Research Article
Journal of Intelligient Systems: 7(2) (2024) 95-101
DOI: 10.38016/jista.1365609

SOM Clustering of OECD Countries for COVID-19 Indicators and Related Socio-economic Indicators

Pakize Yiğit[1*]

[1] İstanbul Medipol University,Medical School, Department of Medical Statistics and Medical Informatics, İstanbul, Türkiye

pyigit@medipol.edu.tr

Abstract

The coronavirus disease is one of the most severe public health problems globally. Governments need policies to better cope with the disease, so policymakers analyze the country's indicators related to the pandemic to make proper decisions. The study aims to cluster OECD (Organisation for Economic Co-operation and Development) countries using COVID-19, health, socioeconomic, and environmental indicators. A self-organizing map (SOM) clustering method, an unsupervised artificial neural network (ANN) method and a hierarchical clustering method are used. The data comprises 38 OECD countries, and 16 different variables are selected. As a result, the countries are grouped into 3 clusters. Cluster 1 contains 33 countries, the USA is Cluster 2, and Cluster 3 has 4 countries, including Turkey. COVID-19 mortality is highly related to mortality from chronic respiratory diseases. In addition, environmental indicators show differences in clusters.

Keywords: COVID-19, OECD countries, Kohonen SOM, clustering

1. Introduction

The recent coronavirus disease (COVID-19) pandemic is one of the serious public health problems in the World, causing 6,905,763 deaths worldwide on August 10th, 2020 (Worldometer, 2023). The World Health Organization (WHO) officially declared it as a Public Health Emergency of International Concern (PHEIC) on 30 January 2020. After three years, on 5 May 2023, the WHO Emergency Committee on the pandemic accepted that the disease did not fit a PHEIC. However, they warned that the condition is not over, so they continue giving suggestions to countries on how to manage the disease at the current time (WHO, 2023a).

The impacts of the pandemic caused difficulties for countries, especially in the field of health and economic systems. However, there were huge differences between countries reporting COVID-19 cases and death statistics due to these systems. Examining the variations between the countries is crucial for controlling the disease and reducing its burden (Gohari et al., 2022; WHO, 2023b).

Several studies have investigated the variations between the countries in terms of COVID-19 and related indicators. The studies examine the differences using different features such as social inequality and disease prevalence (Cardoso et al., 2020; Islam et al., 2021; Kumru et al., 2022), age and gender differences(Calderón-Larrañaga et al., 2020; Gebhard et al., 2020), environmental factors (Rizvi et al., 2021), lifestyle habits (smoking prevalence, alcohol consumption) (Aydin and Yurdakul, 2020; Kumru et al., 2022; Rizvi et al., 2021), health expenditures (Khan et al., 2020; Micah et al., 2021), healthcare capacity (Khan et al., 2020) and so many different aspects.

Furthermore, researchers have mostly used cluster analysis to examine differences between countries according to COVID-19 variables. Hussein and Abdulazeez (Hussein and Abdulazeez, 2021) reviewed the clustering algorithms applied to COVID-19 pandemic data. They stressed that K-means is the most widely used algorithm in this field with high accuracy. The using algorithms to detect variability of the countries COVID 19 related factors are K-Means clustering (Abdullah et al., 2022; Aydin and Yurdakul, 2020; Carrillo-Larco and Castillo-Cara, 2020; Gohari et al., 2022; Imtyaz et al., 2020; Rizvi et al., 2021; Siddiqui et al., 2020), hierarchical clustering (Aydin and Yurdakul, 2020; Sadeghi et al., 2021; Zarikas et al., 2020), fuzzy clustering (Mahmoudi et al., 2020), Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm (Shuai et al., 2020), Kohonen-SOM clustering (Boluwade, 2020).

