pagepress

Geospatial Health 2023; volume 18:1197

The spatial distribution of interleukin-4 (IL-4) reference values in China based on a back propagation (BP) neural network

Zhao Rong Huang,[1] Miao Ge,1 XinRui Pang,[1] Pu Song,[2] Congxia Wang[3]

[1]School of Geographic Sciences and Tourism, Shaanxi Normal University, Xi 'an; [2]Xi'an Shiyou University, Xi'an; [3]Medical College of Xi'an Jiao tong University, Xi 'an, Republic of China

Abstract

This study aimed to investigate the geospatial distribution of normal reference values of Interleukin 4 (IL-4) in healthy Chinese adults and to provide a basis for the development of standard references. IL-4 values of 5,221 healthy adults from 64 cities in China were collected and analyzed for a potential correlation with 24 topographical, climatic and soil factors. Seven of these factors were extracted and used to build a back propagation (BP) neural network model that was used to predict IL-4 reference values in healthy individuals from 2,317 observation sites nationwide. The predicted values were tested for normality and geographic distribution by analytic Kriging interpolation to map the geographic distribution of IL-4 reference values in healthy Chinese subjects. The results showed that IL-4 values generally decreased and then increased from the South to the North. We concluded that the BP neural network model applies to this approach, where certain geographical factors determine levels of various biochemical and immunological standards in healthy adults in regions with different topography, climate and soil indices.

Correspondence: Miao Ge, School of Geographic Sciences and Tourism, Shaanxi Normal University, Xi 'an 710119, Republic of China.
E-mail: gemiao@snnu.edu.cn

Key words: IL-4; geographical factors; correlation analysis; BP neural network; Kriging interpolation; China.

Conflict of interest: the authors declare no potential conflict of interest, and all authors confirm accuracy.

Availability of data and materials: all data generated or analyzed during this study are included in this published article.

Received: 23 March 2023.
Accepted: 27 August 2023.

Geospatial Health 2023; 18:1197
doi:10.4081/gh.2023.1197

Introduction

Interleukin 4 (IL-4) is a multifunctional cytokine that activates, regulates and helps to differentiates the immunological defence system, in particular the T cell and B cells. It plays an essential role by effectively controlling the function of T lymphocytes and the activation of monocytes/macrophages that regulate the inflammatory effect (Wang, Li *et al.*, 2020). It has been shown that an imbalance in the ratio of the two types of T helper (Th) cells (Th1 and Th2) can significantly impact the immune system, with Th2 particularly characterized by secreting interleukin 4 (IL-4) (Huang *et al.*, 2015). Tomar *et al.* (2021) showed that IL-4 is involved in allergic reactions mediated by immunoglobulin E (IgE). It has also been proposed that IL-4 produced by lymphocytes and T cells is a major player in causing inflammation by the activation of Th2 cells and their release of also other cytokines with different functions (Braun, 2020).

Over the past few decades, several scholars have studied the factors that may affect IL-4, such as temperature, hypoxic conditions, altitude and other environmental factors (Facco *et al.*, 2005; Miao *et al.*, 2020). For example, when the body is in a hypoxic state, immune function decreases with a decrease in IL-4 mRNA (Wang *et al.*, 2006) and serum IL-4 levels are significantly higher at high altitudes than at middle altitudes (Hou, 2014). It has also been suggested that humidity affects the expression of immune factors (Zahedi, 2021), while surrounding environmental media, such as soil and air, influence human serum factor concentrations (Li, 2013). In addition, since different soil conditions produce different crops, some scholars have suggested that dietary factors, such as potential environmental factors, can affect mechanisms such as those involved in immunity (Niu *et al.*, 2022).

