



Big data analytics and COVID-19: investigating the relationship between government policies and cases in Poland, Turkey and South Korea

Mert Erkan Sözen¹, Görkem Sanyer^{2,*} and Mustafa Gökalp Ataman³

¹Business Development & Budget Planning Chef, İzmir Metro Company, 2844 Street, No. 5, 35110, İzmir, Turkey

²Selçuk Yaşar Campus, Üniversite Street, Ağaçlı Yol No: 37-39, 35100 İzmir, Turkey

³Department of Emergency Medicine, İzmir Bakırçay University Çiğli Training and Research Hospital, Gazi Mustafa Kemal, Kaynaklar Street, 35665 İzmir, Turkey

*Corresponding author. Department of Business, Yasar University, Selcuk Yasar Campus, Universite Caddesi, Agacliyol, 35100 Izmir, Turkey.

E-mail: gorkem.ataman@yasar.edu.tr

Accepted on 26 July 2021

Abstract

We used big data analytics for exploring the relationship between government response policies, human mobility trends and numbers of coronavirus disease 2019 (COVID-19) cases comparatively in Poland, Turkey and South Korea. We collected daily mobility data of retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential areas. For quantifying the actions taken by governments and making a fairness comparison between these countries, we used stringency index values measured with the ‘Oxford COVID-19 government response tracker’. For the Turkey case, we also developed a model by implementing the multilayer perceptron algorithm for predicting numbers of cases based on the mobility data. We finally created scenarios based on the descriptive statistics of the mobility data of these countries and generated predictions on the numbers of cases by using the developed model. Based on the descriptive analysis, we pointed out that while Poland and Turkey had relatively closer values and distributions on the study variables, South Korea had more stable data compared to Poland and Turkey. We mainly showed that while the stringency index of the current day was associated with mobility data of the same day, the current day’s mobility was associated with the numbers of cases 1 month later. By obtaining 89.3% prediction accuracy, we also concluded that the use of mobility data and implementation of big data analytics technique may enable decision-making in managing uncertain environments created by outbreak situations. We finally proposed implications for policymakers for deciding on the targeted levels of mobility to maintain numbers of cases in a manageable range based on the results of created scenarios.

Keywords: Stringency index, COVID-19, big data analytics, mobility, number of cases

Introduction

The novel coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreak and consequent coronavirus disease 2019 (COVID-19) pandemic have shocked the world (Kaplan, 2020) and has been named as a black swan event causing unprecedented impacts on health systems, public health, societies, and individuals globally (The Lancet Public Health, 2020). This outbreak looks set to be the worst infectious disease in terms of infection and mortality rates. All over the world, people, patients, health service providers and administrators, and policymakers are dealing with high levels of uncertainties and various challenges to limit the spread of this virus and to protect the functioning of health systems (Gilson *et al.*, 2020). Given the high transmissibility and limited epidemiological understanding of this outbreak, governments have espoused various social distancing measures to restrict human mobility and mass gatherings. These measures include, but are not limited to, stay at home campaigns, closure of schools, works and entertainment places, cancellation and postponement of events, travel and public transport bans, etc.

Various studies have been published in this context to report, analyse and interpret different policy measures taken by governments to fight COVID-19 (i.e. Cheng *et al.*, 2020; Gilson *et al.*, 2020; Haldane and Morgan, 2021; Hadjidemetriou *et al.*, 2020; Hensen *et al.*, 2021). Additionally, an Oxford research team developed a scale to measure how strict governments are in fighting this outbreak quantitatively (Hale *et al.*, 2020). This scale is labelled as ‘Oxford COVID-19 government response tracker’ and it evaluates nine different policies as school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements and international travel controls in quantifying the levels of action taken by governments.

These actions inevitably resulted in behavioural changes and this highlighted measuring and using human mobility data. To measure human mobility, Google recently released a time-limited sharing of mobility data (Google, 2020), from all around the world in terms of summary statistics. The data are calculated for six areas as retail and recreation, grocery

Key messages

- Outbreak situations create highly uncertain environments that are challenging for decision-makers to deal with.
- With the advances in technologies, big data analytics techniques receive increasing attention of researchers and decision-makers for analysing high volumes and varieties of data.
- In fighting with COVID-19 outbreak situations, different policies and measures are taken by governments that affect the human behaviour and thus mobility trends.
- In order to measure the government response policies quantitatively, 'Oxford COVID-19 government response tracker' was used.
- We used big data analytics to investigate the relations between government response policies, human mobility and number of cases in selected three countries comparatively.
- We also developed a model for predicting number of cases based on mobility data for Turkey.
- We mainly showed that while the stringency index of day t is significantly related to mobility data of the same day, mobility data of day t is associated with number of cases of day $t + 30$.
- By creating different scenarios based on the descriptive statistics of the selected countries, we also proposed implications for policy makers on deciding the targeted levels of mobility to maintain numbers of cases in a manageable range.

and pharmacy, parks, transit stations, workplaces, and residential. The effects of government response measures can be observed through the changes in mobility patterns. Therefore, there exists a growing literature analysing how different measures adopted by governments affect mobility trends (Engle *et al.*, 2020; Brzezinski *et al.*, 2020; Frey *et al.*, 2020; Maloney and Taskin, 2020; Mendolia *et al.*, 2021). Existing studies have also shown a clear effect of the decrease in these mobility data on the spread rates of the pandemic (Fang *et al.*, 2020; Ferguson *et al.*, 2020; Prem *et al.*, 2020; Soucy Jean-Paul *et al.*, 2020; Yilmazkuday, 2020; Zhao *et al.*, 2020). Additionally, some studies have shown that today's mobility data affects the number of COVID-19 cases at least a month later (i.e. Bryant and Elofsson, 2020).

Although various studies have been published in this context, a gap is identified in analysing and quantifying the relations between government response measures and human mobility, respectively calculated based on the Oxford COVID-19 government response tracker and the Google mobility data in this study. On the other hand, while existing literature supports the reduction effect of mobility on the spread of COVID-19, only a few of them aimed to model this effect and use this model for predicting COVID-19 spread in terms of numbers of cases (infected people) or deaths (Bryant and Elofsson, 2020; Davies *et al.*, 2020; Flaxman *et al.*, 2020; Mandal *et al.*, 2020).

