



OPEN

Quantitative relationships between national cultures and the increase in cases of novel coronavirus pneumonia

Ningyao Yu , Le Tao & Guilin Zou

Support vector machine (SVM) and genetic algorithm were successfully used to predict the changes in the prevalence rate (ΔPR) measured by the increase of reported cases per million population from the 16th to the 45th day during a nation's lockdown after the COVID-19 outbreak. The national cultural indices [individualism–collectivism (Ind), tightness–looseness (Tight)], and the number of people per square kilometer (Pop_density) were used to develop the SVM model of $\ln \Delta PR$. The SVM model has R^2 of 0.804 for the training set (44 samples) and 0.853 for the test set (11 samples), which were much higher than those (0.416 and 0.593) of the multiple linear regression model. The statistical results indicate that there are nonlinear relationships between $\ln \Delta PR$ and Tight, Ind, and Pop_density. It is feasible to build the model for $\ln \Delta PR$ with SVM algorithm. The results suggested that the risk of COVID-19 epidemic spread will be reduced if a nation implements severe measures to strengthen the tightness of national culture and individuals realize the importance of collectivism.

According to the data released by WHO (<https://www.who.int/>), the number of confirmed cases of COVID-19 worldwide reached 562 672 324, of which 6 367 793 died, since the outbreak of COVID-19 in Wuhan in December 2019 to 21 July 2022 at 00:17 am GMT+8. The COVID-19 pandemic has seriously threatened people's lives and health, and brought about serious impacts on global societies, economies and politics¹. In the face of the unprecedented global public health crisis of COVID-19, governments with different national cultures will adopt different response measures, resulting in different epidemic prevention effects. The establishment of prediction models for COVID-19 prevalence rate is of great significance for assisting the government to develop a more effective pandemic prevention measures that are based on its own national cultural characteristics.

Since the outbreak of COVID-19, some researchers have used mathematical models to explore the spread models about COVID-19^{2,3}. These models can be roughly divided into two categories, propagation dynamics models and phenomenological models⁴. These studies that belong to the propagation dynamics models include susceptible-infected-recovered (SIR) models^{5–7}, conceptual mathematical model derived from the susceptible-exposed-infected-recovered (SEIR) isolation models⁸, dynamic model of nonlinear equations based on fractional order⁹. Most transmission dynamics models rely heavily on data and parameters that are difficult to obtain precisely, such as population movements, behavioral habits, the basic number of regenerations R_0 , the number of current infections, the mortality rate of infections, or the proportion of asymptomatic (infectious) individuals. In general, any small change in these input data can result in models producing a prediction bias of several tens of times⁴. The phenomenological models require relatively few parameters, such as environmental risk to which the city or region is exposed^{10,11}, population density, temperature¹², and absolute humidity¹³, lockdown time, gross domestic product¹⁴ or GDP per capita, health care expenditure, air pollution levels¹⁵. Recently, Haouari and Mhir¹⁶ proposed a phenomenological model, which iteratively solved the quadratic programming problem using the particle swarm optimization method to predict the number of deaths from the COVID-19. The input data on which the model relies are easy to obtain precisely.

Besides the above factors, national culture is an important factor affecting the effectiveness of COVID-19 epidemic prevention and control measures. In the face of the global spread of the epidemic, governments have implemented various interventions to mitigate or even stop the spread of COVID-19, including strict lockdowns of cities, regions or borders, home isolation, tracking and isolating COVID-19 patients, observing social distancing, mask mandates, frequent hand washing, etc. The implementation of these measures requires not only the firm determination of governments at all levels, but also the conscious compliance of the public. Culture is

College of Marxism, Xiangtan University, Xiangtan 411105, Hunan, China. ✉email: yuny19910602@126.com

a system of behavioral norms that reflect how people behave, and citizens with different cultures may respond differently to the same challenge¹⁷. The group of Cao¹⁸ used hierarchical regression to detailly analyze the relationships between the efficiency of COVID-19 anti-epidemic measures and national tightness–looseness (Tight), individualism–collectivism (Ind), patient median age, the number of people per square kilometer (Pop_density) and other indices. The contributions of this study are:

- Utilize support vector machine (SVM) algorithm to establish a quantitative relationship model between national culture indices and increased cases of novel coronavirus pneumonia, measured by the changes in the prevalence rate (ΔPR): the increase of reported cases per million population from the 16th to the 45th day during a nation's lockdown after the COVID-19 outbreak.
- Predict the increase in cases of novel coronavirus pneumonia in a country (or region), for assisting governments to develop more effective prevention and control measures based on their own national cultural characteristics to curb the spread of COVID-19.

