ISSN 2616-5872 Vol.6, Issue 4: 46-52, DOI: 10.25236/AJEE.2024.060407

Prediction of urban environmental pollution based on the optimal prediction model

Huijun Yu*

School of Chemical and Environmental Engineering, China University of Mining and Technology (Beijing), Beijing, 100083, China
**Corresponding author: sqt1800302067@student.cumtb.edu.cn*

Abstract: *This paper evaluated the weight of influencing factors on the development of urban environmental quality by studying the relationship between existing environmental pollution and its influencing factors. The prediction models of the combination of water, atmosphere, and solid waste were constructed, based on the non-biased grey model (GM (1,1)), Back Propagation neural network (BPNN), Autoregressive Integrated Moving Average Model (ARIMA), and machine learning algorithms such as gradient descent were used to improve the fitting accuracy. The data from 2001 to 2020 were selected as the training set and the validation set to predict the environmental pollution trend from 2021 to 2028. Research shows that the BPNN has good performance of the three emissions in three models.*

Keywords: *ARIMA; GM (1,1); BPNN; urban environmental quality*

1. Introduction

With the development of industrialization, urban pollution problems are becoming increasingly prominent. Among them, the total emissions of industrial wastewater, industrial sulfur dioxide, and industrial solid waste, led by industrial "three wastes", continue to increase[1]. These environmental pollution problems not only cause severe damage to production and the economy but also have immeasurable negative impacts on ecological systems, wildlife, and human health. With the proposal of the "14th Five-Year Plan" goals for ecological and environmental protection, the future of environmental protection in China will further develop in the direction of pollution control, clean production, and green low-carbon. The emission of "three wastes" serves as a representative indicator for measuring environmental quality, which is affected by many social and economic factors, such as secondary industry, population density, energy consumption, afforestation, etc. Therefore, accurate prediction of "three wastes" pollution emissions is of great significance for further quantifying the sources of pollution, reducing total pollution emissions, and improving environmental carrying capacity[2]. At the same time, it also provides a supporting force for the formulation and adjustment of environmental protection strategies in China.

Recently, researchers mainly focus on three categories of prediction models: statistical analysis models, machine learning models, and grey prediction models[3-5]. However, these models still have certain limitations. Statistical analysis models usually require a large amount of historical data, while in practice, data could be challenging to obtain, and even when available, missing data or data that lack credibility are often occurrences. In addition, the data for statistical analysis models also needs to meet certain conditions, such as they need to follow a normal distribution, otherwise, the established model may not achieve ideal prediction results. As a relatively mature theoretical model, the time series model (ARIMA) can be applied to practical cases simply and conveniently, but its theoretical foundation is mostly based on linear prediction analysis, so the prediction accuracy is not high in dealing with nonlinear prediction problems. To better solve nonlinear prediction problems, scholars have put forward the theory of machine learning theories. Compared with the time series model, machine learning has significant advantages in processing nonlinear data, such as automatic processing of large amounts of data, high generalization ability, and self-learning ability. However, when the data dimension is high and the sample size is insufficient, the model is prone to overfitting and underfitting. BP neural network model, as one of the most efficient machine learning models, can meet different data scenarios and be used in small data sets with many features. The grey prediction model was established by Deng [6] in 1982, which handled many uncertain problems well in the condition of a relatively small amount of data. Therefore, grey prediction models are usually used to solve problems with small, incomplete, and discrete data

ISSN 2616-5872 Vol.6, Issue 4: 46-52, DOI: 10.25236/AJEE.2024.060407

whose characteristics are uncertain[7].

In this paper, the time series algorithm ARIMA, BP neural network, and grey GM (1,1) model were used to construct the optimal prediction models, respectively. The model verification and gradient descent algorithm were used to improve the accuracy of the model, so as to screen out the optimal prediction model. On this basis, the paper will analyze the deep-seated social and economic factors affecting environmental quality, in order to find a balance point for the coordinated development of environment and economy. And to provide a theoretical basis for the formulation of pollution prevention and control policies by predicting the emission of pollutants in the next five years.

