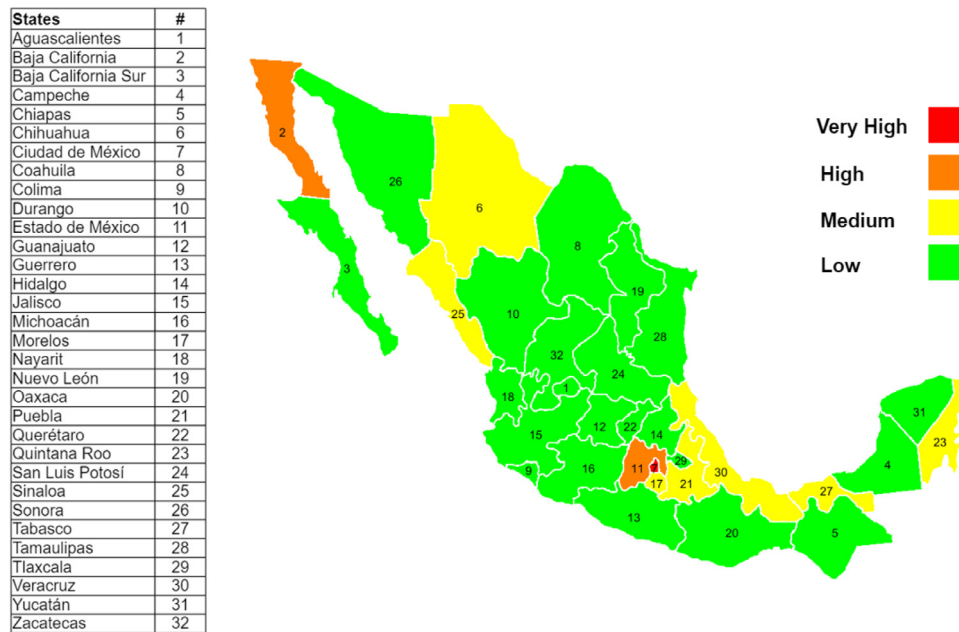
States	#
Aguascalientes	1
Baja California	2
Baja California Sur	3
Campeche	4
Chiapas	5
Chihuahua	6
Ciudad de México	7
Coahuila	8
Colima	9
Durango	10
Estado de México	11
Guanajuato	12
Guerrero	13
Hidalgo	14
Jalisco	15
Michoacán	16
Morelos	17
Nayarit	18
Nuevo León	19
Oaxaca	20
Puebla	21
Querétaro	22
Quintana Roo	23
San Luis Potosí	24
Sinaloa	25
Sonora	26
Tabasco	27
Tamaulipas	28
Tlaxcala	29
Veracruz	30
Yucatán	31
Zacatecas	32

Fig. 8. Classification of states in Mexico according to death Coronavirus cases.