The recent advances in technologies have received the significant attention of researchers aiming to model uncertain environments. These emerging technologies additionally provide great opportunities to collect and store big volumes of

data. Since outbreak situations create highly uncertain environments and big data sets by their nature, during COVID-19, as one of the most popular emerging technologies, big data analytics has been used to detect surface indicators related to this pandemic (Guo *et al.*, 2020). These techniques have also been well implemented for modelling the spread of COVID-19 (Mohamadou *et al.*, 2020; Wang *et al.*, 2020; Yang *et al.*, 2020).

In the light of these, we mainly aimed to analyse the impact of government response policies on the human mobility trends and to model the spread of COVID-19, in terms of numbers of cases, based on mobility data by the implementation of big data analytics. To enrich the analysis and to demonstrate the applicability of the model in different data sets, we analysed three countries, Poland, South Korea and Turkey, comparatively. We discussed the reason for sampling these countries in the methods section. To emphasize the significance of government actions in fighting with COVID-19, we additionally presented some created scenarios based on the descriptive statistics of mobility data on these countries and generated predictions on numbers of COVID-19 cases. We then compared the predicted values and proposed implications for policymakers for deciding on the targeted levels of mobility to maintain numbers of cases in a manageable range. As a gap has been identified to model the effect of human mobility on COVID-19 spread indicators and big data analytics technologies provide opportunities for efficient modelling, this research contributes to the literature by implementing big data analytics to analyse and model COVID-19 spread indicators based on human mobility data. Thus, we aimed to answer the following research questions in this study.

RQ1. How governmental response policies, human mobility and numbers of COVID-19 cases were related?

RQ2. How COVID-19 case numbers can be modelled by defining human mobility data as input variables and implementing big data analytics?

With the presented analysis and proposed big data analytics integrated model, we addressed these research questions in the remaining parts of this study.

Materials and methods

Data sources

We collected data on selected governments' measures in response to COVID-19, by using the 'Oxford COVID-19 government response tracker' (Hale *et al.*, 2020). For each day, this tracker considers the availability of nine different policies in terms of school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements and international travel controls in countries and assigns a number between 0 and 100 to quantify the level of the measures taken by the government. The quantified value of the stringency index approaching 100 shows that the government behaves so strict and takes many measures in response to COVID-19 on this day. For collecting data on human mobility, we used Google Mobility data and measured the mobility in six areas: retail and recreation, grocery and pharmacy, parks, transit stations, workplaces and residential (Google, 2020). For the COVID-19 spread index, we used daily numbers of new cases (per million) and collected data on study countries

from the Our World in Data database (OWID, 2020). Thus, our data set included daily data on eight pre-defined variables, six mobility data, stringency index and numbers of cases (per million) for three countries.

The dates on which the first COVID-19 case was confirmed in study countries were, respectively, as 22nd of January, 4th of March, 11th of March 2020 for South Korea, Poland and Turkey. To make meaningful comparisons, we started our study period from the 11th of March, in which all of the study countries were exposed to COVID-19, until the end of the 25th of February 2021.

Study population

We used data sets of three countries as Poland, Turkey and South Korea. Poland and Turkey were selected since these two European countries have similar economic indicators, listed as emerging market economies based on IMF classification and taking relatively similar measures in response to COVID-19. In comparison, we decided to select South Korea as an Asian country, since it is defined as the leading country of emerging economies, which achieved high-income level, and has taken some additional measures in response to COVID-19. Compared to Poland and Turkey, South Korea adopted extensive testing policies and digital tracking and warning systems. Besides, Turkey and South Korea are listed as MIST (Mexico, Indonesia, South Korea and Turkey) countries as an acronym originally coined by Fidelity Investment in 2011. Also, according to COVID-19 spread indexes, while Poland and Turkey were relatively similar, South Korea had significantly lower values in terms of these indexes.

Analysis

Big data analytics is used in statistical analysis. We analysed the relation between the study variables with Pearson correlation analysis and showed them in a heat map form where different levels of correlation coefficient were shown with different colours. For modelling numbers of cases per million with the mobility data, we implemented one of the most popular big data analytics techniques, multilayer perceptron (MLP) neural networks. This technique is also classified as a machine learning technique that follows a supervised learning procedure relying on learning by observation. In the training data set, which is generally labelled as train data set, learning or training model is formed by using a backpropagation algorithm that is the most computationally straightforward algorithm for training MLP (Rumelhart *et al.*, 1986). MLP provides a multilayer-feed-forward neural network for learning and computation of connected weights of the network. Set of weights in the layers are iteratively trained by MLP until optimal scores of weights are found (Arora and Suman, 2012). Layers are key components of the MLP model. An input layer, hidden layer and output layers are the main types of layers. The input layer consists of the variables of the data set. For the given input variables $x_1, x_2, x_3, \dots, x_n$, and the target or output variable, y , MLP learns a non-linear function approximate for regression. The input layer does not play a computational role but serves to transmit the input vector to the network. More than one hidden layer may exist between input and output layers. MLP is described as being fully connected, with each node connected to every node in the next and previous layers. The analysis was performed in Python.

MLP techniques are superior in pattern classification and capture the complex behaviours of the input data set. Another important characteristic of MLP networks is their abilities of approximation of any linear or non-linear functions. A neural network architecture in a dense approach is critical for improving accuracy (specifically in the case when dealing with high numbers of input features) in comparison to other classification techniques in machine learning. As a disadvantage, hyperparameter tuning plays an important role to avoid overfitting/underfitting and exploding gradient problems; the requirement of time to train the data set is another disadvantage.

Results

Descriptive results

For each of the selected countries, we comparatively show the daily values of each of the study variables in Figure 1.

From the mobility data shown in Figure 1a through 1e, we mainly observed that Poland and Turkey had similar patterns. In both of these countries, mobility values in the first five areas (retail and recreation, grocery and pharmacy, parks, transit stations, and workplaces) showed the most decreasing trends and lower values in March through May and October through December. These periods include first and second peak periods of COVID-19 in these countries. The second peak periods when the disease spread index values were started to increase after having relatively lower values. Although we observed similar patterns in these graphs for the two countries, we also noticed that the decreasing trend was started earlier in Poland compared to Turkey particularly in the second peak periods. In each of these figures, we circled these periods with the same colour representing the country lines. In comparison, these two countries residential's mobility data had the opposite trends, wherein similar periods, mobility data of residential's had the most increasing trends with the higher values. On the other hand, mobility data of South Korea had a more stable trend when compared with Poland and Turkey. Especially, in Figure 1a, b and d, representing mobility data of retail and recreation, grocery and pharmacy, and transit stations, we observed relatively similar values in South Korea. For the others, parks, workplaces and residential, as in Figure 1c, e, and f, although we observed sharp decreases or increases in mobility data for some days, such trend was not continued for long periods as in other countries.