Clustering analysis is one of the data mining methods in Big Data methodologies that use data

* Corresponding Author.
E-mail: pyigit@medipol.edu.tr

Recieved : 23 Sep 2023
Revision : 19 Dec 2023
Accepted : 5 Apr 2024

segmentation. In COVID-19 literature, traditional clustering methods (K-Medoids and Hierarchical clustering) are commonly used (Hussein and Abdulazeez, 2021) . In addition, Arunachalam and Kumar (Arunachalam ve Kumar, 2018) find that ANN-based SOM clustering finds hidden structures in the data set better than hierarchical, K-Medoids, and fuzzy clustering techniques. Bloom (2004) also found that SOM is better than the hierarchical clustering method, overcoming its limitations and better dealing with missing data (Brida vd., 2012). ANN makes predictions by mathematically modeling the way the human brain thinks. They do not require any statistical assumptions. Additionally, they succeed at identifying nonlinear models, so they are highly recommended for their flexibility, robustness, and higher prediction accuracy abilities to solve real-life problems. The SOM clustering method also applies a two-stage clustering method using both ANN and traditional clustering sequentially, which is more robust than other methods. Countries' COVID-19 data do not show normal distribution, and the relationships between variables are nonlinear. Therefore, SOM clustering analysis is effective in examining the differences and similarities between countries' COVID-19 and related variables.

Therefore, the present study aims to cluster OECD countries using COVID-19 and related socioeconomic indicators using SOM clustering method. It uses a two-level approach based on using a SOM in sequence, followed by hierarchical clustering analysis.

The rest of the paper is organized as follows: Section 2 presents material and methods and SOM analysis, Section 3 introduces findings, and Section 4 presents the conclusion.

2. Material and Methods

2.1. Dataset

In this paper, the analysis are performed on a data set of 38 OECD countries and 16 features. The data are obtained from different sources, OECD stat, World Bank, and our World in Data. The variables and their sources are presented in Table 1. The variables selected for the analysis are set according to related literature on COVID-19, mentioned in the introduction. They are examined under three different headings: COVID-19 variables (cum confirmed deaths, cum confirmed cases, cum vaccinations, cum tests), socioeconomic variables (life expectancy, elderly population, share of GDP, current PPPs, out-of-pocket health expenditures, smoking pr, alcohol consumption), environmental factors (EPI, HLT), diseases indicators (deaths from chronic respiratory diseases, deaths from cardiovascular diseases, diabetes pr). The newest available data is used for all countries.

The Environmental Performance Index (EPI) is calculated by researchers from Yale and Colombia Universities (Wolf et al., 2022). It was calculated from 40 indicators with 11 categories and three headings: environmental health (air quality, waste management, water, and sanitation, heavy metals), climate (climate change mitigation), ecosystem validity (biodiversity and habitat, ecosystem services, fisheries, agriculture, acid rain, and water resources). It uses both EPI and environmental health (HLT) indicators.

2.2. Kohonen SOM Analysis

SOM, also called Kohonen SOM, is an unsupervised ANN algorithm and introduced by Kohonen (Kohonen, 1982). The background of SOM comes from functions of neurons like other ANN methods. SOM can learn from multi-dimensional data and transform them into low-dimensional (mainly two-dimensional) topological order, preserving the original relations. The topological ordering map easily visualizes the similarities between the units according to their distance.

Table 1. Study Indicators

Variable name	Description	Data Source
Cum confirmed deaths	Total confirmed cases due to COVID-19 per million people	COVID-19 Indicator Our World in Data(Our World in Data, 2023)
Cum confirmed cases	Total confirmed deaths due to COVID-19 per million people	
Cum vaccinations	Total vaccinations per hundred	
Cum tests	Total COVID-19 test thousand people	
Life expectancy	The average measure of how long a born baby lives	Socioeconomic Indicators OECD(OECD, 2023)
Elderly population	percentage of aged 65 and over in the population	
Share of gross domestic product	The ratio of total health expenditures in gross domestic product (GDP)	
Current PPPs	Current health expenditure per capita	
Out-of-pocket health expenditures	Share of out of out-of-pocket health expenditure per capita	
Smoking pr	Daily smokers, % of population aged 15+	
Alcohol consumption	yearly sales of alcohol in liters per person aged 15+	
HLT	Measure of Environmental Health (HLT)	Environmental Performance Indicators(Wolf et al., 2022)
EPI	Measure of Environmental Performance Index (EPI)	
Deaths from Chronic respiratory diseases	Chronic respiratory diseases death rates (Sex: Both - Age:	Disease Mortality Our World In Data(Our World in Data, 2023)