2. Setting for model testing

The common model testing indicators include correlation coefficient (R^2), mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean relative error (MAPE), explained variance score, etc. In order to compare the fitting effect of each model, R^2 and MSE were selected for the correlation and error test of the model, respectively[8].

R-squared is a measure used to assess the goodness-of-fit of a regression model. It represents the proportion of the variance in the data that is captured by the model, with values ranging from 0 to 1. A higher R-squared value closer to 1 indicates a better fit of the model to the data, while a lower R-squared value closer to 0 indicates a poorer fit of the model. The formula is as follows:

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} = 1 - \frac{\sum(\hat{y}^{(i)} - y^{(i)})^2}{\sum(\bar{y} - y^{(i)})^2} \tag{1}$$

Here $y^{(i)}$ is the true value of the *i* th sample, $\hat{y}^{(i)}$ is the predicted value of the model, and $\bar{y}$ is the average value of the samples.

Mean Squared Error (MSE) is a common statistical metric used to measure the difference between predicted values and true values. It is calculated as the average of the squared differences between predicted and actual values, and squaring the differences helps to ignore the positive and negative signs. Calculating the average is done to consolidate the MSE across all samples into a single metric for easier comparison. The range of MSE is [0, +∞]. A lower MSE value indicates a better fit of the model to the data, while a higher MSE value approaching +∞ indicates a poorer fit of the model.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y^{(i)} - \hat{y}^{(i)})^2 \tag{2}$$

Where m is the number of samples, $y^{(i)}$ is the true value of the *i* th sample, $\hat{y}^{(i)}$ is the predicted value of the model.

3. Prediction model

3.1 GM (1, 1) model

The modeling process of the GM (1,1) model is as follows [3]:

Consider a system with a sequence of characteristic data:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \tag{3}$$

New generated data is obtained by the above single summation:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)) \tag{4}$$

Among them: $x^{(1)}(m) = \sum_{i=1}^{n} x^{(0)}(i)$, m = 1, 2, …, n, Let $z^{(1)}$ be the immediate neighbor of the sequence $x^{(1)}$, $z^{(1)} = \left(z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n)\right)$, including: $z^{(1)}(m) = \delta x^{(1)}(m) + (1 - \delta)x^{(1)}(m-1)$, m = 2, 3,... , n and δ=0.5.

$x^{(0)}(m) + az^{(1)}(m) = b$ is the basic form of the GM (1, 1) model (k = 2, 3, ... n), where in the GM (1,1) model, a is called the development coefficient and b is called the driving coefficient.

ISSN 2616-5872 Vol.6, Issue 4: 46-52, DOI: 10.25236/AJEE.2024.060407

Introduce matrix-vector notation:

$$u = (a,b)^T, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ x^{(0)}(4) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, Y = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & 1 \\ -z^{(1)}(n) & 1 \end{bmatrix} \tag{5}$$

u can be obtained by least square method:

The prediction model can be solved by the differential equation:

$$\hat{x}^{(1)}(k+1) = \left[x_1^{(0)}(1) - \frac{b}{a}\right] e^{-ak} + \frac{b}{a} \tag{6}$$

By reducing and reducing the eq.6, the grey prediction model of the original series can be obtained as follows:

$$\hat{x}^{(0)}(k) = x^{(1)}(k) - \hat{x}^{(0)}(k-1) \tag{7}$$

3.2 Time sequence model:

The ARIMA model (Autoregressive Integrated Moving Average model) is one of the time series forecasting and analysis methods. It transforms data into stationary data and then builds a model by regressing the dependent variable on its lagged values and random error terms. In ARIMA (p, d, q), AR stands for "autoregressive" with p being the number of autoregressive terms, MA stands for "moving average" with q being the number of moving average terms, and d is the order of differencing needed to make the series stationary. The mathematical expression of the model is shown as follows:

$$\phi(B)\nabla_{x_t}^{a} = \Theta(B)\varepsilon_t \tag{8}$$

Where x_t(t = 1, 2, …, n) is the time series data, ε_t(t = 1, 2, ..., n) is the residual term, B is the lag operator, where $B^n x_t = \mathrm{x_{t-n}}$; Φ(B) is the autoregressive coefficient polynomial, ∇^t is the differencing operator, where $\nabla^t = (1-B)^d$, and Θ(B) is the moving average coefficient polynomial[9].