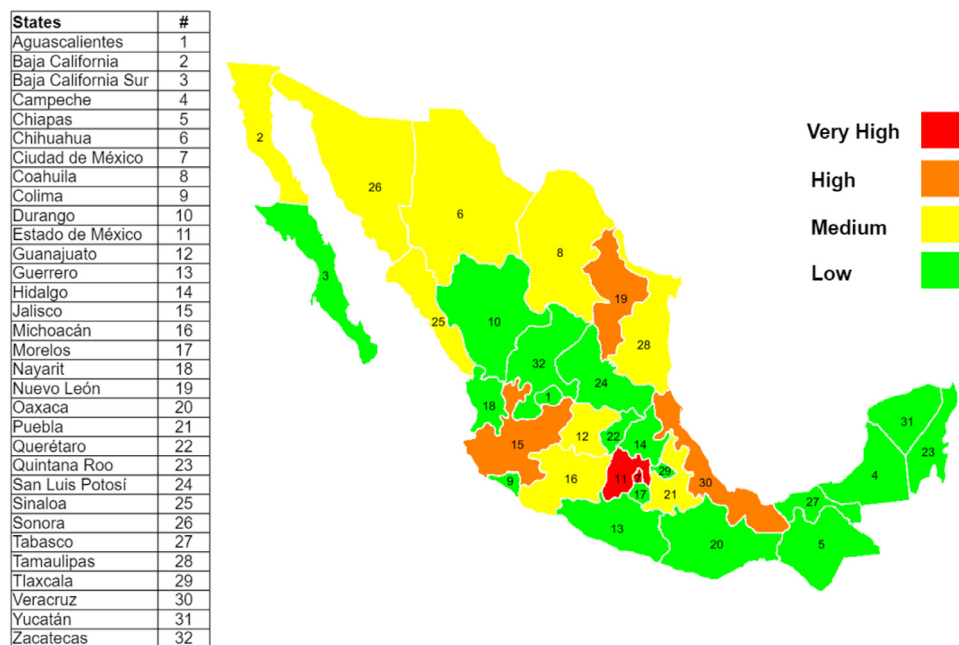| States	#
Aguascalientes	1
Baja California	2
Baja California Sur	3
Campeche	4
Chiapas	5
Chihuahua	6
Ciudad de México	7
Coahuila	8
Colima	9
Durango	10
Estado de México	11
Guanajuato	12
Guerrero	13
Hidalgo	14
Jalisco	15
Michoacán	16
Morelos	17
Nayarit	18
Nuevo León	19
Oaxaca	20
Puebla	21
Querétaro	22
Quintana Roo	23
San Luis Potosí	24
Sinaloa	25
Sonora	26
Tabasco	27
Tamaulipas	28
Tlaxcala	29
Veracruz	30
Yucatán	31
Zacatecas	32

Fig. 9. Classification of states in Mexico according to the number of Hypertension cases.

method, clearly indicating the classes for the Covid-19 death cases for the February 27 of 2020 to May 13 of 2020 period of time

In this case, we were also interested in the possible relation of propensity of Coronavirus deaths to the chronic degenerative Hypertension and Diabetes diseases. Based on this, we also applied SOM clustering to the publicly available data in Mexico of these cases [18,19] In Figure 9 we can find the results of clustering the states of Mexico according to the number of Hypertension cases from 2000 to 2018.

If we compare Figures 8 and 9 we can find that there is a similarity between states with higher number of deaths to the states

with higher number of Hypertension cases, confirming a relation between these variables.

In addition, in Figure 10 we can find the results of clustering the states of Mexico according to the number of Diabetes cases from 2000 to 2018.

Once again, if we compare Figures 8 and 10 we can find that there is a similarity between states with higher number of deaths to the states with higher number of Diabetes cases, confirming a relation between these variables. In this regard, we believe a model could be constructed using the number of cases of hypertension and diabetes to estimate the number of
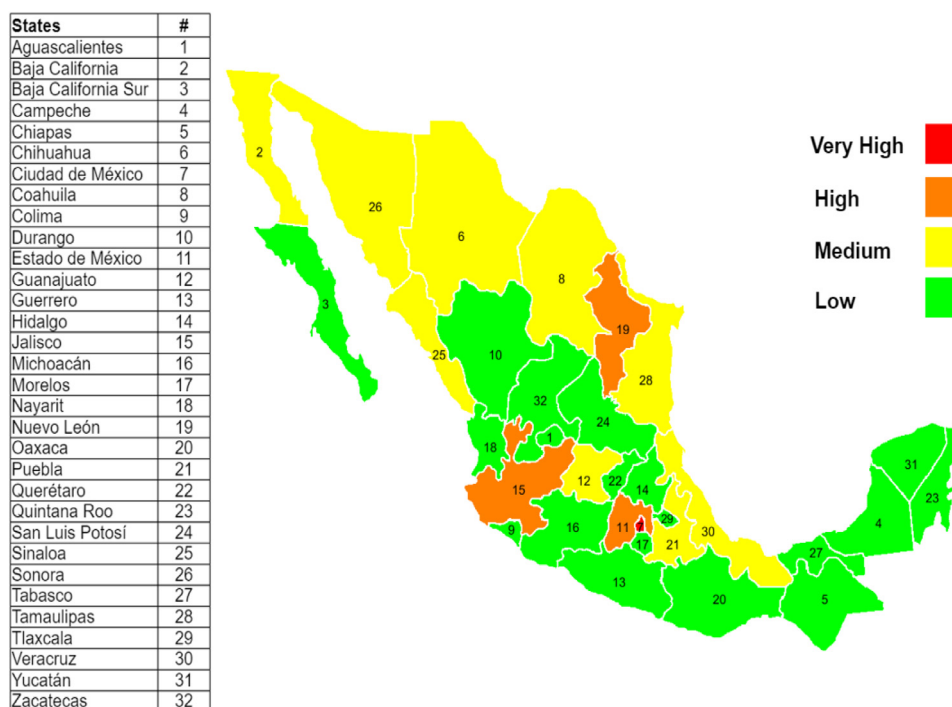
States	#
Aguascalientes	1
Baja California	2
Baja California Sur	3
Campeche	4
Chiapas	5
Chihuahua	6
Ciudad de México	7
Coahuila	8
Colima	9
Durango	10
Estado de México	11
Guanajuato	12
Guerrero	13
Hidalgo	14
Jalisco	15
Michoacán	16
Morelos	17
Nayarit	18
Nuevo León	19
Oaxaca	20
Puebla	21
Querétaro	22
Quintana Roo	23
San Luis Potosí	24
Sinaloa	25
Sonora	26
Tabasco	27
Tamaulipas	28
Tlaxcala	29
Veracruz	30
Yucatán	31
Zacatecas	32