	Age-standardized-2019)	
Deaths from Cardiovascular diseases	Cardiovascular disease death rates (Sex: Both - Age: Age-standardized-2019)	
Diabetes pr	Diabetes prevalence (% of population ages 20 to 79)	Disease Prevalence World Bank(The World Bank, 2023)

Like other ANN methods, it has input layer neurons (input data) and output layer neurons (topological order: hexagonal or rectangular lattice). The output layer neurons are connected to every neuron in the input nodes with weight vectors. The SOM algorithm is summarized into five stages (Haykin, 2008):

1. **Initialization:** Set the starting weight vectors wj(0) to random values.
2. **Sampling:** Draw a sample x with a certain probability from the input space. The activation pattern applied to the lattice is represented by the vector x. The dimension of the vector x is m.
3. **Similarity matching:** Utilizing the minimum-distance criterion, determine the best-matching (winning) neuron i(x) at time-step n:

$$i(x) = \arg\min_{j} \left\| x(n) - w_j \right\|, \quad j = 1,2,\ldots,l \qquad (1)$$

4. **Updating:** Using the update formula, modify the synaptic-weight vectors of all stimulated neurons:

$$w_j(n+1) = w_j(n) + \eta(n) h_{j,i(x)}(n) \big(x(n) - w_j(n)\big) \qquad (2)$$

5. **Continuation:** Use step 2 until there are no changes in the feature map.

2.3. *Application of SOM*

The Kohonen package (Wehrens, 2018) in R is used to perform SOM clustering. Firstly, the data is normalized and transformed into matrix form. In normalization, z values were used. The shape of the topological order should be chosen: hexagonal or circular. According to Kohonen's suggestion, the hexagonal topological order was selected (Kohonen, 2013). The number of nodes is decided as 5 $\sqrt{n}$ rule (Bruwer et al., 2018; Huiyan et al., 2008). The iteration demonstrates the iterative process and how distances arrive at their smallest value. After several experiments, a 5x5 SOM grid (25 neurons) with 1000-time iterations is created with hexagonal topologies. The learning rate was between 0.05 and 0.01.

A two-level approach based on using in sequence a SOM is used, followed by hierarchical clustering analysis, proposed by Vesanto and Alhoniemi (Vesanto and Alhoniemi, 2000). In this approach, initially, the SOM method applies and has SOM codes, and then SOM codes are clustered by hierarchical or partitive clustering methods. It helps to obtain more robust clusters. In this study, 25 neuron SOM codes clustered by Euclidean distance and Ward's agglomerative linkage method. Silhouette Index (Rousseeuw, 1987) and the Davies–Bouldin Index (Davies, D and Bouldin, D, 1979) are used to determine cluster size. Optimal clusters are found as three (Figure 1).

Spearman Correlation analysis is also conducted to measure the relationship between COVID-19 indicators and socioeconomic, environmental, and disease factors.

3. Results

The correlation matrix is given in Table 2. There are high positive correlations between COVID-19 cases and smoking pr (0.46) and alcohol consumption (0.52). Moderate correlation is found between COVID-19 cases and EPI (0.348). Moderate and negative correlation is also found between COVID-19 cases and life expectancy (-0.34). It is found high positive correlations between COVID-19 deaths and cardiovascular mortality rates (0.58) and smoking pr (0.52). In addition, high and negative correlation exists between COVID-19 deaths and life expectancy (-0.71) and HLT (-0.63). There is moderate correlation between COVID-19 deaths and alcohol consumption (0.39). High positive correlations are found between COVID-19 vaccinations and life expectancy (0.533) and HLT (0.51). There are moderate correlations between COVID-19 vaccinations and share of GDP (0.41), per capita current prices (0.41), and a negative, moderate correlation with cardiovascular mortality (-0.492).