It follows from the studies mentioned above, that there should be some association between IL-4 levels and geographical factors, including regional differences. The reference values of IL-4 in different regions have been discussed (Guo *et al.*, 2018; Lan *et al.*, 2018; Luo, 2021), but the boundaries of IL-4 values are blurred since China is a vast country with significant geographical differences, which can impact the accuracy of future risk assessment for some diseases (Jin *et al.*, 2020; Monteiro, 2022; te Velde *et al.*, 1990; Wei *et al.*, 2003; Zhou *et al.*, 2014) caused by inflammation. Notably, there are few domestic and international studies on the relationship between IL-4 reference values and geographic environment, and there needs to be more uniformity in measuring reference values. In addition, China spans several temperature zones, and there are significant differences in geographic environments between regions. Therefore, the influence of geographic environment in the clinical diagnosis of reference values should be considered as it would realize the prediction of reference values in different regions. This can be achieved by quantitatively analyzing the relationship between geographic and environmental factors

and reference values that would establish a set of the standard system of IL-4 reference values suitable for all parts of China. The geographic environment contains many variables, such as air, water, organisms and minerals. The distribution pattern and characteristics of the human physique are closely related to the natural social environment (Liu *et al.*, 2019). Therefore, this paper integrates the influence of various geographical factors, such as climate, topography and soil performing spatial autocorrelation analyses of IL-4 data. We aimed to achieve this by constructing and applying a back propagation (BP) neural network model to map the geographical distribution of IL-4 reference values among healthy adults in China. Based on the excavation of the association between IL-4 reference values in healthy adults and the geographical environment in which they are located, regional differences in the distribution of IL-4 reference values were analyzed, thus providing a complete basis for the development of IL-4 reference values.

Material and Methods

IL-4 reference values

In this paper, we used literature from databases such as China Knowledge Network (CNKI); Wanfang Medical Network; Journal of the Chinese Medical Association; China Science Citation Database; China Important Conference Full Text Database; and China Important Conference Full Text Database, as well as actual data of some hospitals to investigate and analyze the IL-4 reference value data. Based on the keywords "interleukin 4" and "IL-4" combined with "health" and "reference value," a large amount of literature was obtained for the period 2010-2022. We based our work on IL-4 data for eligible healthy adults were collected and analyzed in publications including ethical reviews.

Nationwide, 5,221 medical data from 64 counties and municipalities were collected after excluding information with incomplete registrations; 43.6% were males and 56.4% females. We also carried out physical examinations to avoid the presence of systemic inflammatory symptoms and elevated white blood cell counts (WBC). We also excluded various acute diseases of essential organs of the heart, liver, brain, kidneys and lungs, as suggested by Pang *et al.* (2005).

Geographical factors

Three major geographic indicators were selected, namely, terrain, meteorological, and soil indicators, with a total of 24 small geographic factors, as shown in Table 1.

Design

One week before treatment, 5 mL of fasting venous blood was collected from patients, and serum was centrifuged (speed 3,000 r/min, centrifugation radius 10 cm, centrifugation time 10 min), and IL-4 levels were measured by enzyme-linked immunosorbent assay. After the data had been collated, a spatial autocorrelation analysis was performed in ArcGIS, v. 10.8 (ESRI, Redlands, CA, USA) to determine whether there was a specific association between IL-4 levels and geographic factors. Afterwards, the Chinese healthy adult IL-4 reference values were matched with 24 geographic indicators one by one, and indicators related to healthy adult IL-4 values were screened. Data were analyzed for normality in SPSS, v. 25.0 and Spearman method was selected for correlation analysis. To further test the effect of covariance between factors, severe covariance was eliminated by multicollinearity diagnosis.

Secondly, a BP neural network program was produced to establish a nonlinear mapping relationship between geographic factor

Table 1. Factors of potential influence

Indicator	Geographical factor	Type of indicator	Source of indicator
Topographic index	X1	Longitude	China State Bureau of Surveying and Mapping*
	X2	Latitude	
	X3	Elevation	
Climate index	X4	Annual sunshine duration	China Meteorological Data Science Data Sharing Network**
	X5	Annual mean air temperature	
	X6	Annual mean relative humidity	
	X7	Annual precipitation	
	X8	Annual range of air temperature	
	X9	Annual mean wind speed	
Soil index	X10	Percentage of topsoil gravel	World and Spectrum Soil Database***
	X11	Percentage of topsoil silt particles	
	X12	Percentage of topsoil clay particles	
	X13	Topsoil reference capacity	
	X14	Topsoil gravel content	
	X15	Topsoil organic carbon	
	X16	Topsoil pH	
	X17	Topsoil cation exchange capacity-clay	
	X18	Topsoil cation exchange capacity-silt	
	X19	Topsoil base saturation	
	X20	Topsoil total exchange capacity	
	X21	T_CaCO4	
	X22	T_CaSO4	
	X23	Topsoil alkalinity	
	X24	Topsoil salinity	