3.3 BP neural network model

BP neural network (Back Propagation neural network) is a concept proposed by scientists led by Rumelhart and McClelland in 1986. It is a multi-layer feedforward neural network trained according to the error inverse propagation algorithm. It is one of the widely used neural network models. BP neural network has arbitrary complex pattern classification ability and excellent multi-dimensional function mapping ability, which can solve the Exclusive OR, XOR, and some other problems that simple perceptron cannot solve[10].

The basic BP algorithm includes two processes signal forward propagation and error back propagation[4]. That is, the error output is calculated in the direction from input to output, and the adjustment weight and threshold are adjusted in the direction from output to input. In forward propagation, the input signal acts on the output node through the hidden layer and generates the output signal after nonlinear transformation. If the actual output is inconsistent with the expected output, the error is transferred to the reverse propagation process. Error backpropagation is to back-transmit the output error to the input layer by layer through the hidden layer, and allocate the error to all the elements of each layer, and the error signal obtained from each layer is used as the basis for adjusting the weight of each unit. By adjusting the connection strength between the input node and the hidden layer node and the connection strength between the hidden layer node and the output node and the threshold value, the error is reduced along the gradient direction. After repeated learning and training, the network parameters corresponding to the minimum error (weight and threshold) are determined, and the training is stopped. At this time, the trained neural network can process the non-linear transformed information with minimal output error for the input information of similar samples[11].

STEP 1. Forward propagation.

Let $x_1, x_2, \ldots, x_n$ be the input variable, y is the output variable, u_i is the output of the hidden layer neuron, and f is the mapping relationship of the activation function. Let v_{ij} be the weight of the i th input variable and the j th hidden neuron, and let θ_j^u be the threshold of the j th neuron of the hidden layer u.

ISSN 2616-5872 Vol.6, Issue 4: 46-52, DOI: 10.25236/AJEE.2024.060407

Then the expression of u_j is:

$$u_j = f(\sum_{i=1}^{n} v_{ij} x_i + \theta_j^u) \quad j=1,2,\ldots,m \tag{9}$$

STEP 2. Objective function

There are M objects and N features:

$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_i^{(1)} & x_N^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_i^{(2)} & x_N^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1^{(k)} & x_2^{(k)} & \cdots & x_i^{(k)} & x_N^{(k)} \\ x_1^{(M)} & x_2^{(M)} & \cdots & x_i^{(M)} & x_N^{(M)} \end{bmatrix} \cdots \begin{bmatrix} \tilde{y}^{(1)} \\ \tilde{y}^{(2)} \\ \vdots \\ \tilde{y}^{(k)} \\ \tilde{y}^{(M)} \end{bmatrix} \tag{10}$$

Where $x_i^{(k)}$ represents the input of the i th feature. Let x be the true feature of the data, the output value of the true value is $\tilde{y}$, and y is the value predicted by the neural network. In order to minimize the error between the true value and the predicted value through the network, the objective function is set as follows:

$$J = \sum_k \left(y^{(k)} - \tilde{y}^{(k)}\right)^2 \tag{11}$$

STEP 3. Gradient descent

Let y be the target, w be the parameter to be optimized, and u be the rate of decline, i.e. the learning rate.

$$\widetilde{w} = w - u\frac{dy}{dw} \tag{12}$$

The weights are updated, where μ is the learning rate.

$$v_{ij}' = v_{ij} - u\frac{\partial J}{\partial v_{ij}}, \quad w_j' = w_j - u\frac{\partial J}{\partial w_i} \tag{13}$$

The bias update is as follows, the formula for each layer is the same.

$$\theta_j^{u'} = \theta_j^u - u\frac{\partial J}{\partial \theta_i^u}, \quad \theta^{y'} = \theta^y - u\frac{\partial J}{\partial \theta^y} \tag{14}$$

v_{ij}, θ_j^u, w_j, θ^y is the network parameter to be determined, J is the objective function, which can be optimized by gradient descent method.