Methods

The SVM algorithm is a supervised learning method that can be applied to classification (or regression) tasks. Since the SVM appeared in the late 1960s, this technology has been deeply improved and applied in many different fields. SVMs are called maximum interval classifiers because they find the optimal hyperplane determined by many support vectors between the two classes. Their excellent generalization ability is attributed to the introduction of penalty factor C , which allows a certain number of misclassifications to prevent the effects of outliers.

The SVM algorithm initially provided only linear classification. Due to the introduction of kernel functions, SVMs can also deal with nonlinear problems. The principle is to use kernel functions to map data points to high-dimensional feature spaces, and then carry out linear classification or linear regression. Thus, each dot product is replaced by a nonlinear kernel function, allowing the technique to find the largest spaced hyperplane in transformed high-dimensional spaces. There are many kernel functions available, among which the radial basis function (RBF)-based method is widely used.

SVM algorithms based on RBF kernel function have two parameters (penalty factor C and width coefficient γ) that must be optimized because they affect the predictive performance of a model. The parameter C is a regularization parameter, which controls the trade-off between achieving a low error on the training data and minimizing the norm of the weight. If the value is too larger, it increases the penalty for inseparable points, resulting in the use of too many support vectors and overfitting. Conversely, if C is too smaller, there may be a poor fit. Similar to C , the γ parameter describing the width of the RBF kernel results in overfit or underfit for the model when the γ value is too larger or too smaller^{19,20}.

Usually, genetic algorithm (GA) and particle swarm optimization (PSO) algorithms can be used in C and γ optimization. The former is superior to the latter in the optimization speed, so genetic algorithm was used to find the best C and γ parameters in this paper. This algorithm mimics the process of natural evolution and genetics, where the process of mutation, selection, and crossover results in the fittest individuals surviving. Figure 1 shows the flow chart of SVM based on genetic algorithm²¹. Firstly, the problem to be solved is encoded as a chromosome or individual in the genetic space; Then the genetic manipulation steps are followed that include (1) selection of the winning individual from the population, (2) crossover by choosing a random position on the binary string and exchanging the segments with another string partitioned similarly, and (3) variation (a small probability of mutation randomly changing certain genetic of individuals in the population). Finally, the optimization process

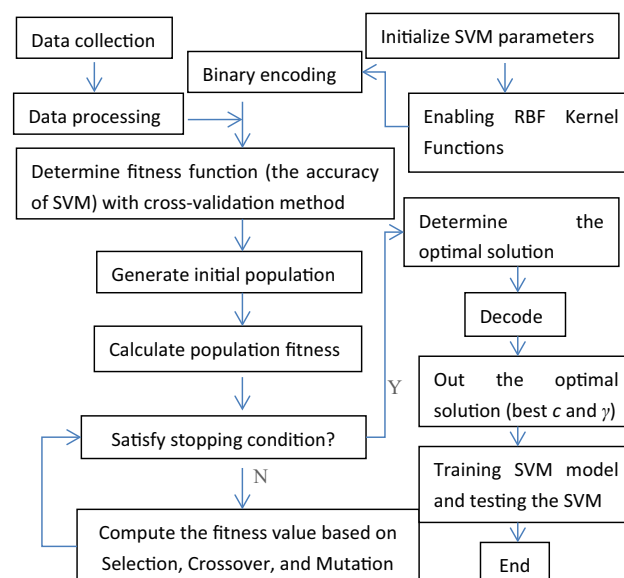


Figure 1. Flow chart of SVM SVM based on genetic algorithm.

stops when the fitness of the optimal individual reaches a given threshold or the number of iterations reaches a preset number. The candidate solutions of the next generation (individual or chromosome) are superior to the previous generation because they inherit the excellent traits of the previous generation's solutions. That is to say, the candidate solutions evolve in the direction of optimal solutions. In optimization, genetic algorithm parameters such as the maximum number of generations, the maximum population size, the crossover probability, and the mutation probability need to be set in advance²⁰.

The data in Table 1 is derived from the literature¹⁸, where increased cases of novel coronavirus pneumonia (Δ PR) are measured by the increase of reported cases per million population from the 16th to the 45th day during a nation's lockdown after the COVID-19 outbreak. Tight is a national culture index of tightness–looseness. The culture index Ind denotes the individualism–collectivism. Pop_density is the number of people per square kilometer. The Δ PR values are serious uneven distribution, e.g., the maximum value (Singapore Δ PR = 3536) is 7072 times of the lowest value (Vietnam Δ PR = 0.5). Thus, Δ PR values are converted to the natural logarithmic form, $\ln\Delta$ PR, in this paper.