In addition, we observed a step-wise trend in stringency index values of the study countries in Figure 1g. In Turkey and Poland, we observed that the stringency index achieved and stayed in the highest values between March through May, and then, it decreased and stayed in lower values until the second peak was achieved. Thereafter, the stringency index was sharply increased in mid-September in Turkey, and at the end of October in Poland. Although we observed some decrease and increase steps in these countries' stringency index values after these periods, we mainly observed that they had relatively high values thereafter. In comparison, the stringency index increased and achieved the maximum level at the end of March; it then decreased and stayed in a range of similar values during the study period for South Korea.

For the study output variables, daily numbers of new cases (per million), we required to add a secondary vertical axis in

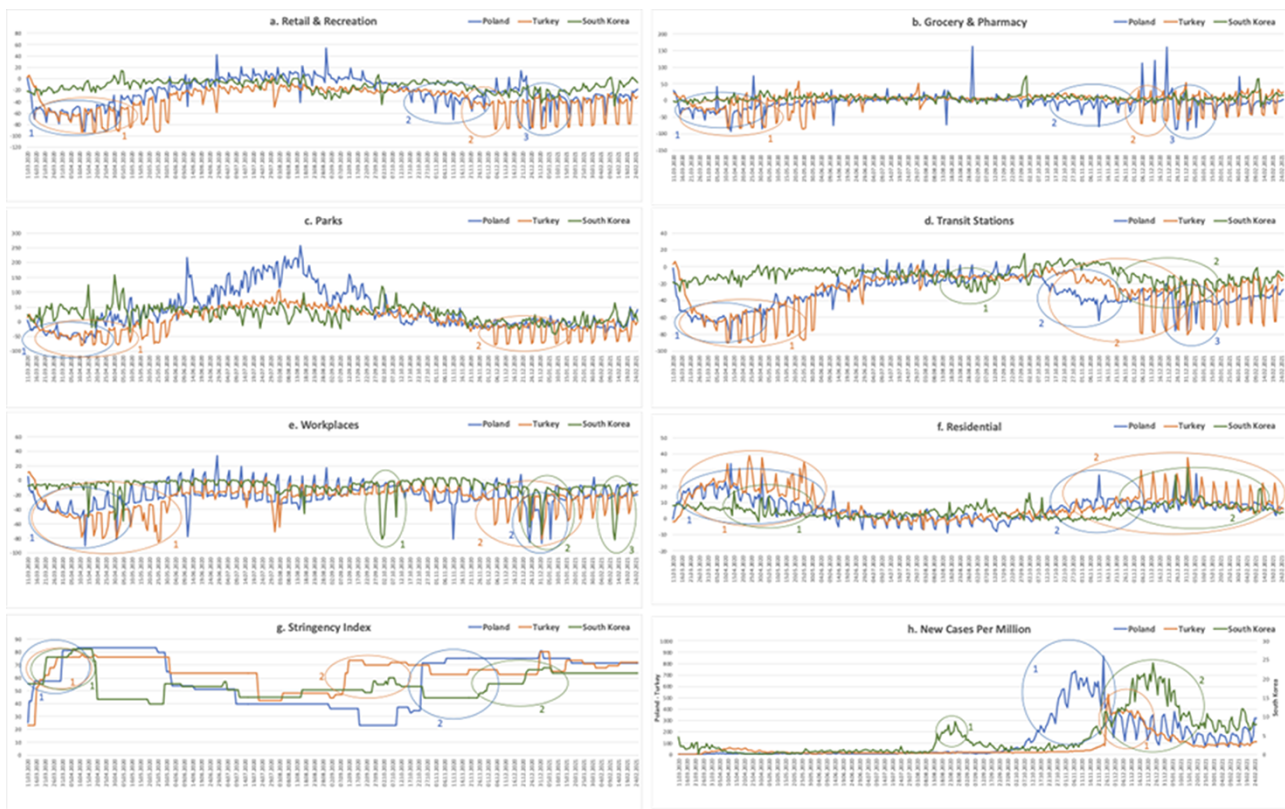


Figure 1. Daily values of the study variables for Poland, Turkey and China

Figure 1h, since the values of this variable were significantly different in South Korea compared to Poland and Turkey. While the values of new cases (per million) were changed in the range of 0 to more than 500 in Poland and Turkey, in South Korea, the values changed just in the range of 0 to 25. While, in the first peak periods, Poland and Turkey were successful in keeping the values of cases per million lower than 100, these countries were not able to keep this success in the second peak periods. While the numbers of cases (per million) started to be more than 100 in early October in Poland, these values were seen in late November in Turkey.

We also present descriptive statistics of the study variables during the study period in **Table 1**.

In **Table 1**, we observed that Poland and Turkey had similar levels in the minimum values of the study variables. However, since the maximum levels of the study variables were comparatively different in these countries, where maximum values were higher in Poland, the other descriptive statistics (25%, 50%, 75%, average) were different in these countries. We also observed that on average the mobility in the first five areas was higher in South Korea compared to other countries, except the parks' mobility, where Poland had a higher average statistics in parks' mobility compared to South Korea. We had the opposite finding on the residential's mobility data, lower statistics were seen in South Korea in this study variable compared to other countries. We observed that on average Turkey achieved the highest statistics on the stringency index, where it was followed by Poland and South Korea. While the maximum values of the stringency index were very close in these countries, the minimum stringency index level of South Korea was higher compared to other countries. We also depicted that

the standard deviation of the stringency index was the highest in Poland, where it was followed by Turkey and South Korea. For the output variable, new cases (per million) as was clearly seen in **Figure 1h**, significantly lower values were obtained in South Korea, and thus, descriptive statistics of this country were much lower compared to other countries. In terms of this output variable, Poland had the highest values according to average, maximum and deviation statistics.

Relation between stringency index and mobility data

We analysed how stringency index and mobility data of the same day were related to each other on the selected countries based on Pearson correlation analysis and represented the correlation coefficients in a heat map form as in **Figure 2**.