Table 2. Spearman correlations for study Indicators

	Cum confirmed deaths	Cum confirmed cases	Cum Vaccinations	Cum tests
Life expectancy	-.712**	-.337*	.533**	-0.008
Share of GDP	-0.169	0.021	.412*	0.141
Current PPPs	-.394*	0.12	.409*	0.259
Out of Pocket Health Exp.	0.23	-0.179	-0.112	-0.098
HLT	-.631**	-0.113	.511**	0.248
EPI	-0.09	.348*	0.222	.464**
Elderly Population	0.196	0.159	0.085	0.199
Smoking pr	.519**	.461**	-0.113	.366*
Alcohol Consumption	.388*	.518**	-0.107	.400*
Deaths from Chronic respiratory diseases	-0.056	-0.102	0.141	-0.19
Deaths from Cardiovascular diseases	.579**	0.279	-.492**	0.068
Diabetes pr	0.13	-0.314	-0.247	-.369*

*p<0.05; **p<0.01

SOM quality is visually measured with node counts, node quality (distance), and SOM neighbor distance plots (Arunachalam and Kumar, 2018). They can be seen in Figure 2. The counts plot visualizes the number of countries in each node. There were 1-4 countries in each node. Grey nodes show empty nodes. The quality plot displays the average distance between countries. SOM neighbour distance plot (U-matrix) represents the distance between each node and its neighbors. It gives the idea of determining cluster numbers. The red color means closer neurons with similar characteristics, and the straw yellow indicates neighboring neurons with different characteristics.

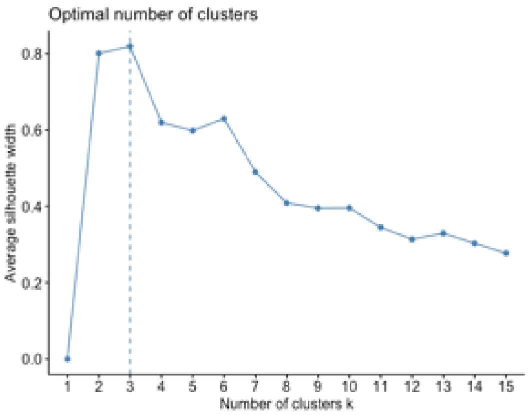

Figure 1. Silhouette Index

The membership of the 38 countries in the three clusters is provided in Table 3 and Figure 2.

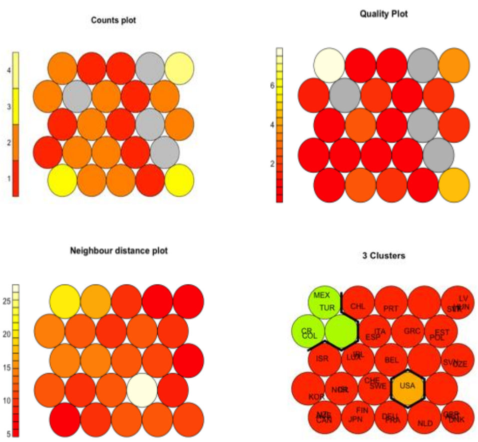

Figure 2. Counts Plot, Quality Plot, Neighbour distance plot and Cluster Plot

According to SOM visualization, it can be seen that The USA was the most different country across OECD countries. Secondly, Mexico and Turkey differ from other countries and are in the same nodes, meaning they have quite different characteristics from other countries, but highly similar each other.

The means of the study variables in the clusters are given in Table 4. Cluster 1 comprises 31 developed and two developing countries (Chile and Poland) (WEO Groups and Aggregates Information, 2023). Cluster 1 has the highest mean of cum confirmed cases (119,596), cum vaccinations (172), cum tests (2,966), life expectancy (80.7), the elderly population (19.23), HLT (77.06), EPI (60.38) and the lowest cum confirmed deaths (1,577), deaths from chronic respiratory diseases (20.72) and diabetes pr (6.25).