*http://cdc.cma.gov.cn/;**http://data.cma.cn/);***http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/index.html?sb=1)

data and IL-4 reference values. Next, IL-4 reference values for healthy individuals were predicted nationally. Finally, geostatistical analysis was performed in ArcGIS selecting the analytical Kriging interpolation method to calculate the optimal difference points with the help of (semi-)variational function theory and cross-validation analysis so that a map of the geographical distribution of IL-4 reference values of healthy adults in China could be generated.

Statistical analysis

Autocorrelation

Spatial autocorrelation was measured by Moran's *I*, using both global and local spatial autocorrelation (Chen, 2009; Jiang *et al.*, 2011). Among them, global spatial autocorrelation is a spatial characterization of geographic element attribute values across regions, and estimates by Global Moran's *I* are usually used to analyze the average degree of overall regional spatial association and spatial variation (Nie *et al.*, 2013). $I>0$ means that data under consideration are positively correlated, $I<0$ that they are negatively correlated and $I=0$ that the distribution is completely random (Yang *et al.*, 2020). This analysis was used to obtain *Z* and *P* values to evaluate the significance of the index (Ge *et al.*, 2005) and refers to the potential interdependence between observations of some variables in the same region reflecting the self-correlation of the values of a spatial variable (Feng *et al.*, 2022).

Correlation analysis

Correlation analysis was done to quantify the degree of correlation between variables by introducing specific statistical indicators (Liang *et al.*, 2016). Correlation analysis either uses Pearson's coefficient, Spearman's rank coefficient or Kendall's rank coefficient to determine the correlation between indicators (Valero *et al.*, 2019). The Kolmogorov-Smirnov test (K-S) test can be used to test whether a sample obeys a specified distribution by analyzing the difference between two distributions (Yang *et al.*, 2019). Before conducting the correlation analysis, the K-S test was utilized to conclude that the sample data were non-normally distributed (Zhu *et al.*, 2009), and therefore Spearman correlation analysis was chosen. If data are close to a normal distribution under the K-S test, then the *p*-value is more significant than 0.05; conversely, if less than 0.05, the opposite. The IL-4 reference values of healthy Chinese adults were matched with a database of 24 geographic indicators in SPSS 25.0 software to filter out geographic factors associated with IL-4 reference values of healthy adults.

Multicollinearity diagnosis

In regression analysis high levels correlation between variables may lead to a decrease in model accuracy (Wei, 2019). The variance inflation factor(VIF) measures whether there is a correlation between geographic indicators and is therefore used to determine the presence of multicollinearity, which is defined as:

$$VIF_j = \frac{1}{(1-R_j^2)} \quad \text{(Eq. 1)}$$

where *j* represents the individual variables. VIF_j values between 0 and 5 indicate a slight degree of covariance between the independent variables; values between 5 and 10 that there is moderate degree of covariance but values exceeding 10 indicates severe multicollinearity making it necessary to remove certain variables to improve the accuracy (Li, Zhao, *et al.*, 2006 ; Liu, 2013). However, if the correlations among the various variables are the same, a slight multicollinearity problem will not significantly impact the final model results.

BP neural network model

This is a multilayer forward neural network that uses the error gradient descent method for learning to minimize the total or average error of the network's computational output (Yang *et al.*, 2017). It is characterized by forward propagation of the signal, followed by backward propagation of the error (Li *et al.*, 2021) and is currently the most widely used neural network. It has good self-learning, adaptive and generalization capabilities; in theory, the three-layer BP neural network can approximate any nonlinear function with arbitrary accuracy (Wang *et al.*, 2018). When this algorithm is used, the disease prediction model's accuracy improves substantially (Liu *et al.*, 2021).