We mainly observed from **Figure 2** that most of the mobility data were related to each other in each of the countries. The mobility in retail and recreation of a day was positively related to grocery and pharmacy', parks', transit stations' and workplaces' mobility and was negatively related to residential' mobility of the same day. We had similar observation for parks' and transit stations' mobility data. In Poland and Turkey, a similar result was obtained for the mobility of grocery and pharmacy; however, in South Korea, this mobility was not related to the mobility of workplaces and residence. Similarly, while workplaces' mobility was related to all of the other types of mobility in Poland and Turkey, it was not highly related to grocery and pharmacy' and parks' mobility in South Korea. Finally, residential' mobility was negatively and highly related to all of the other types of mobility in Poland and

Table 1. Descriptive statistics of study variables

Variables	Country	Min	25%	50%	75%	Max	Mean	Std.
Retail and recreation' mobility	Poland	-92	-37	-22	-3	54	-22.99	24.17
	Turkey	-93	-49	-30	-20	6	-37.26	23.58
	South Korea	-47	-19	-11	-5.75	15	-12.85	9.91
Grocery and pharmacy' mobility	Poland	-94	-13.5	-4	3	162	-6.25	25.96
	Turkey	-86	-1	8	13	57	-1.29	27.50
	South Korea	-18	2	7	11	73	7.31	10.35
Parks' mobility	Poland	-71	-12.25	10.5	84.25	257	38.51	69.93
	Turkey	-81	-28	2.5	38	109	2	42.85
	South Korea	-42	4	27.5	44	158	25.28	28.64
Transit stations' mobility	Poland	-86	-42	-34	-19	9	-32.48	17.75
	Turkey	-90	-44	-21	-12	6	-30.08	24.62
	South Korea	-46	-16	-10	-4	16	-10.10	9.55
Workplaces' mobility	Poland	-90	-29	-24	-12	34	-22.30	16.78
	Turkey	-85	-37	-21	-16	12	-26.61	18.17
	South Korea	-82	-10	-6	-2	3	-8.74	12.78
Residential' mobility	Poland	-9	2	7	11	34	6.89	6.94
	Turkey	-7	2	6	14.25	39	9.14	9.18
	South Korea	-2	2	4	7	31	4.76	4.18
Stringency index	Poland	23.15	39.81	71.30	75	83.33	59.99	19.18
	Turkey	23.15	62.50	66.20	73.61	80.09	65.06	10.70
	South Korea	39.81	44.91	53.24	63.89	82.41	54.58	10.33
New cases (per million)	Poland	0.24	9.39	19.04	186.76	864.89	125.60	177.79
	Turkey	0	14.34	20.66	75.00	527.70	63.44	94.48
	South Korea	0	0.84	1.98	6.99	24.13	4.51	5.35



Figure 2. Correlation matrix between stringency index and mobility variables

Turkey, it was not highly related to grocery and pharmacy' mobility in South Korea.

On the other hand, when the relation between the stringency index of a day and mobility data of the same day was analysed, it was mainly observed that all types of mobility data were related to the stringency index in Poland and Turkey. In these countries, retail and recreation', parks', transit station' and residential' mobility had the highest correlation coefficients with the stringency index. However, in South Korea, we showed that while the stringency index was related to retail and recreation', transit stations' and residential' mobility, it was not related to grocery and pharmacy', parks' and workplaces' mobility. Besides, while the stringency index was positively related to residential' mobility, it was negatively related to all other types of mobility data.

Modelling results

In this section, we aimed to model COVID-19 spread based on the mobility data. By using Turkey as a template, we developed a prediction model for Turkey. Since the findings of the

previous section showed that the stringency index and mobility data were associated with each other, we decided to use only daily mobility data as model input variables where the target variable is the daily numbers of new cases (per million). On the other hand, since COVID-19 is a long-term continuing outbreak, which is affected by various other factors, such as culture, season, technology, etc., we decided to limit our study period to October, 28 to December 19 in this modelling part. Finally, as previous studies (i.e. Bryant and Elofsson, 2020) suggest that a decrease in mobility trend may cause a significant decrease in numbers of cases after some period passes, we used mobility data at time t , while the target variable was defined as numbers of cases (per million) at $t + 30$. This was also a good assumption for short-term planning whereby using current day's data, decision-maker may plan 1 month's after.

We showed the correlation matrix between the model's input and output variables in Figure 3.

In Figure 3, we mainly observed that all types of mobility data of the current day associate with the target variable, new cases (per million) in 1 month later. While mobility of the



Figure 3. Correlation matrix between mobility and new cases (per million) for Turkey case

first five areas was positively related, the last one residential’s mobility was negatively related to new cases (per million). We additionally observed that residential and recreation’, residential’ and transit station’ mobility has the highest level of relationship with the target variable.

Concluding that the mobility data were significant in modelling numbers of new cases (per million), we developed a prediction model by implementing MLP neural networks. We used ‘MLPRegressor’ in the neural network package of ‘sklearn’ module as a machine learning library for the model implementation. Activation and solver functions of the algorithm were set as ‘relu’ and ‘lbfgs’. We used a random train test split of 66% rule, thus while randomly selected 66% of the data set was used for training the algorithm, remaining was used to test the model performance. By changing the values of model parameters, learning rate, momentum, number of hidden layers, we repeated the experimentation until the best model performance was achieved. According to the trial and error technique, we identified the best model parameters as ‘momentum = 0.001, learning rate = 0.01 and $n_hiddenlayer = 5$ ’ having the prediction accuracy of 0.893 in the randomly split test data set. Thus, by the implementation of the MLP algorithm, 89.3% of the target variable, daily numbers of new cases (per million), was explained with the used mobility data. For the test data set, the distribution of

the predicted numbers of new cases (per million) around the obtained regression line is presented in Figure 4.

In Figure 4, we observed that the predicted values of the randomly selected 19 days (according to the 66% rule) of the study period identified in the modelling part were close to the actual numbers of new cases (per million). Once again, we concluded that the obtained model with the use of mobility data as model input variables, and the implementation of a superior big data analytics technique, performed well for COVID-19 spread modelling in terms of the number of cases (per million).

Before closing the analysis section, we aimed to create some scenarios by replacing the actual mobility data of the selected day in Turkey with the obtained descriptive statistics in Table 1 for obtaining predicted scores of numbers of cases (per million) and to make further comparative analysis. We believed that this analysis sheds light on investigating the effects of mobility data directly and the stringency index indirectly on COVID-19 spread.