Table 3. Countries in the Clusters

Cluster-1			Cluster-2	Cluster-3
Australia	Greece	Latvia	United States	Colombia
Austria	Hungary	Israel		Costa Rica
Belgium	Iceland	Italy		Mexico
Canada	Lithuania	Japan		Turkey
Chile	Luxembourg	Portugal		
Czech Republic	Netherlands	Slovak Republic		
Denmark	New Zealand	Slovenia		
Estonia	Norway	Spain		
Finland	Poland	Sweden		
France	Ireland	Switzerland		
Germany	Korea	United Kingdom		

The USA, one of the developed countries, constitutes Cluster 2. The cluster-2 has the highest value of cum confirmed deaths (2,421), the share of GDP (17.36), current PPPs (12,196), and deaths from chronic respiratory diseases (37.72), and the lowest life expectancy (76.4), out-of-pocket health expenditures (10.70).

The cluster 3 comprises four developing countries: Colombia, Costa Rica, Mexico, and Turkey. They have the highest mean of out-of-pocket health expenditure (23.00), diabet pr (12.13) and the lowest cum confirmed cases (87,886), cum vaccinations (136.3), elderly population (8.85), share of GDP, current PPPs (1,506), smoking pr (15.00), alcohol consumption (3.43), HLT (48.60), EPI (40.13).

Smoking pr and alcohol consumption values of Cluster 1 and Cluster 2 are found to be very close to each other (23.70; 23.00 - 9.21; 9.50, respectively). Likewise, deaths from cardiovascular diseases also had very similar values for the three clusters (158.97; 157.01; 158.68, respectively). In addition, developed countries have the highest HLT and EPI, while developing countries have the lowest.

Table 4. Cluster mean of variables

Study Variable	Cluster1	Cluster2	Cluster3
Cum confirmed deaths	1,578	2,420	1,817
Cum confirmed cases	119,596	158,249	87,886
Cum vaccinations	171.82	157.20	136.30
Cum tests	2,966	2,155	628
Life expectancy	80.73	76.40	77.90
Elderly population	19.23	16.83	8.85
Share of gross domestic product	9.83	17.36	6.81
Current PPPs	4,877	12,196	1,506
Out-of-pocket health expenditures	18.00	10.70	23.00
Smoking pr	23.70	23.00	15.00
Alcohol consumption	9.21	9.50	3.43
HLT	77.06	76.80	48.60
EPI	60.38	51.10	40.13
Deaths from chronic respiratory diseases	20.72	37.72	34.27
Deaths from cardiovascular diseases	158.97	157.01	158.68
Diabetes pr	6.25	10.70	12.13

4. Conclusion

COVID-19 has affected many people globally, making it one of the most significant challenges to humankind recently. Since the beginning of the pandemic, various studies have been conducted to help policymakers make better decisions for countries. The study investigates and presents the topic using a robust clustering technique and using various indicators related the pandemic. Therefore, the study proposes to cluster OECD countries using COVID-19, health, socioeconomic, and environmental indicators.

This study uses the Kohonen SOM clustering method to cluster 38 OECD countries based on COVID-19 confirmed cases, deaths, vaccinations, tests, and health, socioeconomic, and environmental variables. The data set used in the study consists of 16 variables. The study conducted a two-level approach of clustering SOM: SOM and hierarchical clustering. Silhouette and the Davies–Bouldin Index methods were used to decide the optimal number of clusters, and the number of optimal cluster is three.

Cluster 1 has 33 countries, with 31 developed and two developing countries (Chile and Poland) showing the lowest mean of confirmed COVID-19 deaths and the highest confirmed cases, vaccinations, and tests. Cluster 2 consists of only USA. It distinguishes itself from other countries by having the highest number of COVID-19 deaths. Cluster 3 contains four developing countries: Colombia, Costa Rica, Mexico, and Turkey. It shows the lowest number of COVID-19 confirmed cases, vaccinations, and tests.