Results and discussions

By using SPSS 19.0 software, stepwise multiple linear regression (MLR) analysis was performed for the increase in cases of novel coronavirus pneumonia ($\ln\Delta$ PR) and the individualism–collectivism (Ind), tightness–looseness (Tight), and the number of people per square kilometer (Pop_density), national policy severity index Govt_Strgcy, and median age Median_age¹⁸. A regression equation containing three variables (Tight, Ind, and Pop_density) was obtained. The data set was randomly divided into a training set with 44 samples (80%) and a test set having 11 samples (20%), which are listed in Table 1. The training set is used to build a model, and the test set is used to validate the model. Table 2 shows the descriptive statistics of variables. Both of the Skewness and Kurtosis statistic values for Pop_density are greater than 3, suggesting that its data is seriously abnormal.

The coefficients of the linear regression equation obtained are shown in Table 3. The coefficients R^2 of the training set and the test set are 0.416 and 0.593, respectively. It is generally regarded that a regression model has statistical significance if its coefficient of determination R^2 is greater than 0.5. But the linear model in this paper has a determination coefficient R^2 of $0.416 < 0.5$, indicating that the MLR model possesses poor statistical results. Therefore, the SVM algorithm was used to improve the statistical quality of the models for the increase in cases of novel coronavirus pneumonia $\ln\Delta$ PR, by applying the LIBSVM toolbox on the MATLAB R2014a platform. During the modeling process, the parameter values in GA were set as follows: the maximum number of generations being 200, the maximum population size being 20. In addition, the default values were adopted for the crossover probability and the mutation probability. The leave-one-out (LOO) method was used for cross-validation in the optimization of SVM parameters (C and γ). This method uses one observation from the training set as the validation data, and the remaining $n - 1$ (n : the total number of samples in the training set) observations as the training data²². This is repeated n times such that each observation in the training set is used once as the validation data. The mean squared error (MSE) from cross-validation was used to evaluate the SVM parameters. The search range of SVM parameter C was $1 \sim 200$, and the γ parameter varied from 0 to 0.1. In the end, the optimal parameters of $C = 126.0$ and $\gamma = 1.80 \times 10^{-5}$ were obtained.

Based on the optimal SVM parameters ($C = 126.0$; $\gamma = 1.80 \times 10^{-5}$), the 11 samples in the test set were predicted. The prediction results from the optimal SVM model are listed in Table 1 and the statistical results are listed in Table 4. The coefficient of determination R^2 from the optimal SVM mode are 0.804 for the training set and 0.853 for the test set, which are much higher than the acceptance criteria ($R^2 = 0.5$), indicating that the optimal SVM model is statistically significant. For the training set, the SVM model yielded the Akaike information criterion (AIC) value of -7.22 , the Bayesian information criterion (BIC) value of -1.86 , and the mean absolute percentage error (MAPE) of 11.80%, which are respectively lower than that (39.85, 45.20 and 37.92%) from the MLR model. In addition, the coefficient of determination R^2 in the optimal SVM mode are much larger than that (0.416 and 0.593, respectively) in the MLR model, indicating that there are nonlinear relationships between the dependent variable $\ln\Delta$ PR and the independent variable (Tight, Ind, Pop_density). Therefore, it is appropriate to use SVMs to model the increase in cases of novel coronavirus pneumonia. Figure 2 shows the relationships between the predicted prevalence rate $\ln\Delta$ PR and the actual $\ln\Delta$ PR.

Table 3 shows that the two national culture indices including individualism–collectivism (Ind), tightness–looseness (Tight), and the number of people per square kilometer (Pop_density) have sig.-values less than 0.05, indicating that all the three parameters are significantly related to the increase in cases of novel coronavirus pneumonia ($\ln\Delta$ PR). In addition, the VIF value of the three parameters is less than five, indicating that there is no serious multicollinearity problem among the three parameters, Ind, Tight, and Pop_density, that is, the three parameters can independently reflect the influencing factors of the increase in cases of novel coronavirus pneumonia.

Model validation from the 11 samples in test set leads to the determination coefficient between the predicted and reported $\ln\Delta$ PR being $R^2 = 0.853$; the validation coefficient $q_{\text{ext}}^2 = 0.673$; at intercept = 0, the slopes k' of regression line (experimental versus predicted $\ln\Delta$ PR) = 0.953 and k (predicted versus experimental $\ln\Delta$ PR) = 1.033; determination coefficient R_0^2 (experimental versus predicted $\ln\Delta$ PR) = 0.922; R_0^2 (predicted versus experimental $\ln\Delta$ PR) = 0.768; and $r_m^2 (= R^2 \times [1 - (R^2 - R_0^2)^{1/2}]) = 0.604$. The optimal SVM model meets the criteria²³: $q_{\text{ext}}^2 > 0.5$; $R^2 > 0.6$; $(R^2 - R_0^2)/R^2 (= 0.0996) < 0.1$; $0.85 \leq k' \leq 1.15$; and $r_m^2 > 0.5$, which suggests that the optimal SVM model proposed is successful.