We applied Matlab and its language 2018 component (https://www.mathworks.com/products/matlab/programming-with-matlab.html) to program the available geographic data for simulating the IL-4 reference values for other counties and cities in China. The BP neural network used consisted of input, implicit and output layers. Seven geographic factors diagnosed by multicollinearity (see below) were used as the input nodes for the input layer. The measured IL-4 values of healthy people were used as the output layer nodes as shown in the schematic diagram in Figure 1. Then it was trained based on the inputs and outputs to obtain the network weight parameters with the slightest error.

Network training is the step in which weights and thresholds are continuously adjusted to reduce the network error to a predetermined minimum or stop at a predetermined error. The prediction samples are then fed into the trained network to obtain the prediction results. In this paper, we used the advantages of the BP neural network algorithm and combined the medical indicators and geographic factors for modelling. The work flow chart is shown in Figure 2.

Each input parameter was normalized before performing the network training step. The original data were then processed into the interval between [-1 and 1]. Next, model training and network simulation were performed, and the output simulation results back-normalized: *i.e.* the normalized data are mapped back to the origi-

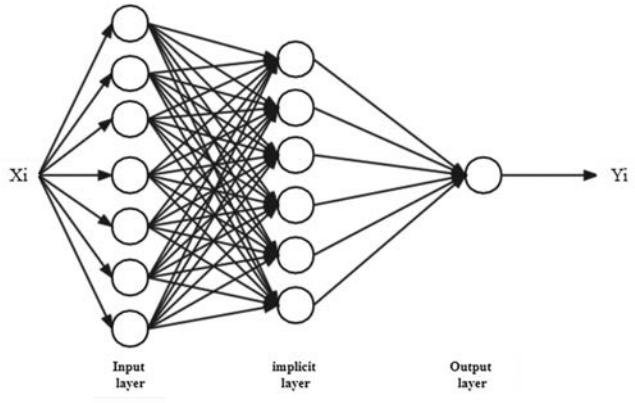

Figure 1. Schematic diagram of the network structure.

nal data, thus obtaining data that can be used for geostatistical analysis. There is no precise method to determine the number of neuron nodes in the hidden layer, so we used 'trial-and-error' to determine the optimal number of neurons in the hidden layer. The empirical formula is as follows:

$$L = \sqrt{n+m} + a \quad \text{(Eq. 2)}$$

where L is the number of nodes in the hidden layer; n the number of nodes in the input layer; m the number of nodes in the output layer; and a a regulatory constant with a value between 1 and 10.

A three-layer BP neural network model with a structure of 7-L-1 form means that there are 7 nodes in the input layer, L nodes in the hidden layer, and 1 node in the output layer. The number of nodes in the hidden layer is determined using an empirical formula. The optimal number of neurons in the hidden layer is determined using 'trial-and-error'; normally 3-12 hidden layer nodes are taken for network training. Figure 3 shows a partial network training error plot for the number of neurons in the hidden layer from 5 to 10. The number of neurons here is determined together with the network training mean square error (MSE) (Table 2). Following training, the neural network automatically generated the model construction between the input, implicit and output layers. After several training sessions, it can be concluded that when the number of neurons in the hidden layer is 6, the BP neural network approximates the function well, and the combined network training speed and performance are optimal.

The activation function of the implicit layer of the BP neural network uses the logsig function, the transfer function of the input layer and the implicit layer the tansig function, the output layer the purelin function, and the training function trainlm. The maximum number of training sessions was set at 2000, the expectation error at 0.1 and the error between the predicted and actual values of the network arrived at by the MSE that was chosen as an evaluation index to visualize the effectiveness of the modelling method. It can be calculated as follows:

$$MSE = \frac{1}{2}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \quad \text{(Eq. 3)}$$

where y_i is the network output value; andis actual value.