It is found that deaths from cardiovascular diseases are not distinctive in separating clusters because it is the leading cause of death for all countries in the World (WHO, 2020). It has almost similar risks for all countries. On the other hand, chronic respiratory disease mortality is strongly associated with COVID-19 indicators and confirmed by several studies (Kumru et al., 2022; Rizvi et al., 2021). The developed countries have higher EPI and HLT, lower COVID-19 mortality, confirming by other studies (Coccia, 2021; Rizvi et al., 2021). However, The USA has the highest COVID-19 mortality, GDP, and health expenditures. Studies show that the USA did not prevent COVID-19 cases surveillance and provide equal health services during the pandemic because of it substantial regional differences (Bergquist et al., 2020; Bollyky et al., 2023). In addition, Turkey is in the same cluster with three other developing countries. Developing countries have lower GDP, and health expenditures so their health system is weak to deal with the pandemic (Coccia, 2021). The COVID-19 statistics data is not also well documented in developing countries because of their health system (Levin et al., 2022).

The study differs from other literature to cluster countries. (1) it uses COVID-19 cases, deaths, tests and vaccinations, and related socioeconomic and environmental indicators. (2) It uses an ANN-based SOM clustering technique. Other studies have used different clustering methods and variables when investigating the hidden structures of COVID-19-related factors across countries.

There were some limitations in this study. First, the study examined only 38 OECD countries. Hence, the results may not fully reflect global trends in COVID-19 and related factors. Second, our data set did not contain all the variables related to the pandemic. Despite these limitations, the study uses an ANN-based SOM two-phased clustering approach with various related COVID-19 features. It has focused on OECD countries to have more quality comparatives especially for Turkey's situation. The hope is that the study helps policymakers make regulations about emergencies in our country and other countries and to plan new studies.

References

Abdullah, D., Susilo, S, Ahmar A.S., et al., 2022. 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 and Quantity* 56(3). Springer Netherlands: 1283–1291. DOI: 10.1007/s11135-021-01176-w.

Arunachalam, D., Kumar, N., 2018. Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence of data-driven decision-making. *Expert Systems with Applications* 111. Elsevier Ltd: 11–34. DOI: 10.1016/j.eswa.2018.03.007.

Aydin N and Yurdakul G., 2020. Assessing countries' performances against COVID-19 via WSIDEA and machine learning algorithms. *Applied Soft Computing Journal* 97. Elsevier B.V.: 106792. DOI: 10.1016/j.asoc.2020.106792.

Bergquist, S., Otten, T., Sarich, N., 2020. COVID-19 pandemic in the United States. *Health Policy and Technology* 9(4). Elsevier Ltd: 623–638. DOI: 10.1016/j.hlpt.2020.08.007.

Bollyky, T.J., Castro, E., Aravkin, A.Y., et al., 2023. Assessing COVID-19 pandemic policies and behaviours and their economic and educational trade-offs across US states from Jan 1, 2020, to July 31, 2022: an observational analysis. *The Lancet* 401(10385). The Authors. Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license: 1341–1360. DOI: 10.1016/S0140-6736(23)00461-0.

Boluwade, A., 2020. Regionalizing Partitioning Africa's Coronavirus (COVID-19) Fatalities Using Environmental Factors and Underlying Health Conditions for Social-economic Impacts. *2nd Novel Intelligent and Leading Emerging Sciences Conference, NILES 2020*: 439–443. DOI: 10.1109/NILES50944.2020.9257875.

Bruwer, J., Prayag, G., Disegna, M., 2018. Why wine tourists visit cellar doors: Segmenting motivation and destination image. *International Journal of Tourism Research* 20(3): 355–366. DOI: 10.1002/jtr.2187.