Systematic bias test was further carried out for the optimal SVM model. A regression model should not meet any one of the following conditions, otherwise it possesses systematic bias in predictions²⁴: (1) $\text{NPE}/\text{NNE} > 5$ or $\text{NNE}/\text{NPE} > 5$ (NPE: the number of positive errors, NNE: the number of negative errors); (2) $\text{ABS}(\text{MPE}/\text{MNE}) > 2$ or $\text{ABS}(\text{MNE}/\text{MPE}) > 2$ (ABS: absolute value, MPE: the mean positive error, MNE: the mean negative error);

No	Country	Tight	Ind	Pop_ density	lnΔPR/ reported	lnΔPR/SVM	lnΔPR/MLR
Training set							
1	Japan	42.8	46	347.8	3.68	3.67	5.37
2	Belgium	98.1	75	375.6	8.01	8.00	7.67
3	Argentina	73.4	46	16.2	4.62	5.41	5.91
4	Netherlands	62.6	80	508.5	7.36	7.35	7.10
5	Portugal	67.3	27	112.4	7.37	5.71	5.14
6	Austria	71.6	55	106.7	6.46	6.47	6.23
7	Lithuania	46.1	60	45.1	4.10	6.63	5.80
8	Venezuela	46.0	12	36.3	1.95	4.68	4.08
9	India	81.9	48	450.4	3.67	3.68	6.37
10	Indonesian	3.0	14	145.7	3.63	3.62	3.23
11	Vietnam	30.7	20	308.1	−0.69	−0.68	4.15
12	Iran	23.6	41	49.8	6.86	5.43	4.62
13	Singapore	42.2	20	7915.7	8.17	8.16	7.93
14	China	38.2	20	147.7	3.54	3.55	4.24
15	Luxembourg	94.7	60	231.4	7.77	7.76	6.99
16	South Korea	29.7	18	528.0	4.07	4.08	4.16
17	South Africa	71.5	65	46.8	4.84	5.58	6.56
18	Hungary	43.7	80	108.0	5.28	5.29	6.49
19	United Kingdom	74.5	89	272.9	7.71	7.70	7.58
20	Jordan	3.9	30	109.3	2.87	4.34	3.80
21	Pakistan	7.0	14	255.6	4.59	4.58	3.37
22	Canada	77.0	80	4.0	7.02	7.01	7.20
23	Morocco	6.1	46	80.1	4.69	4.70	4.41
24	Chile	82.0	23	24.3	6.67	6.66	5.29
25	Estonia	51.3	60	31.0	6.72	6.71	5.91
26	Egypt	2.6	25	98.0	3.97	3.96	3.59
27	Bangladesh	3.2	20	1265.0	4.40	4.41	3.97
28	Saudi Arabia	20.4	25	15.3	5.86	5.87	3.96
29	Tanzania	23.3	25	64.7	2.08	3.51	4.05
30	Mexico	80.7	30	66.4	5.18	5.19	5.53
31	Slovenia	74.6	27	102.6	5.84	5.85	5.30
32	Nigeria	19.4	30	209.6	3.11	3.12	4.20
33	Ireland	62.1	70	69.9	8.01	6.62	6.53
34	Turkey	21.5	37	104.9	6.39	5.32	4.45
35	Bulgaria	60.5	30	65.2	4.98	4.97	5.07
36	Peru	47.8	16	25.1	6.91	6.05	4.26
37	Slovakia	65.8	52	113.1	5.26	6.11	5.99
38	Greece	61.8	35	83.5	4.47	5.64	5.28
39	Iceland	42.4	60	3.4	7.73	7.74	5.70
40	Poland	50.9	60	124.0	5.60	5.59	5.95
41	Denmark	60.2	74	136.5	7.03	7.02	6.66
42	Finland	66.2	63	18.1	6.36	6.10	6.35
43	Ukraine	59.0	25	77.4	5.42	5.30	4.86
44	Spain	80.6	51	93.1	8.01	6.41	6.28
Test set							
45	Philippines	39.3	32	351.9	4.09	2.62	4.79
46	Czech Rep	58.8	58	137.2	6.06	5.78	6.06
47	Germany	71.8	67	237.0	7.00	7.59	6.72
48	Italy	66.4	76	205.9	7.62	9.28	6.91
49	Russia	53.6	39	8.8	7.34	7.03	5.21
50	France	82.8	71	122.6	7.08	8.31	7.06
51	United States	60.8	91	35.6	7.89	8.68	7.24
52	Albania	35.0	20	104.9	5.23	5.87	4.15
Continued							