For this final analysis, we selected 3 days having high, moderate and low numbers of new cases (per million) from Turkey’ test data set (as shown in Figure 4). Since we identified that retail and recreation’, transit stations’, parks’ and residential’ mobility were strongly related to the target variable, we decided to modify actual values of these mobility data and

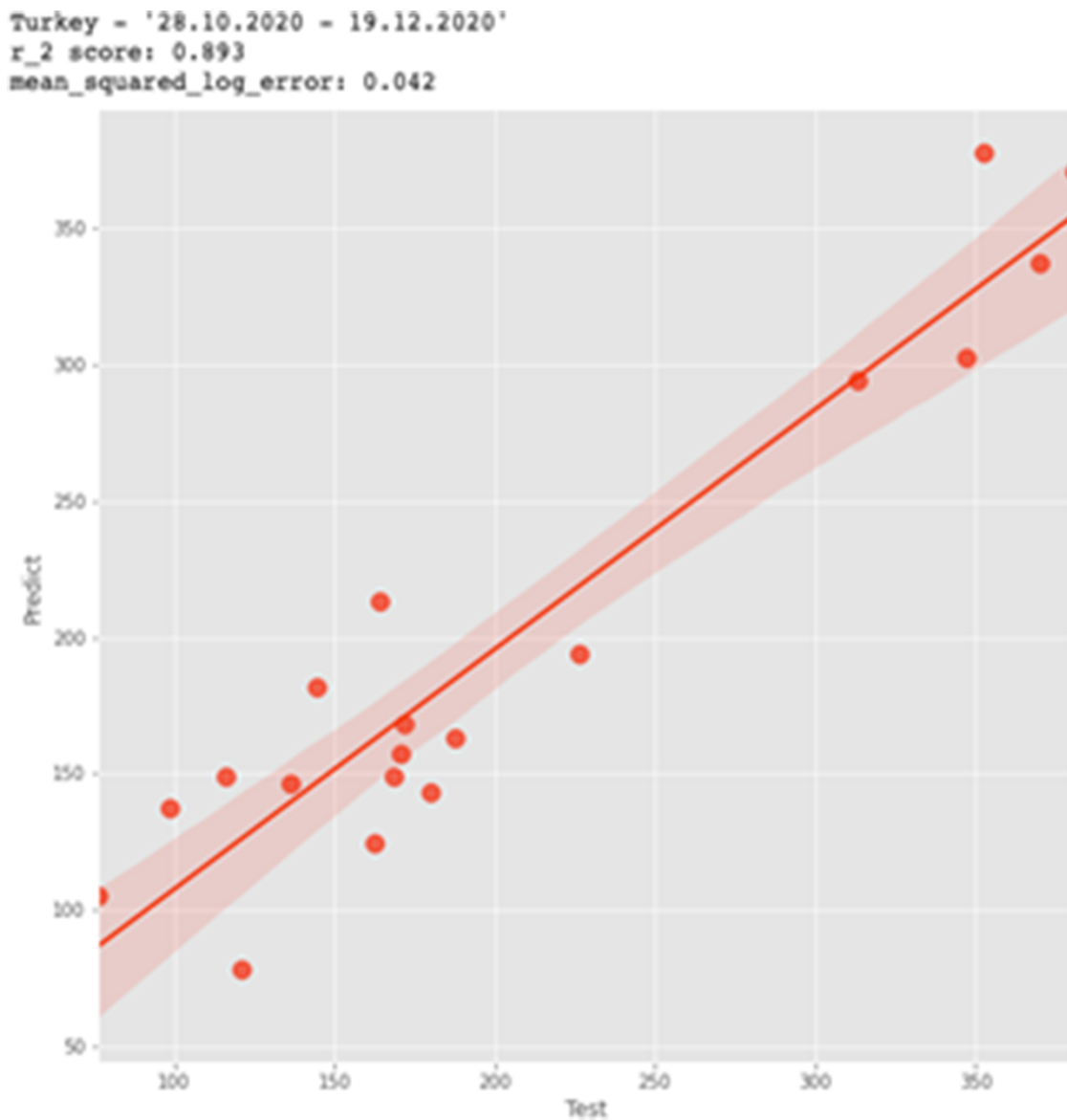


Figure 4. Distribution of numbers of new cases (per million) predictions in test data set for Turkey case

keep the others' values as they were actual. The values of the identified input variables were replaced with the minimum, average and maximum statistics of these variables of Turkey, Poland and South Korea and further comparative discussions were performed. By replacing these values, we created three scenarios: best case, avg. case and worst case. Best-case scenarios were obtained by replacing retail and recreation', transit stations' and park' mobility with minimum statistics and residential' mobility with maximum statistics of the countries. Since reducing non-residential' mobility as much as possible while increasing residential' mobility as much as possible is one of the most effective situations in the fight against COVID-19, we labelled this scenario as the best case one. Just the opposite replacement was done for creating a worst-case scenario, where retail and recreation', transit stations', and parks' mobility were replaced with maximum statistics and residential' mobility with minimum statistics of the countries. Similarly, we defined this scenario as the worst-case one, since such a situation makes things more challenging in the fight

against COVID-19. For creating avg. case, we replaced each of the three mobility data with countries' average statistics. We present the results of this analysis in [Table 2](#).

From the values in [Table 2](#), we observed that in worst-case scenarios, while retail and recreation', transit station' mobility values were much higher from the corresponding actual data, residential' mobility was lower for each of the selected days. Therefore, predicted values of the number of cases were much higher compared to actual ones. Besides, when predicted values of three countries were compared for the worst-case scenarios, we observed that the values were lowest with Turkey' statistics compared to Poland' and South Korea'. We, therefore, concluded that worst-case scenarios may create high numbers of cases, and it is highly crucial for policymakers to maintain values of retail and recreation' and transit station' mobility in the negative values and residential' mobility in higher values.