Calderón-Larrañaga, A., Dekhtyar, S., Vetrano, D.L., et al., 2020. COVID-19: risk accumulation among biologically and socially vulnerable older populations. *Ageing Research Reviews* 63(May). DOI: 10.1016/j.arr.2020.101149.

Cardoso, E.H.S., Silva, M.S., Da, Júnior, FEDAF, et al., 2020. Characterizing the Impact of Social Inequality on COVID-19 Propagation in Developing Countries. *IEEE Access* 8: 172563–172580. DOI: 10.1109/ACCESS.2020.3024910.

Carrillo-Larco, R.M., Castillo-Cara, M., 2020. Using country-level variables to classify countries according to the number of confirmed COVID-19 cases: An unsupervised machine learning approach. *Wellcome Open Research* 5: 56. DOI: 10.12688/wellcomeopenres.15819.1.

Coccia, M., 2021. High health expenditures and low exposure of population to air pollution as critical factors that can reduce fatality rate in COVID-19 pandemic crisis: a global analysis. *Environmental Research* 199(January). Elsevier Inc.: 111339. DOI: 10.1016/j.envres.2021.111339.

Davies, D.L., Bouldin, D.W., 1979. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2): 224–227.

Gebhard, C., Regitz-Zagrosek, V., Neuhauser, H.K., et al., 2020. Impact of sex and gender on COVID-19 outcomes in Europe. *Biology of Sex Differences* 11(1). Biology of Sex Differences: 1–13. DOI: 10.1186/s13293-020-00304-9.

Gohari, K., Kazemnejad, A., Sheidaei, A., et al., 2022. Clustering of countries according to the COVID-19 incidence and mortality rates. *BMC Public Health* 22(1). BioMed Central: 1–12. DOI: 10.1186/s12889-022-13086-z.

Haykin, S., 2008. *Neural Networks and Learning Machines*. DOI: 978-0131471399.

Huiyan, S.B., Gelfand, A.E., Chris, L., et al., 2008. Interpreting self-organizing maps through space–time data models. *The Annals of Applied Statistics* 2(4): 1194–1216. DOI: 10.1214/08-AOAS174.

Hussein, H.A., Abdulazeez, A.M., 2021. Covid-19 Pandemic Datasets Based on Machine Learning Clustering Algorithms: A Review. *Journal Of Archaeology Of Egypt/Egyptology* 18(4): 2672–2700. Available at: https://archives.palarch.nl/index.php/jae/article/download/6703/6488.

Imtyaz, A., Abid Haleem, Javaid, M., 2020. Analysing governmental response to the COVID-19 pandemic. *Journal of Oral Biology and Craniofacial Research* 10(4). Elsevier: 504–513. DOI: 10.1016/j.jobcr.2020.08.005.

Islam, N., Lacey, B., Shabnam, S., et al., 2021. Social inequality and the syndemic of chronic disease and COVID-19: County-level analysis in the USA. *Journal of Epidemiology and Community Health* 75(6): 496–500. DOI: 10.1136/jech-2020-215626.

Khan, J.R., Awan, N., Islam, M.M., et al., 2020. Healthcare Capacity, Health Expenditure, and Civil Society as Predictors of COVID-19 Case Fatalities: A Global Analysis. *Frontiers in Public Health* 8(July): 1–10. DOI: 10.3389/fpubh.2020.00347.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43(1): 59–69. DOI: 10.1007/BF00337288.

Kohonen, T., 2013. Essentials of the self-organizing map. *Neural Networks* 37. Elsevier Ltd: 52–65. DOI: 10.1016/j.neunet.2012.09.018.

Kumru, S., Yiğit, P., Hayran, O., 2022. Demography, inequalities and Global Health Security Index as correlates of COVID-19 morbidity and mortality. *International Journal of Health Planning and Management* 37(2): 944–962. DOI: 10.1002/hpm.3384.

Levin, A.T., Owusu-Boaitey, N., Pugh, S., et al., 2022. Assessing the burden of COVID-19 in developing countries: Systematic review, meta-Analysis and public policy implications. *BMJ Global Health* 7(5): 1–17. DOI: 10.1136/bmjgh-2022-008477.