STEP 4. Find the gradient in reverse

(1) Find the gradient of w_j $(\frac{\partial J^{(k)}}{\partial w_j})$

$$J^{(k)} = (y^{(k)} - \widetilde{y^{(k)}}) \tag{15}$$

$$\frac{\partial J^{(k)}}{\partial w_i} = 2(y^{(k)} - \widetilde{y^{(k)}})\frac{\partial J^{(k)}}{\partial w_i} \tag{16}$$

Let $G = \sum_{j=1}^{m} w_j u_i + \theta^y$, $y = f(G)$

$$\frac{\partial J^{(k)}}{\partial w_i} = \frac{\partial J^{(k)}}{\partial G}\frac{\partial G}{\partial w_j} \tag{17}$$

(2) Find the gradient of θ^y $(\frac{\partial J^{(k)}}{\partial \theta^y})$

$$\frac{\partial J^{(k)}}{\partial \theta^y} = 2(y^{(k)} - \widetilde{y^{(k)}})\frac{\partial J^{(k)}}{\partial G}\frac{\partial G}{\partial \theta^y} \tag{18}$$

(3) Find the gradient of v_{ij} $(\frac{\partial J^{(k)}}{\partial v_{ij}})$

ISSN 2616-5872 Vol.6, Issue 4: 46-52, DOI: 10.25236/AJEE.2024.060407

$$\frac{\partial J^{(K)}}{\partial v_{ij}} = \frac{\partial J^{(K)}}{\partial u_j}\frac{\partial u_j}{\partial v_{ij}}, \tag{19}$$

(4) Find the gradient of θ_j^u $(\frac{\partial J^{(k)}}{\partial \theta_j^u})$

Let $P = \sum_{i=1}^{n} v_{ij}x_i + \theta_j^u,\ u_j = f(P)$,

$$\frac{\partial J^{(K)}}{\partial \theta_j^u} = \frac{\partial J^{(K)}}{\partial u_j}\frac{\partial u_j}{\partial P}\frac{\partial P}{\partial \theta_j^u}, \tag{20}$$

4. Case study

4.1 Selection of subjects and data

According to the China Bureau of Statistics, the permanent population of Beijing will be 21.843,000 in early 2023. Gross Domestic Product (GDP) of Beijing will be about 4.180 billion yuan (~6230 trillion US dollars) in 2022, ranking seventh in the world. As the capital of China, Beijing is one of the representative cities in the world, which not only has a high policy responsiveness in China, but also has a strong influence in the international community[12]. Therefore, the collection of various data in Beijing can not only accurately and quickly reflect the impact of different indicators on regional environmental pollution, but also represent the current level of environmental pollution control in China and the global level of pollution control, so the prediction has far-reaching significance.

In this paper, the annual emissions of three wastes (chemical oxygen demand (COD) of wastewater, sulfur dioxide emission, and smoke dust emission) in Beijing from 1999 to 2022 were selected as the research objectives. The gross domestic product (GDP), total fuel consumption (TFC), Fixed Asset Investment (FAI) and Employment to Population of Beijing were collected ratio (EPOP), Green coverage rate (GCR) and other indicators were used as the factors affecting the annual emissions of three wastes, in which to find the key factors affecting the annual emissions of three wastes in Beijing. The data were selected from the China's National Bureau of Statistics, China Statistical Yearbook, and Beijing Statistical Source Annual Report. The descriptive statistics of the above indicators are shown in Table 1.