Fig. 10. Classification of states in Mexico according to the number of Diabetes cases.

Coronavirus cases, that could reflect the interaction among these variables.

5. Conclusions

In this paper an analysis of the spatial evolution of coronavirus pandemic around the world by using a particular type of unsupervised neural network was presented. Based on the clustering abilities of self-organizing maps we were able to spatially group together countries that are similar according to their coronavirus cases, in this way being able to analyze which countries are behaving similarly and thus can benefit by using similar strategies in dealing with the spread of the virus. Publicly available datasets of coronavirus cases around the globe from the last months were used in the analysis. Interesting conclusions have been obtained, that could be helpful in deciding the best strategies in dealing with this virus. In addition, the proposed approach was tested with the spatial distribution of cases around the country of Mexico and its relation to the Diabetes and Hypertension cases. Most of the previous papers dealing with data of the Coronavirus have viewed the problem on its temporal aspect, which is also important, but this is mainly concerned with the forecast of the numeric information. However, we believe that the spatial aspect is also important, so in this view the main contribution of this paper is the use of unsupervised self-organizing maps for grouping together similar countries in their fight against the Coronavirus pandemic, and thus proposing that strategies for similar countries could be established accordingly. As future work, we envision integrating both the spatial and temporal aspects of the Coronavirus spread problem in a unified manner to achieve a complete view and solution to the problem. We can also consider applying other intelligent techniques (like fuzzy logic, evolutionary algorithms and swarm intelligence) that could help in dealing in a better way with this complex problem. Finally, we could also consider other recent approaches, as the ones presented in [20, 21], and other recent interesting works related to evolving fuzzy models and chaos, like in [22-26]. In summary, we envision that there are many potential beneficial lines of research that could be engaged.

Declaration of competing interest

The authors of the above manuscript whose names are listed above certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

CRediT authorship contribution statement

Patricia Melin: Methodology, Data curation, Writing - review & editing. **Julio Cesar Monica:** Formal analysis, Methodology, Writing - review & editing. **Daniela Sanchez:** Validation, Writing - review & editing. **Oscar Castillo:** Formal analysis, Data curation, Writing - review & editing.

References

[1] Mostafa MM. Clustering the ecological footprint of nations using Kohonen's self-organizing maps. Expert Systems with Applications 2010;37(4):2747–55.
[2] Kotu V, Deshpande B. Data Science: Concepts and Practice. Morgan Kaufmann; 2018.
[3] Malone J, McGarry K, Wermter S, Bowerman C. Data mining using rule extraction from Kohonen self-organising maps. Neural Computing & Applications 2006;15(1):9–17.
[4] Haselbeck V, Kordilla J, Krause F, Sauter M. Self-organizing maps for the identification of groundwater salinity sources based on hydrochemical data. Journal of Hydrology 2019;576:610–19.
[5] Wolski GJ, Kruk A. Determination of plant communities based on bryophytes: The combined use of Kohonen artificial neural network and indicator species analysis. Ecological Indicators 2020;113:106160.
[6] Wyns B, Boullart L, Sette S, Baeten D, Hoffman I, De Keyser F. Prediction of arthritis using a modified Kohonen mapping and case based reasoning. Engineering Applications of Artificial Intelligence 2004;17(2):205–11.