No	Country	Tight	Ind	Pop_density	lnΔPR/ reported	lnΔPR/SVM	lnΔPR/MLR
53	Sweden	61.2	71	24.7	7.31	6.93	6.53
54	Romania	46.0	30	85.1	6.26	5.53	4.75
55	Croatia	53.6	33	73.7	5.61	5.23	5.02

Table 1. Increased cases of novel coronavirus pneumonia in 55 countries and parameters used.

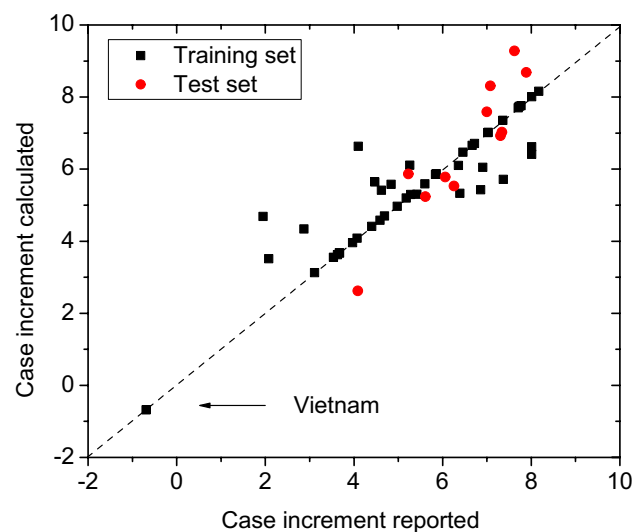
Variable	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Tight	55	2.60	98.10	50.9309	24.87817	-0.407	-0.575
Ind	55	12.00	91.00	44.9455	22.44204	0.337	-1.140
Pop_density	55	3.40	7915.70	297.8291	1064.77005	7.044	51.066
lnΔPR/reported	55	-0.69	8.17	5.6193	1.86034	-0.864	1.004

Table 2. Descriptive statistics.

Variable	Coefficient	S.E	t	Sig	VIF
Constant	2.595	0.582	4.462	0.000	—
Tight	0.023	0.010	2.226	0.032	1.352
In	0.036	0.013	2.829	0.007	1.386
Pop_density	4.63E-4	2.01E-4	2.306	0.026	1.031

Table 3. Coefficients for linear regression equation.

Model	Subset	N	R ²
MLR	Training set	44	0.416
	Test set	11	0.593
SVM	Training set	44	0.804
	Test set	11	0.853

Table 4. Comparison of model statistics in MLR and SVM models.**Figure 2.** Relationships between the predicted and reported prevalence rates (generated by the OriginPro 8.5.0 SR1 software—<http://www.OriginLab.com>).

(3) $AAE - ABS(AE) < 0.5 \times AAE$ (AAE: the average absolute error, AE: the average error); (4) $R^2(\text{ith vs } (i-1)\text{th residuals}) > 0.5$ for residuals sorted on Y_{obs} ; (5) $R^2(Y \text{ vs. residuals}) > 0.5$.

The results of systematic bias test are as follow: $NPE/NNE = 6/5 < 5$; $ABS(MPE/MNE) = 0.982/0.592 < 2$; $AAE - ABS(AE) = 0.769 - 0.124 = 0.645 > 0.5 \times 0.769 = 0.385$; $R^2(\text{ith vs } (i-1)\text{th residuals}) = 0.002 < 0.5$ for residuals sorted on Y_{obs} ; $R^2(Y \text{ vs residuals}) = 0.390 < 0.5$. These results suggest that the optimal SVM model ($C = 126.0$; $\gamma = 1.80 \times 10^{-5}$) does not meet above circumstances, that is, it has no systematic error in predicting $\ln\Delta PR$ ²⁴.