Table 1: Description of each index

Index name/index serial number	indicator unit	MAX	MIN	MEAN	SD	ANOVA	Me	Kurt	Skew	CV
GDP/x_1	100 million Yuan	40269.6	2759.8	17318.5	11901.2	14964	1.4E+08	-1.057	0.477	0.687
TFC/x_2	10 thousand tons of standard coal	7360.32	3906.6	5942.39	1108.32	6359.49	1228371	-1.115	-0.523	0.187
FAI/x_3	100 million Yuan	8948.1	1170.6	5201.79	2719.38	5493.5	7395041	-1.597	-0.136	0.523
EPOP/x_4	10 thousand people	1198.9	456.1	858.835	252.727	905.4	63870.8	-1.249	-0.361	0.294
GCR/x_5	%	49.3	36.3	44.465	4.035	45	16.28	-0.617	-0.58	0.091
COD/y_1	mg/L	19.3	4.9	12.301	4.752	11.6	22.582	-1.313	-0.113	0.386
SO_2 emission	10 thousand tons	28	0.1	11.509	8.064	11.5	65.03	-0.91	0.128	0.701
Dust emission	10 thousand tons	19.399	0.542	7.352	5.15	6.585	26.517	0.415	0.894	0.7

4.2 Comparison of algorithm prediction

According to the three models constructed above, the error sizes under various algorithms are finally measured (Table 2), based on SPSSRPO prediction analysis and combined with the test index system constructed above. As shown in the Table 2 and Fig.1, In all algorithmic sorting, the R^2 sizes in descending order are: BP neural network > ARIMA > GM (1,1), and the MSE sizes are: GM (1,1) > BP neural network > ARIMA. Since the MSE indicator may be affected by the dimension of the forecast indicator, we use R^2 to compare the forecast results, which show BP neural network algorithm has the best prediction effect, followed by time series ARIMA model, and GM (1,1) model. Among the three types of emissions, the prediction of SO_2 emission is the most accurate, followed by dust emission, and finally COD emission.

ISSN 2616-5872 Vol.6, Issue 4: 46-52, DOI: 10.25236/AJEE.2024.060407

Table 2: Detection index

Index name	Testing indicators	GM (1,1)	ARIMA	BPNN	Mean
COD emission	MSE	1440390001	207397.4	159417193.78	533338197.4
	R^2	0.577	0.682	0.904	0.721
SO_2 emission	MSE	139665325	16534698.1	76239880.77	77479968
	R^2	0.989	0.977	0.988	0.984667
dust emission	MSE	244697140	1174.776	33558504.60	92752273
	R^2	0.951	0.963	0.98	0.964667
MSE Mean		608250822	5581090	89738526.38	234523479.5
R^2 Mean		0.839	0.874	0.957	0.89

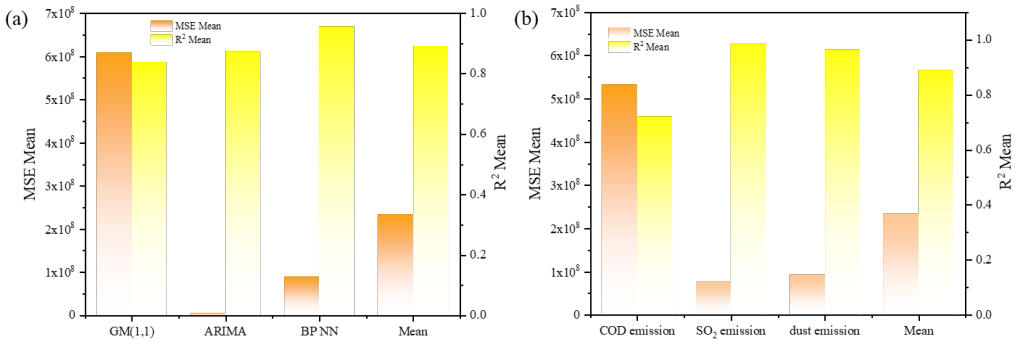

Figure 1: (a) R^2 and MSE of three prediction algorithms, (b) Three emissions of R^2 and MSE

4.3 Prediction model analysis

COD, SO_2 and dust emissions are evaluated by R^2 and MSE to assess the accuracy of the model's prediction. In figure 1, R^2 of BP neural network in the COD emissions prediction model is closer to 1 and MSE is relatively small than that of the ARIMA and GM (1, 1) models, indicating that the BP model has a higher degree of fitting original data and the accuracy of the model is higher. In the same way, the BP model in SO_2 emissions and dust emissions has shown a relatively good forecast. In addition, from Fig.1a, the overall mean of MSE of the BP model is less than ARIMA and GM (1, 1) model, while the R^2 mean value is closer to 1 than the ARIMA and the GM (1, 1) model. The good performance of the BP neural network is due to the ability of independent learning, which has the ability to learn from positive propagation and reverse propagation of error. In addition, compare to the linear prediction of GM (1, 1) model and one variable parameter (like year) of the time series model of ARIMA, the BP model has relatively rich parameters, such as the GDP, TFC, FAI, EPOP and GCR, and nonlinear matching ability, which makes it more accurate to predict the pollutants.