For the average-case scenario, based on Turkey and Poland, while replaced statistics of retail and recreation' and transit

Table 2. Predicted values of number of cases (per million) in Turkey with the created scenarios

Selected day number	Actual						Created scenarios						Predicted new cases (per million) with actual data	Predicted new cases (per million) with replaced data
	Retail and Recreation	Grocery and Pharmacy	Park	Transit stations	Workplaces	Residential	New cases (per million)	Based on	Retail & Recreation	Transit stations	Park	Residential		
DAY 1	-20	8	27	-19	-42	12	356.93	Turkey, best case Poland, best case South Korea, best case Turkey, avg. case Poland, avg. case South Korea, avg. case Turkey, worst case Poland, worst case South Korea, worst case Turkey, best case Poland, best case South Korea, best case Turkey, avg. case Poland, avg. case South Korea, avg. case Turkey, worst case Poland, worst case South Korea, worst case	-93	-90	-81	39	74.24	
									-92	-86	-71	34	81.60	
									-47	-46	-42	31	89.87	
									-37	-30	2	9	144.59	
									-23	-32	38,5	7	284.70	
									-13	-10	25,3	5	510.50	
									6	6	109	-7	816.91	
									54	9	257	-9	1318.99	
									15	16	158	-2	935.06	
DAY 2	-41	2	-15	-27	-26	10	214.63	Turkey, best case Poland, best case South Korea, best case Turkey, avg. case Poland, avg. case South Korea, avg. case Turkey, worst case Poland, worst case South Korea, worst case	-93	-90	-81	39	84.29	
									-92	-86	-71	34	81.64	
									-47	-46	-42	31	122.89	
									-37	-30	2	9	182.50	
									-23	-32	38,5	7	315.36	
									-13	-10	25,3	5	541.16	
									6	6	109	-7	847.57	
									54	9	257	-9	1349.64	
									15	16	158	-2	965.72	
DAY 3	-43	10	-20	-29	-24	12	98.58	Turkey, best case Poland, best case South Korea, best case Turkey, avg. case Poland, avg. case South Korea, avg. case Turkey, worst case Poland, worst case South Korea, worst case	-93	-90	-81	39	70.33	
									-92	-86	-71	34	75.69	
									-47	-46	-42	31	77.26	
									-37	-30	2	9	152.02	
									-23	-32	38,5	7	288.84	
									-13	-10	25,3	5	514.64	
									6	6	109	-7	821.05	
									54	9	257	-9	1323.13	
									15	16	158	-2	939.20	

station' mobility were better compared to actual values of the first selected day, replaced values of residential' mobility were worse. Although residential' mobility was replaced with worse values, since the others were replaced with better ones, predicted values of the number of cases significantly decreased compared to actual ones. This showed that, if for this selected day, the policymaker should keep mobility data in average-scenario values of especially Turkey, the number of cases could be decreased. However, in South Korea, since the average statistics of corresponding mobility data were worse compared to actual ones, predicted values were higher. We concluded that the government in Turkey should take measures to avoid keeping mobility data at these values regarding the average statistics of South Korea.

On the other hand, for the selected second and third days, the average statistics of all the study countries were worse compared to actual mobility data of these days. Thus, higher predictions were generated compared to the actual number of cases in those days. We, therefore, concluded that policymaker was required to take additional measures to have better average statistics on corresponding mobility data to decrease the number of cases (per million).

For the best-case scenarios, replaced statistics were all better compared to actual values of mobility data and lower values of cases were predicted compared to actual ones in all selected days. Thus, it should be provided for the policymaker if possible keeping the values of mobility data in such ranges could significantly decrease the number of cases (per million).

Finally, by comparing predicted values of all scenarios based on Turkey' and Poland's statistics, we observed that predicted values based on Turkey's statistics were lower. Since we showed that Turkey and Poland had relatively close values in numbers of cases in the previous analysis, we obtained an important finding from this comparison. If mobility data of Turkey were as in Poland, the numbers of cases would be much higher in Turkey.

Discussion

The main findings of descriptive analysis showed that while selected two European countries with emerging market economies had relatively similar values and characteristics on the study variables, South Korea had many differences. We found that, in Poland and Turkey, in general, while values of stringency index and residential mobility achieved the highest values in the first and second peak periods of COVID-19, the other five types of mobility data achieved the minimum levels. However, we emphasise that difference between these countries data was related to comparison between achieved values of the new cases (per million) in corresponding peak periods. While the first peak periods of the two countries were very similar, the second peak period was started comparatively earlier in Poland. Additionally, while the numbers of cases (per million) was relatively higher in Turkey in the first peak period, in the second peak period, the numbers were much higher in Poland. These findings may be reconciled with comparatively lower values of stringency index and higher values of mobility data, particularly retail and recreation', parks' and workplaces' mobility, in the time between two peak periods (May to October 2020). Based on these important findings, we highlighted the importance of keeping the measures taken

at a certain level even when the numbers of cases decreased in outbreak situations.

For the other study country, South Korea that is labelled as a leader of medium-income countries, the situation was highly different from Poland and Turkey. Although deviations exist in mobility and stringency index values, in general, we observed that these data were more stable when compared with Poland and Turkey. Additionally, although descriptive statistics of the stringency index of the three countries were close to each other, the minimum level of stringency index was comparatively higher in South Korea. Since South Korea had very lower values of cases (per million) compared to not only the study countries but also many of the others, we also showed the importance of being more stable in terms of mobility and being relatively strict even when minimal measures are taken.

Based on the descriptive analysis we also noted that while the values of the stringency index were not different between these countries, the selected spread index of COVID-19 was significantly different in South Korea compared to others. This should be due to additional measures adopted in South Korea that are not taken into consideration while calculating the value of the stringency index. We may list extensive testing policies, digital tracking and warning systems, comprehensive contact tracking, and ability to manage electronic health records as some of the additional measures adopted in South Korea to fight COVID-19. Besides, such a huge difference in the numbers of cases in South Korea may be associated with some other factors such as culture or health system. For example, in most of the Asian countries including South Korea, people used to masks even before the COVID-19 outbreak; thus, the government was not required to spend efforts for public campaigns on informing the people on the necessity of masks. Besides, in terms of main health system performance indicators, South Korea had better values compared to Turkey and Poland. For example, while the numbers of hospital beds available (per 1000 population) were 6.62 and 2.81 in Poland and Turkey respectively, this value was 12.27 in South Korea. Similarly, numbers of nursing personnel (per 10 000 population) were, respectively, as 68.93, 73 and 27.11 for Poland, South Korea and Turkey. South Korea additionally had significantly higher expenditures for its health system. While current health expenditure as a percent of gross domestic product was 6.5 and 4.2 for Poland and Turkey respectively, this value was 7.6 in South Korea. According to International Health Regulations, core capacity score was measured based on 13 core capacities as national legislation, policy and financing, and coordination; national focal point communications; surveillance; response; preparedness; risk communication; human resources; laboratory; points of entry; zoonotic events; food safety; chemical events; radionuclear emergencies and was rated in the range of 0 to 100, South Korea achieved the highest score. While it achieved the score of 97, scores of Poland and Turkey were, respectively, as 70 and 77. We collected health-system-based statistics from the World Health Organization official website (WHO, 2020).