Statistics

Compared with traditional statistics, geostatistics considers both the size of the sample value, its spatial location and the distance between the samples, which corrects traditional statistics' deficiency of omitting the spatial orientation. The spatial interpolation algorithm can transform the input point data into continuous surface data, thus reflecting the spatial distribution pattern of the selected index data (Li *et al.*, 2013; Liu *et al.*, 2016; Zhu *et al.*, 2004). Of the different Kriging variations (Oliver, 1990; Yue *et al.*, 2005), the disjunctive Kriging method does not require too much distribution of the original data and can directly observe the spatial variation pattern of the factors (Zhang *et al.*, 2021). Therefore, this method was chosen for interpolation in this paper.

In geostatistical interpolation, the choice of the variance function is crucial. The basic properties of the unknown point can be inferred from the variation function. Theoretically, when the observed point distance is 0, its variation function should also be 0; however, due to the existence of measurement error and spatial

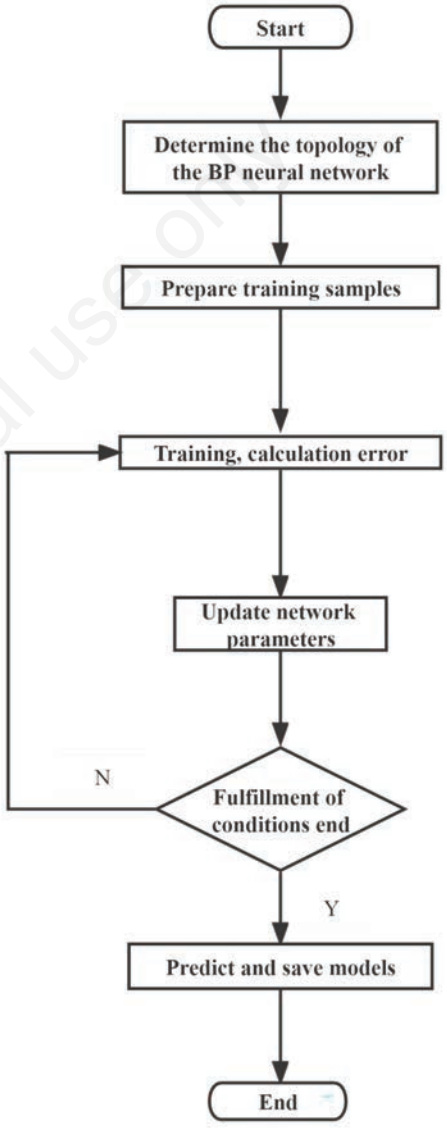

Figure 2. BP neural network algorithm flow chart.

Table 2. Partial network training mean square error.

Number of implicit layer neurons	5	6	7	8	9	10
Network training mean square error	0.0204	0.0191	0.0315	0.0192	0.0222	0.0254

variation, it will cause the values of the various functions of two spatially close points to be non-zero, a phenomenon called the nugget effect (Liu *et al.*, 2011; Zhang *et al.*, 2016). Generally, low nugget values indicate better method applicability.

Results

Spatial autocorrelation

The IL-4 reference value is spatially clustered (Table 3).The significance test performed with z-score was obtained at 6.5702 (p=0.05), which indicated a significant spatial autocorrelation of the reference values of IL-4. If the z-value is more significant than

Table 3. Summary of spatial autocorrelation results.