Figure 3 shows the Williams plot, which reflects the relationships between the standardized residuals predicted by the optimal SVM model and the leverage values. The Williams plot is used to describe the application domain of the optimal model. The prediction results are reliable only if the prediction points fall into this domain, where the absolute values of the standardized residual are not more than 3 and the leverage values are less than the warning leverage value h^* . h^* is calculated with $h^* = 3 \times (p+1)/n$, where p is the number of parameters and n is the number of samples in the model. Figure 3 shows that only the sample point of Singapore is outside the applicability domain, indicating that the argument sample is an "outlier" compared to other sample variables. But its standardized residuals being close to 0 indicates that the optimal SVM model has a strong predictive power for the increase in cases of novel coronavirus pneumonia $\ln\Delta PR$ ²⁰. Similarly, according to the median absolute deviation method^{25,26}, the sample point of Vietnam (No. 11) has a determination coefficients D of 3.22, above the threshold of 2.5, and belongs to outliers. However, its $\ln\Delta PR$ value predicted with the SVM model is accurate compared with that from the MLR model.

In this paper, the SVM model of the increase in cases of novel coronavirus pneumonia $\ln\Delta PR$ contains three parameters, individualism–collectivism (Ind), tightness–looseness (Tight), and the number of people per square kilometer (Pop_density). As is obvious, there are correlations between population density Pop_density and $\ln\Delta PR$, because an increasing population density, Pop_density, leads to people more likely to be in close contact, and results in higher ΔPR . Thus, this article focuses on the relationships between $\ln\Delta PR$ and the national cultural indices Ind and Tight.

For decades, cultural studies have attracted scholars from a variety of disciplines, including anthropologists, sociologists, and psychologists. Despite its universality, ambiguity, and diversity, it is widely believed that culture encompasses values, beliefs, norms, and other factors that guide and influence human behavior in society¹⁸. The famous Dutch social psychologist, Geert Hofstede^{27,28} pioneered six dimensions to measure the cultural systems of different countries, of which, the individualism–collectivism culture is the most commonly used one, with the strongest predictive power. The individualism–collectivism index Ind represents the extent to which the individual sees himself primarily as an autonomous entity (individualism) or embedded in a closely connected group (collectivism). Collectivist culture prioritizes the social role of the individual in the group and places it above individual achievement. Conversely, an individualistic culture refers to an individual acting in his or her own interests, without any social affiliation or obligation, and with no intention of protecting the interests of other members^{18,29}. The individualism–collectivism culture index Ind varies from 1 to 100 points. The higher the score, the more the national culture is displayed as individualistic. In a society emphasizing collective achievement, people tend to adhere strictly to government restrictive directives and recommendations. Conversely, individualists may be more concerned about the interests of themselves and their immediate family. The higher the Ind value, the more individualistic the nation, which naturally hinders efforts to contain the spread of the virus, resulting in an increase in ΔPR . Therefore, ΔPR is positively correlated with Ind.

Professor Michele Gelfand, a prominent cultural psychologist at the University of Maryland, has studied the national culture index, tightness–looseness³⁰ and argued that cultural tightness comes from the strength of social norms and the intensity of sanctions. Cultures in which groups require individuals to adhere to group norms and have less tolerance for behavior that deviates from group norms are defined as tight cultures, while cultures

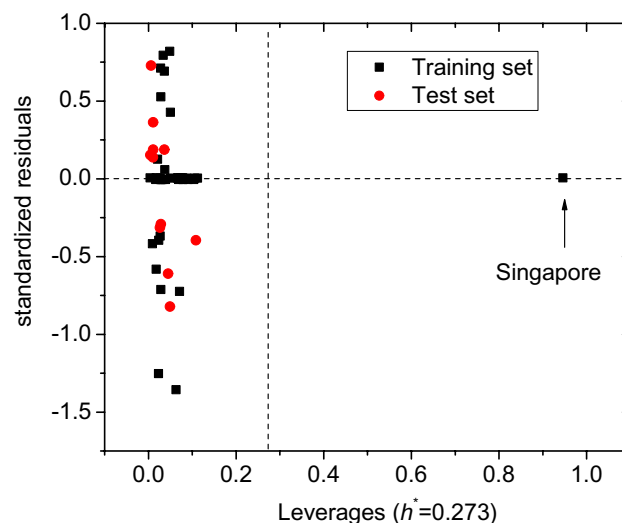


Figure 3. The standardized residuals vs the leverage values (generated by the OriginPro 8.5.0 SR1 software—<http://www.OriginLab.com>).