To measure governments' response policies against COVID-19, we used the 'Oxford COVID-19 government response tracker' scale that calculates daily values of stringency indexes of all governments in terms of nine indicators described in the introduction section. When we investigated the relations between mobility data and stringency index, we

concluded that these data were associated with each other. This finding was in line with existing studies (Engle *et al.*, 2020; Brzezinski *et al.*, 2020; Frey *et al.*, 2020; Maloney and Taskin, 2020; Mendolia *et al.*, 2021) which explored how different policies adopted by governments affect the human mobility. To contribute to these existing studies, we used stringency index definition evaluating responses of governments from various perspectives and enables making fair comparisons between different countries. We presented the relation between mobility and stringency index data as a heat map form for the three study countries. From these figures, we also found out that although associations between stringency index and many types of mobility data exist, levels of relation changes between study countries. While similar levels of correlation exist between stringency index and mobility data in Poland and Turkey, the levels differ in South Korea. This finding may also be reconciled with the additional responses taken by the government of South Korea but which were not evaluated while calculating the stringency index.

From the analysis, since we observed that the numbers of cases in Turkey stay in the middle of Poland and Turkey, we took Turkey as a case country for modelling COVID-19 spread in terms of numbers of cases based on mobility data. We developed the prediction model by implementing an emerging technology of big data analytics. We used one of the most superior big data analytics technology that is also listed as a machine learning technique, MLP neural networks in modelling. Although outbreak situations create highly uncertain environments that are challenging for modelling and managing, we showed that the use of this technique performs well particularly in short-term modelling. This finding was also supported by existing literature (Ofli *et al.*, 2016; Akter and Wamba, 2019; Jia *et al.*, 2020; Song *et al.*, 2020). By using this technique particularly for COVID-19 spread modelling based on human mobility data that emphasizes data on mobility as surface indicators of outbreak situations, this study aimed to contribute to these existing studies. In this model, we also identified that mobility of current day' was a significant predictor for 1 month later' number of cases by which we supported the model proposed by Bryant and Eloffsson (2020).

Finally, by creating comparative scenarios on the actual descriptive statistics of the study countries and developing predictions on numbers of cases (per million) for these scenarios, we aimed to further investigate and clearly show the effect of mobility data on COVID-19 spread. We presented that while for the worst-case scenarios predicted numbers of cases were much higher compared to actual ones, they were much lower for the best-case scenarios. We generated the worst-case scenarios based on the maximum values of retail and recreation and transit stations and the minimum values of residential mobility in the selected countries during the study period. These three types of mobility were selected since they were identified as the ones having the highest relations with the numbers of cases. Best-case scenarios were generated just in the opposite way. Since we also showed the relation between stringency index and the mobility data, the results of these analyses shed light on the direct effect of mobility and the indirect effect of governments responses to fight with COVID-19. We, therefore, provided implications for policymakers for deciding on the targeted levels of mobility to maintain the numbers of cases in a manageable range.

In terms of the presented analysis, this study differs from existing studies by (1) not only describing how government response policies, human mobility and numbers of COVID-19 cases were related but also proposing efficient predictions for numbers of COVID-19 cases, (2) implementing big data analytics as one of the most popular emerging technologies for modelling, (3) creating scenarios based on study countries' actual data and exploring how these scenarios affect predicted values of COVID-19 cases.

Limitations

The main limitation of our study was defined as the numbers of selected countries. We studied only three countries data although, COVID-19 was a global outbreak situation. However, presenting such analysis for all countries should be complicated and we believed that using such models and analysis as a template, other countries' data should be efficiently analysed in future research studies. The other limitation was identified as developing a short-term model. Although we analysed data sets of the selected countries including two peaks and downs of this outbreak for around 1 year, in the modelling section, we limited the study period to around 2 months' data covering only the second peak period. The reason for making this limitation was explained as developing efficient short-term models for coping with outbreak situations.

Conclusion

In this study, by using big data analytics, we investigated the relations between government response policies, human mobility trend and COVID-19 spread in terms of numbers of cases in Poland, Turkey and South Korea. We selected Poland and Turkey as two European countries having emerging market economies and South Korea as a leader of middle-income countries that may even be listed in upper-income countries more recently. Besides the economical differences, we selected South Korea based on its superiority against fighting with COVID-19. We measured human mobility based on Google mobility data, and government response policies based on a scale labelled as 'Oxford COVID-19 government response tracker' which assigns a score between 0 and 100 in evaluating the stringency of the measures taken. We also developed a prediction model for daily numbers of COVID-19 cases by identifying mobility data as model inputs and implementing MLP neural network, as one of the most superior big data analytics techniques.

We mainly showed that daily values and distributions of study variables were closer in Poland and Turkey. On the other hand, in South Korea study variables were relatively more stable when compared with Poland and Turkey. We also showed that, while the current day's mobility data was associated with the current day' stringency index, it was associated with 1month later' case numbers. Thus, we concluded that measures taken in the current day affect the mobility of that day which then affects the numbers of COVID-19 cases 1 month later. Since numbers of cases were staying between Poland and South Korea in general, we modelled Turkey as a case. By limiting the modelling period to the end of October to mid-December, 2020 which represented the

second peak period of COVID-19 in Turkey and by optimizing the parameters of the MLP algorithm based on the trial and error method, we achieved 89.3% prediction accuracy. Thus, we concluded that by using mobility data and implementing big data analytics technique, high-performing models may be obtained for COVID-19 spread prediction and this may enable decision-making in managing highly uncertain environments created by this outbreak. Finally, we created different scenarios based on the main descriptive statistics of the mobility data of the study countries and generated predictions on the numbers of cases based on the obtained model. By comparatively analysing these predicted values, we proposed implications for policymakers on deciding on the targeted levels of mobility to maintain numbers of cases in a manageable range. In future research studies, we plan to validate the proposed model with data sets of different countries and the following periods' data of selected countries of this study.

Data availability statement

The datasets were derived from sources in the public domain: OWID. 2020. *Our World in Data, Covid Statistics*. <https://covid.ourworldindata.org/data/owid-covid-data.xlsx>, Google. 2020. *Mobility Data*. <https://www.google.com/covid19/mobility>, Hale T, Webster S, Petherick A, Phillips T, Kira B. 2020. *Oxford COVID-19 Government Response Tracker, Blavatnik School of Government*. <https://www.bsg.ox.ac.uk/research/research-projects/oxford-covid-19-government-response-tracker>.