Moran's *I*	1.887492
Expected value	-0.022222
Variance	0.084485
z-score	6.5702

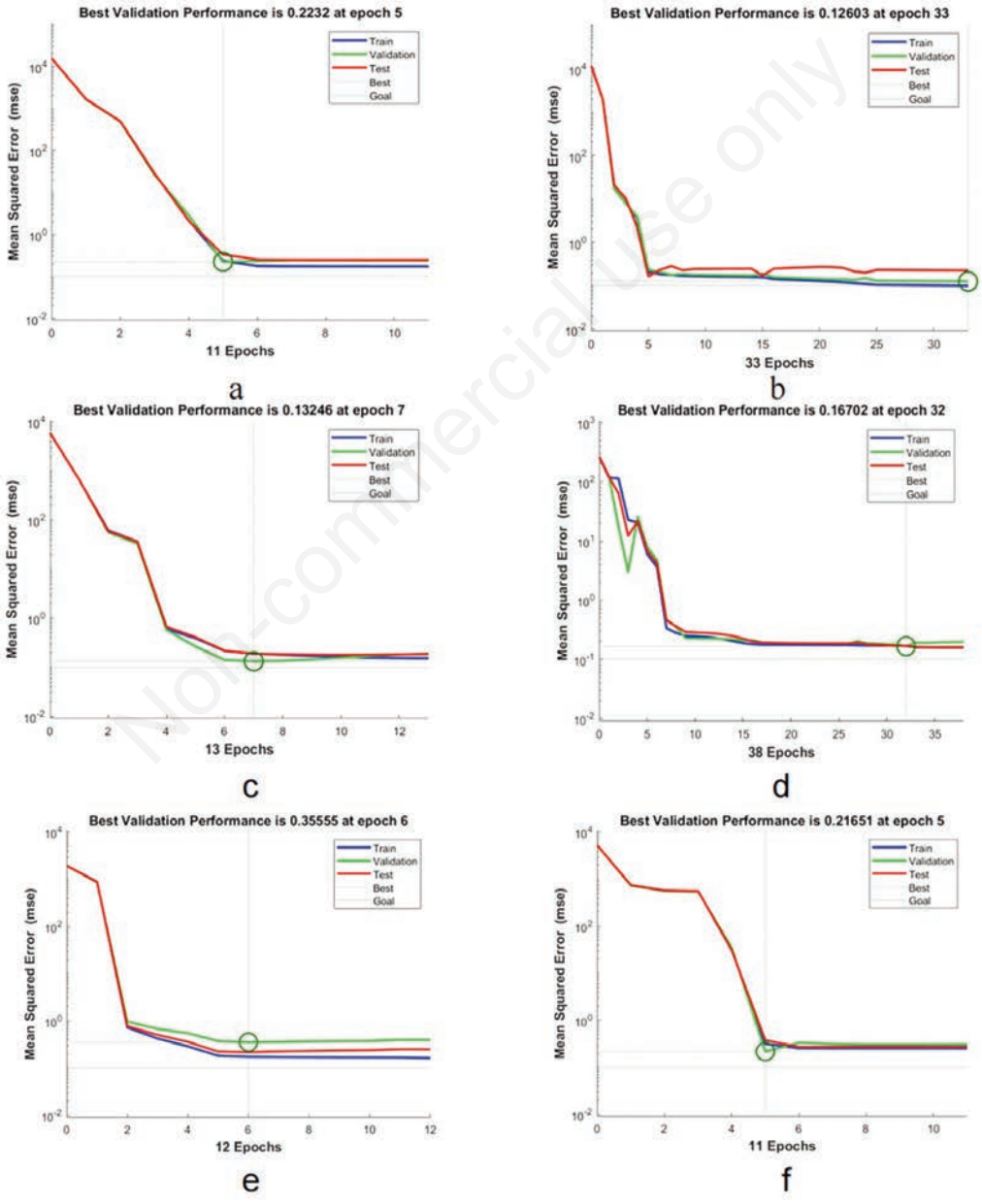

Figure 3. Network training errors when the number of neurons in the hidden layer is 5~10.

1.96 or less than -1.96, the Moran index is significant, *i.e.*, spatial autocorrelation exists.If the z-value is between -1.96 and 1.96, the Moran's index is insignificant, *i.e.* spatial autocorrelation does not exist. Meanwhile, the analysis in Figure 4 yielded that the spatial distribution of IL-4 showed an aggregation pattern.