that are less normative and can tolerate individual deviations are defined as loose cultures. The national tightness–looseness culture index Tight can be examined at three levels: (1) microscopic, individual or community level, (2) mesoscopic, regional, city or provincial level, and (3) macroscopic, or national level¹⁸. Gelfand et al.³¹ based on a survey of nearly 7000 people in 33 countries, asked respondents to assess the veracity of some of their statements, and found that the greater the pressure on a society, whether it is a natural or man-made disaster, the tighter the culture formed by this society. For the first time, they proposed the quantitative description of cultural tightness–looseness, that is, the Gelfand tightness–looseness index. Another cultural psychologist, Irem Uz³² developed the Gelfand tightness–looseness index, proposing the Uz tightness–looseness index. This index has an advantage over the Gelfand tightness–looseness index because Uz et al. surveyed a wider scope and more countries. Thus, the Uz tightness–looseness index has been widely used. The lower the value of the national tightness–looseness index Tight, the greater the intensity of social norms and sanctions in the nation's culture, e.g., Japan, Singapore, and China. The larger the Tight, the more loose the national cultural identity (e.g., Italy, the United States, France), which may have a negative effect on government recommendations such as stay-at-home quarantine, mask mandates, observing social distancing, leading to the more increased cases of novel coronavirus pneumonia Δ PR in the nation³³. Therefore, Δ PR is also positively correlated with Tight.

It should be pointed out that the proposed model possesses a relatively smaller data set, especially the lack of samples from African countries. Collecting more data on Δ PR will make the model more representative.

Conclusions

By using the national culture tightness–looseness, Tight, individualism–collectivism, Ind, together with the number of people per square kilometer, Pop_ density as independent variables, and the increase in cases of novel coronavirus pneumonia $\ln\Delta$ PR as the dependent variable, quantitative relationship models between the independent variables and the dependent variable were successfully established with the GA and SVM algorithms. The accuracy of the SVM model in predicting the increase in cases of novel coronavirus pneumonia $\ln\Delta$ PR from the 16th to the 45th day during a nation's lockdown after the COVID-19 outbreak is much greater than that of the linear model. The use of SVM to establish a $\ln\Delta$ PR model is successful. Based on the analysis of the SVM model, the larger the individualism–collectivism Ind index, the more the national culture is displayed as individualism, which is not conducive to curbing the spread of the virus, resulting in an increase in the number of COVID-19 infection cases. The larger the national culture tightness–looseness index Tight, the looser the cultural identity of the nation, which may reduce people's conscious compliance with government instructions and recommendations and reduce the effectiveness of epidemic prevention measures. The results suggested that the risk of COVID-19 epidemic spread will be reduced if a nation implements severe measures to strengthen the tightness of national culture and individuals are aware of the importance of collectivism. In addition, reducing population density can help reduce people's access, thereby reducing the spread of the epidemic.

Data availability

All data generated or analysed during this study are included in this published article.

Received: 25 July 2022; Accepted: 27 January 2023

Published online: 30 January 2023

References

1. Alsayed, A. R. M. Turkish stock market from pandemic to Russian invasion, evidence from developed machine learning algorithm. *Comput. Econ.* <https://doi.org/10.1007/s10614-022-10293-z> (2022).
2. Arti, M. K. Mathematical modeling of COVID-19 and prediction of upcoming wave. *IEEE J. Sel. Top. Signal. Process.* **1**, 300–306 (2022).
3. Alqahtani, R. T. Mathematical model of SIR epidemic system (COVID-19) with fractional derivative: stability and numerical analysis. *Adv. Differ. Equ.* **2021**, 2 (2021).
4. Valvo, P. S. A bimodal lognormal distribution model for the prediction of COVID-19 deaths. *Appl. Sci.* **10**, 8500 (2020).
5. Abdy, M., Side, S., Annas, S., Nur, W. & Sanusi, W. An SIR epidemic model for COVID-19 spread with fuzzy parameter: The case of Indonesia. *Adv. Differ. Equ.* **2021**, 105 (2021).
6. Aryatama, F. Y., Kurniadi, F. I. & Manik, N. I. Mathematical model estimation and prediction application of Covid-19 infection in Indonesia using Levenberg–Marquardt Algorithm based on Python. *Proc. Comp. Sci.* **216**, 120–127 (2023).
7. Yajada, M., Moridani, M. K. & Rasouli, S. Mathematical model to predict COVID-19 mortality rate. *Infect. Dis. Model.* **7**, 761–776 (2022).
8. Buhat, C. A. H. et al. A mathematical model of COVID-19 transmission between frontliners and the general public. *Netw. Model. Anal. Health Inform. Bioinform.* **10**, 17 (2021).
9. Mohammad, M., Trounev, A. & Cattani, C. The dynamics of COVID-19 in the UAE based on fractional derivative modeling using Riesz wavelets simulation. *Adv. Differ. Equ.* **2021**, 115 (2021).
10. Coccia, M. An index to quantify environmental risk of exposure to future epidemics of the COVID-19 and similar viral agents: Theory and practice. *Environ. Res.* **191**, 110155 (2020).
11. Barcellos, D. D. S., Fernandes, G. M. K. & de Souza, F. T. Data based model for predicting COVID-19 morbidity and mortality in metropolis. *Sci. Rep.* **11**, 24491 (2021).
12. Alsayed, A. R. M. Association between coronavirus cases and seasonal climatic variables in Mediterranean European Region, evidence by panel data regression. *Int. J. Environ. Sci. Technol.* **19**, 8265–8272 (2022).
13. Diao, Y. et al. Influence of population density, temperature, and absolute humidity on spread and decay durations of COVID-19: A comparative study of scenarios in China, England, Germany, and Japan. *One Health* **12**, 100203 (2021).
14. Coccia, M. The relation between length of lockdown, numbers of infected people and deaths of Covid-19, and economic growth of countries: Lessons learned to cope with future pandemics similar to Covid-19 and to constrain the deterioration of economic system. *Sci. Total Environ.* **775**, 145801 (2021).
15. Coccia, M. 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. *Environ. Res.* **199**, 111339 (2021).