Funding

This research does not have any funding.

Acknowledgements

The authors acknowledge Assoc. Prof. Dr Serpil Kahraman for sharing her valuable expertise.

Ethical approval

Ethical approval for this type of study is not required by our institute.

Conflict of interest statement

The authors declare that they have no conflict of interest.

References

Akter S, Wamba SF. 2019. Big data and disaster management: a systematic review and agenda for future research. *Annals of Operations Research* 283: 939–59.

Arora R, Suman S. 2012. Comparative analysis of classification algorithms on different datasets using WEKA. *International Journal of Computer Applications* 54: 21–5.

Bryant P, Elofsson A. 2020. Estimating the impact of mobility patterns on COVID-19 infection rates in 11 European countries. *PeerJ* 8: e9879.

Brzezinski A, Deiana G, Kecht V, Van Dijcke D. 2020. The COVID-19 pandemic: government versus community action across the United States. *Covid Economics: Vetted and Real-Time Papers* 7: 115–56.

Cheng C, Barceló J, Hartnett AS, Kubinec R, Messerschmidt L. 2020. COVID-19 government response event dataset (CoronaNet v. 1.0). *Nature Human Behaviour* 4: 756–68.

Davies NG, Kucharski AJ, Eggo RM *et al.* 2020. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health* 5: e375–85.

Engle S, Stromme J, Zhou A. 2020. Staying at home: mobility effects of COVID-19. *Covid Economics, Vetted and Real-Time Papers* 4: 86–102.

Fang H, Wang L, Yang Y. 2020. Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in China. *Journal of Public Economics* 191: 104272.

Ferguson N, Laydon D, Nedjati Gilani G *et al.* 2020. Report 9: impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. London: Imperial College London.

Flaxman S, Mishra S, Gandy A *et al.* 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 584: 257–61.

Frey CB, Chen C, Presidente G. 2020. Democracy, culture, and contagion: political regimes and countries responsiveness to Covid-19. *Covid Economics* 18: 1–20.

Gilson L, Marchal B, Ayepong I *et al.* 2020. What role can health policy and systems research play in supporting responses to COVID-19 that strengthen socially just health systems? *Health Policy and Planning* 35: 1231–6.

Google. 2020. *Mobility Data*. <https://www.google.com/covid19/mobility>, accessed 15 January 2021.

Guo M, Zhang Q, Liao X, Chen FY, Zeng DD. 2020. A hybrid machine learning framework for analyzing human decision-making through learning preferences. *Omega* 101: 102263.

Hadjidemetriou GM, Sasidharan M, Kouyialis G, Parlikad AK. 2020. The impact of government measures and human mobility trend on COVID-19 related deaths in the UK. *Transportation Research Interdisciplinary Perspectives* 6: 100167.

Haldane V, Morgan GT. 2021. From resilient to transilient health systems: the deep transformation of health systems in response to the COVID-19 pandemic. *Health Policy and Planning* 36: 134–35.

Hale T, Webster S, Petherick A, Phillips T, Kira B. 2020. *Oxford COVID-19 Government Response Tracker, Blavatnik School of Government*. <https://www.bsg.ox.ac.uk/research/research-projects/oxford-covid-19-government-response-tracker>, accessed 28 January 2021.

Hensen B, Mackworth-Young CRS, Simwanga M *et al.* 2021. Remote data collection for public health research in a COVID-19 era: ethical implications, challenges and opportunities. *Health Policy and Planning* 36: 360–68.

Jia Q, Guo Y, Wang G, Barnes SJ. 2020. Big data analytics in the fight against major public health incidents (Including COVID-19): a conceptual framework. *International Journal of Environmental Research and Public Health* 17: 6161.

Kaplan EH. 2020. Covid-19 scratch models to support local decisions. *Manufacturing and Service Operations Decisions* 22: 645–55.

The Lancet Public Health. 2020. COVID-19 puts societies to the test. *The Lancet Public Health* 5: e235.

Maloney WF, Taskin T. 2020. Determinants of social distancing and economic activity during COVID-19: A global view. World Bank Policy Research Working Paper (9242).

Mandal M, Jana S, Nandi SK *et al.* 2020. A model based study on the dynamics of COVID-19: prediction and control. *Chaos, Solitons, and Fractals* 136: 109889.

- Mendolia S, Stavrunova O, Yerokhin O. 2021. Determinants of the community mobility during the covid-19 epidemic: the role of government regulations and information. *Journal of Economic Behavior and Organization* 184: 199–231.
- Mohamadou Y, Halidou A, Kapen PT. 2020. A review of mathematical modelling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Applied Intelligence* 50: 3913–25.
- Ofli F, Meier P, Imran M *et al.* 2016. Combining human computing and machine learning to make sense of big (aerial) data for disaster response. *Big Data* 4: 47–59.
- OWID. 2020. *Our World in Data, Covid Statistics*. <https://covid.ourworldindata.org/data/owid-covid-data.xlsx>, accessed 11 January 2020.
- Prem K, Liu Y, Russell T *et al.* 2020. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health* 5: e261–e70.
- Rumelhart DE, Hinton GE, Williams RJ. 1986. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL (eds). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. 1st edn Cambridge, MA: MIT Press pp. 318–62.
- Song X, Zhang H, Akerkar RA *et al.* 2020. Big data and emergency management: concepts, methodologies, and applications. *IEEE Transactions on Big Data* 14: 1–24.
- Soucy Jean-Paul R, Sturrock SL, Berry I *et al.* 2020. Estimating the effect of physical distancing on the COVID-19 pandemic using an urban mobility index. *medRxiv*. Preprint; [10.1101/2020.04.05.20054288](https://doi.org/10.1101/2020.04.05.20054288).
- Wang P, Zheng X, Li J, Zhu B. 2020. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons, and Fractals* 139: 110058.
- WHO. 2020. *World Health Organization, Global Health Workforce Statistics*. <https://www.who.int/data/gho/data/indicators/indicator-details/GHO>, accessed 08 January 2021.
- Yang Z, Zeng Z, Wang K *et al.* 2020. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease* 12: 165.
- Yilmazkuday H. 2020. Stay-at-home works to fight against COVID-19: international evidence from Google mobility data. *Journal of Human Behavior in the Social Environment* 31: 210–20.
- Zhao Z, Li X, Liu F *et al.* 2020. Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. *Science of the Total Environment* 729: 138959.