Correlation analysis

The K-S test results are shown in Table 4. Since it was clear that IL-4 medical reference values and geographical indicators did not satisfy the normal distribution, the Spearman approach was used for the subsequent correlation analysis. Among the 24 geographic factors, nine were associated with IL-4 reference values for healthy adults: latitude (X_2), elevation (X_3), annual sunshine hours (X_4), annual mean temperature (X_5), mean relative humidity (X_6), annual precipitation (X_7), annual difference in temperature (X_8), percentage of topsoil gravel (X_{10}), and Percentage of topsoil clay particles(X_{12}). The correlation) and significance coefficients are shown in Table 5.

Table 4. Kolmogorov-Smirnov test results.

Geographical indicator	Type of indicator	Statistics	Conspicuousness
X_2	Latitude	0.097	0
X_3	Elevation	0.296	0
X_4	Annual sunshine duration	0.122	0
X_5	Annual mean air temperature	0.08	0.003
X_6	Annual mean relative humidity	0.217	0
X_7	Annual precipitation	0.154	0
X_8	Annual range of air temperature	0.087	0.001
X_{10}	Percentage of topsoil gravel	0.276	0
X_{12}	Percentage of topsoil clay particles	0.205	0

Table 5. Correlation between reference values and geographical factors.

Geographical indicator	Type of indicator	p	Correlation coffecient
X_1	Longitude	0.145	0.073
X_2	Latitude	-0.223**	0.006
X_3	Elevation	-0.202*	0.012
X_4	Annual sunshine duration	-0.271**	0.001
X_5	Annual mean air temperature	0.285**	0
X_6	Annual mean relative humidity	0.289**	0
X_7	Annual precipitation	0.268**	0.001
X_8	Annual range of air temperature	-0.173*	0.032
X_9	Annual mean wind speed	-0.001	0.988
X_{10}	Percentage of topsoil gravel	-0.229**	0.004
X_{11}	Percentage of topsoil silt particles	0.220**	0.006
X_{12}	Percentage of topsoil clay particles	-0.036	0.654
X_{13}	Topsoil reference capacity	-0.038	0.642
X_{14}	Topsoil gravel content	0.100	0.221
X_{15}	Topsoil organic carbon	0.102	0.211
X_{16}	Topsoil pH	0.008	0.921
X_{17}	Topsoil cation exchange capacity-clay	-0.074	0.364
X_{18}	Topsoil cation exchange capacity-silt	-0.081	0.319
X_{19}	Topsoil base saturation	0.092	0.258
X_{20}	Topsoil total exchange capacity	0.146	0.072
X_{21}	T_CaCO4	0.127	0.117
X_{22}	T_CaSO4	-0.056	0.495
X_{23}	Topsoil alkalinity	-0.131	0.106
X_{24}	Topsoil salinity	0.047	0.563

*significant correlation, within a 95% confidence interval; **Significant correlation within a 99% confidence interval.

Multi-collinearity diagnosis

The data of the IL-4 samples from healthy individuals were used as the dependent variable and the geographical indicators associated with them were used as independent variables to diagnose multicollinearity. Table 6 shows that the VIF values of latitude (X_2) and annual mean air temperature (X_5) were greater than 10, indicating their collinearity problem, which needed to be removed for regression statistics to ensure the accuracy of subsequent modelling.

Use of the BP neural network model

The final fit shown by the BP neural network can be seen in Figure 5. The horizontal coordinates of the graph represent the number of random number data points and the vertical ones the range of IL-4 reference values. The closer the two fold lines are, the better the fit. The training and predicted data fitted well with no special outliers. In neural networks, R stands for regression, which represents how the output of the target in a particular dataset compares to the training target, the calibration target. The R-values of the established IL-4 prediction model fitted to the training dataset, validation dataset, test dataset and the overall dataset were 0.93, 0.91, 0.98, and 0.94, respectively, which indicated that the predicted data were highly correlated with the target data and the network simulation was reasonable (Figure 6).

Geostatistical analysis

The data were explored for normality using histograms. When the predicted IL-4 values of the 2, 317 counties fed into the China vector map, the results showed a right-skewed state (1.2237 > 0), with the mean IL-4 prediction 39.074 ng/mL and the median 37.425 ng/ml. These results indicated that the predicted IL-4 values were not have a normal distribution, so the analytic Kriging method was used for interpolation when the graph was produced.