In this paper, the BP neural network model has good prediction performance on the emissions of three kinds of pollutants. As a result, BP neural networks can be used to predict environmental pollution, such as urban sewage emissions, atmospheric pollution emissions, solid waste emissions and so on. In addition, it can also be applied to energy consumption, geological disasters and areas of industry, agriculture, health and health, and provide more favorable help for human life.

5. Conclusion

In this paper, COD, SO_2 and dust emissions were predicted by using the BP neural network, ARIMA and GM (1, 1) model. According to the overall performance, the BP neural network not only has the highest R^2 of a single pollutant discharge, but also shows the good performance of the mean of the three emissions in three models. Although the MSE of BPNN method is slightly higher than that of ARIMA method, the R^2 shown by BPNN method is better than ARIMA, indicating that although BPNN method lacks generalization performance, its robustness is strong, and it can be proved that BPNN is a model suitable for long-term environmental index prediction.

In order to predict environmental pollution, these models can be developed in the future, not only in the resource environment, natural disasters, but also energy development, industrial manufacturing and agricultural production, providing more powerful support for future technology and life.

ISSN 2616-5872 Vol.6, Issue 4: 46-52, DOI: 10.25236/AJEE.2024.060407

References

[1] FAN H, ZHAO C, YANG Y. A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018 [J]. Atmospheric Environment, 2020, 220: 117066.
[2] KUERBAN M, WAILI Y, FAN F, et al. Spatio-temporal patterns of air pollution in China from 2015 to 2018 and implications for health risks [J]. Environmental Pollution, 2020, 258: 113659.
[3] YUAN C, LIU S, FANG Z J E. Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM (1, 1) model [J]. 2016, 100: 384-90.
[4] CHEN J, LIU Z, YIN Z, et al. Predict the effect of meteorological factors on haze using BP neural network [J]. Urban Climate, 2023, 51: 101630.
[5] CEYLAN Z, BULKAN S, ELEVLI S. Prediction of medical waste generation using SVR, GM (1,1) and ARIMA models: a case study for megacity Istanbul [J]. Journal of Environmental Health Science and Engineering, 2020, 18(2): 687-97.
[6] Reddy K , Krishna K .Vehicular pollution prediction using HWTO-ANFIS model in urban areas of Hyderabad City[J].Environmental Progress & Sustainable Energy, 2023, 42.DOI:10.1002/ep.14082.
[7] ZENG B, DUAN H, ZHOU Y. A new multivariable grey prediction model with structure compatibility [J]. Applied Mathematical Modelling, 2019, 75: 385-97.
[8] Zhang X , Zhao M , Dong R .Time-Series Prediction of Environmental Noise for Urban IoT Based on Long Short-Term Memory Recurrent Neural Network[J].Applied Sciences, 2020, 10(3):1144.DOI: 10.3390/ app10031144.
[9] ZHANG G P. Time series forecasting using a hybrid ARIMA and neural network model [J]. Neurocomputing, 2003, 50: 159-75.
[10] Yang Y , Zhang C T , Zhang R .BME prediction of continuous geographical properties using auxiliary variables[J].Stochastic Environmental Research & Risk Assessment, 2016, 30(1):9-26.DOI:10.1007/s00477-014-1005-1.
[11] SUN W, HUANG C. A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network [J]. Journal of Cleaner Production, 2020, 243: 118671.
[12] LU X, ZHANG S, XING J, et al. Progress of Air Pollution Control in China and Its Challenges and Opportunities in the Ecological Civilization Era [J]. Engineering, 2020, 6(12): 1423-31.