[7] "The Humanitarian Data Exchange (HDX)," [Online]. Available: https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases. [Accessed 13 05 2020].

[8] "Gobierno de Mexico," [Online]. Available: https://www.gob.mx/salud/documentos/coronavirus-covid-19-comunicado-tecnico-diario-238449. [Accessed 13 05 2020].

[9] Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. Journal of Advanced Research 2020;24:91–8.

[10] Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Agha R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery. 2020;76:71–6.

[11] Apostolopoulos, I. D., & Bessiana, T. (2020). Covid-19: Automatic detection from X-Ray images utilizing Transfer Learning with Convolutional Neural Networks. arXiv preprint arXiv:2003.11617.

[12] Sarkodie, S. A., & Owusu, P. A. (2020). Investigating the Cases of Novel Coronavirus Disease (COVID-19) in China Using Dynamic Statistical Techniques. Available at SSRN 3559456.

[13] Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Computational and Structural Biotechnology Journal 2020;18:784–90.

[14] Zhong L, Mu L, Li J, Wang J, Yin Z, Liu D. Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China based on Simple Mathematical Model. IEEE Access 2020;8:51761–9.

[15] Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr 2020;19:8. doi:10.1186/s12942-020-00202-8.

[16] Gao P, Zhang H, Wu Z, Wang J. Visualising the expansion and spread of coronavirus disease 2019 by cartograms. Environment and Planning A 2020. doi:10.1177/0308518-20910162.

[17] Rao ASRS, Vazquez JA. Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when Cities/Towns are under quarantine. Infection Control and Hospital Epidemiology, 2020. doi:10.1017/ice.2020.61.

[18] Santosh KC. AI-driven tools for coronavirus outbreak: Need of active learning and cross-population Train/Test models on Multitudinal/Multimodal data. Journal of Medical Systems 2020;44(5). doi:10.1007/s10916-020-01562-1.

[19] "Datos Abiertos IMMS" [Online]. Available http://datos.imss.gob.mx/ [Accessed 01 04 2020]

[20] Robson B. Computers and viral diseases. Preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus. Computers in Biology and Medicine 2020;119:1–19.

[21] Fanelli D, Piazza F. Analysis and Forecast of COVID-19 spreading in China, Italy and France. Chaos, Solitons and Fractals 2020;134:1–5.

[22] Dima, G.C., Copelli, M., Mindlin, G.B. (2018) Anticipated Synchronization and Zero-Lag Phases in Population Neural Models International Journal of Bifurcation and Chaos, 28 (8), art. no. 1830025

[23] Gil, R.P.A., Johanyák, Z.C., Kovács, T. Surrogate model based optimization of traffic lights cycles and green period ratios using microscopic simulation and fuzzy rule interpolation (2018) International Journal of Artificial Intelligence, 16 (1), pp. 20-40.

[24] Precup, R., T. Teban, A. Albu, A. Borlea, I. A. Zamfirache and. Petriu E. M, "Evolving Fuzzy Models for Prosthetic Hand Myoelectric-based Control," in IEEE Transactions on Instrumentation and Measurement, doi: 10.1109/TIM.2020.2983531

[25] Precup R-E, Teban T-A, Albu A, Borlea A-B, Zamfirache IA, Petriu EM. Evolving fuzzy models for prosthetic hand myoelectric-based control using weighted recursive least squares algorithm for identification. In: IEEE International Symposium on Robotic and Sensors Environments, ROSE 2019 Proceedings; 2019. art. no. 8790416.

[26] Sanchez MA, Castillo O, Castro JR, Melin P. Fuzzy granular gravitational clustering algorithm for multivariate data. Information Sciences 2014;279:498–511.