16. Haouari, M. & Mhiri, M. A particle swarm optimization approach for predicting the number of COVID-19 deaths. *Sci. Rep.* **11**, 16587 (2021).
17. Jolly, P. M. *et al.* Quantifying cultural tightness-looseness in Ecuador. *PLoS ONE* **16**(1), e0246064 (2021).
18. Cao, C., Li, N. & Liu, L. Do national cultures matter in the containment of COVID-19? *Int. J. Sociol. Soc. Policy.* **40**(9/10), 939–961 (2020).
19. Yu, X. Quantitative structure-toxicity relationships of organic chemicals against *Pseudokirchneriella subcapitata*. *Aquat. Toxicol.* **224**, 105496 (2020).
20. Yu, X. & Wang, H. Support vector machine classification model for color fastness to ironing of vat dyes. *Text. Res. J.* **91**(15–16), 1889–1899 (2021).
21. Yu, Y. Extrapolation for aeroengine gas path faults with SVM bases on genetic algorithm. *Sound Vib.* **53**(5), 237–243 (2019).
22. Yu, X. & Zeng, Q. Random forest algorithm-based classification model of pesticide aquatic toxicity to fishes. *Aquat. Toxicol.* **251**, 106265 (2022).
23. Yu, X. Prediction of chemical toxicity to *Tetrahymena pyriformis* with four descriptor models. *Ecotox. Environ. Safe.* **190**, 110146 (2020).
24. Fang, Z., Yu, X. & Zeng, Q. Random forest algorithm-based accurate prediction of chemical toxicity to *Tetrahymena pyriformis*. *Toxicology* **48**, 153325 (2022).
25. Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **49**(4), 764–766 (2013).
26. Yu, X. Support vector machine-based model for toxicity of organic compounds against fish. *Regul. Toxicol. Pharm.* **123**, 104942 (2021).
27. Hofstede, G., Hofstede, G. J. & Minkov, M. *Cultures and organizations: Software of the mind, revised and expanded* 3rd edn, 89–134 (McGraw-Hill, 2010).
28. Taras, V., Steel, P. & Kirkman, B. L. Improving national cultural indices using a longitudinal meta-analysis of Hofstede's dimensions. *J. World Bus.* **47**(3), 329–341 (2012).
29. Yan, H., Chen, X. & Chen, W. Chinese and American cultural differences reflected by their totems: Chinese dragon and American eagle. *Creat. Educ.* **9**, 1945–1956 (2018).
30. Gelfand, M. J., Nishii, L. H. & Raver, J. L. On the nature and importance of cultural tightness-looseness. *J. Appl. Psychol.* **91**(6), 1225–1244 (2006).
31. Gelfand, M. J. *et al.* Differences between tight and loose cultures: A 33-nation study. *Science* **332**(6033), 1100–1104 (2011).
32. Uz, I. The index of cultural tightness and looseness among 68 countries. *J. Cross-Cult. Psychol.* **46**(3), 319–335 (2015).
33. Yan, B., Zhang, X., Wu, L., Zhu, H. & Chen, B. Why Do Countries Respond Differently to COVID-19? A Comparative Study of Sweden, China, France, and Japan. *Am. Rev. Public Adm.* **50**(6–7), 762–769 (2020).

Author contributions

N.Y.: Conceptualization, methodology, software, model development, writing-original draft preparation, manuscript revision; L.T.: Model validation, manuscript revision; G.Z.: Data collection and curation, model development, writing-original draft preparation.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023