Mahmoudi, M.R., Baleanu, D., Mansor, Z., et al., 2020. Fuzzy clustering method to compare the spread rate of Covid-19 in the high risks countries. *Chaos, Solitons and Fractals* 140. Elsevier Ltd: 1–9. DOI: 10.1016/j.chaos.2020.110230.

Micah, A.E., Cogswell, I.E., Cunningham, B., et al., 2021. Tracking development assistance for health and for COVID-19: a review of development assistance, government, out-of-pocket, and other private spending on health for 204 countries and territories, 1990–2050. *The Lancet* 398(10308): 1317–1343. DOI: 10.1016/S0140-6736(21)01258-7.

OECD, 2023. OECD Statistics. Available at: https://stats.oecd.org/ (accessed 11 July 2023).

Our World in Data (2023) Data, Coronavirus Pandemic (COVID-19) - Statistics and Research - Our World in. Available at: https://ourworldindata.org/explorers/coronavirus (accessed 8 July 2023).

Rizvi, S.A., Umair, M., Cheema, M.A., 2021. Clustering of countries for COVID-19 cases based on disease prevalence, health systems and environmental indicators. *Chaos, Solitons and Fractals* 151. Elsevier Ltd: 111240. DOI: 10.1016/j.chaos.2021.111240.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20(C): 53–65. DOI: 10.1016/0377-0427(87)90125-7.

Sadeghi, B., Cheung, R.C.Y., Hanbury, M., 2021. Using hierarchical clustering analysis to evaluate COVID-19 pandemic preparedness and performance in 180 countries in 2020. *BMJ Open* 11(11): 1–11. DOI: 10.1136/bmjopen-2021-049844.

Shuai, Y., Jiang, C., Su, X., et al., 2020. A Hybrid Clustering Model for Analyzing COVID-19 National Prevention and Control Strategy. *2020 IEEE 6th International Conference on Control Science and Systems*

Engineering, ICCSSE 2020: 68–71. DOI: 10.1109/ICCSSE50399.2020.9171941.

Siddiqui, M.K., Morales-Menendez, R., Gupta, P.K., et al., 2020. Correlation between temperature and COVID-19 (suspected, confirmed and death) cases based on machine learning analysis. *Journal of Pure and Applied Microbiology* 14(May): 1017–1024. DOI: 10.22207/JPAM.14.SPL1.40.

The World Bank, 2023. The World Bank Data.

Vesanto, J., Alhoniemi, E., 2000. Clustering of self-organizing map. *IEEE TRANSACTIONS ON NEURAL NETWORKS* 11(3): 586–600.

Wehrens, M.R., 2018. Package ‘ kohonen .’

WEO Groups and Aggregates Information, 2023. World Economic Outlook Database - Groups and Aggregates. Available at: https://www.imf.org/en/Publications/WEO/weo-database/2023/April/groups-and-aggregates (accessed 12 August 2023).

WHO, 2020. The top 10 causes of death. Available at: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (accessed 28 December 2023).

WHO, 2023a. Coronavirus disease (COVID-19) pandemic. Available at: https://www.who.int/europe/emergencies/situations/covid-19 (accessed 10 August 2023).

WHO, 2023b. WHO Coronavirus (COVID-19) Dashboard.

Wolf, M.J., Emerson, J.W., Esty, D.C., et al., 2022. Environmental Performance Index. DOI: 10.1002/9781118445112.stat03789.

Worldometer, 2023. COVID - Coronavirus Statistics - Worldometer. Available at: https://www.worldometers.info/coronavirus/ (accessed 10 August 2023).

Zarikas, V., Poulopoulos, S.G., Gareiou, Z., et al., 2020. Clustering analysis of countries using the COVID-19 cases dataset. *Data in Brief* 31. Elsevier Inc.: 105787. DOI: 10.1016/j.dib.2020.105787.