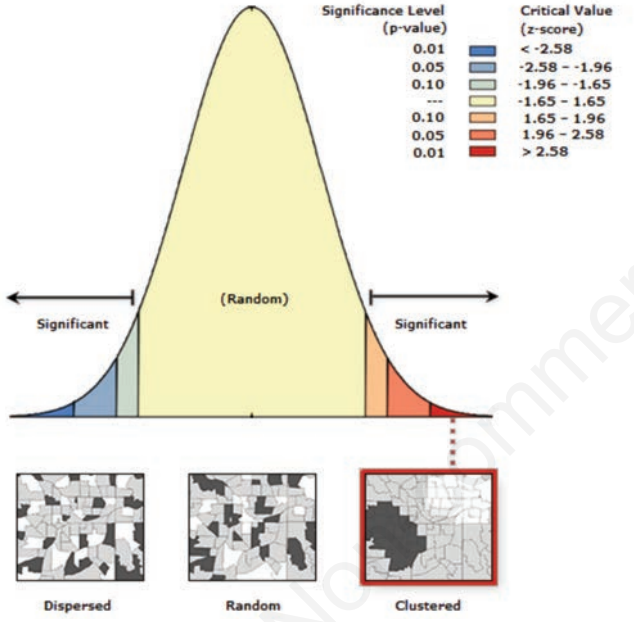

Figure 4. Spatial autocorrelation results.

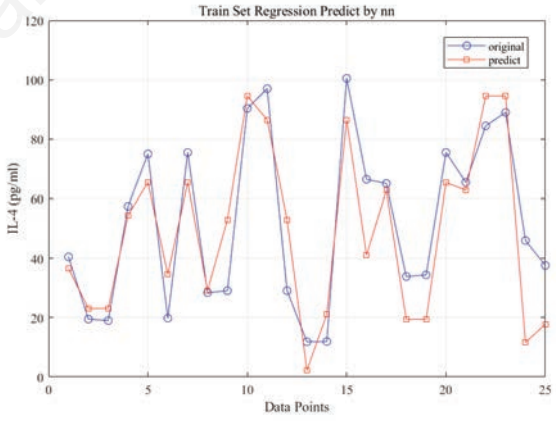

Figure 5. Fitting of training and predicted data.

Table 6. Multi-collinearity diagnosis results.

Geographical indicator	Type of indicator	Tolerance	VIF
X_2	Latitude	0.025	40.215
X_3	Elevation	0.227	4.409
X_4	Annual sunshine duration	0.267	3.742
X_5	Annual mean air temperature	0.048	20.683
X_6	Annual mean relative humidity	0.157	6.352
X_7	Annual precipitation	0.139	7.198
X_8	Annual range of air temperature	0.115	8.674
X_{10}	Percentage of topsoil gravel	0.454	2.203
X_{12}	Percentage of topsoil clay particles	0.465	2.149

VIF, variance inflation factor.

Before interpolation, an exploratory analysis of the interpolated data was performed using the geostatistical tool in ArcGIS to obtain a trend map of the spatial characteristics of the IL-4 reference values (Figure 7). As seen in the figure, the trend surface of the predicted reference values of IL-4 in healthy adults showed an overall second-order polynomial character. As can be seen in the figure, the graph shows an increase and a slower growth rate in the longitudinal direction from west to east, with little trend in the latitudinal direction, while it shows a U-shaped trend from south to north. Through several trials, the exponential function was finally chosen for interpolation. The choice of the variance function is also critical when performing the model construction. The values of the nugget obtained during the construction process are the semi-variance function (0.5585158) and covariance function (0.0711008). Since the nugget value of the semi-variance function was smaller and closer to 0, reflecting the strong spatial correlation of the interpolated data, the final semi-variance function was chosen to interpolate the IL-4 reference values in the map. The interpolation results are shown in Figure 8, which shows that the overall trend of IL-4 reference values in healthy individuals in China is high in the South and low in the North, consistent with trend surface analysis results. IL-4 reference values varied rapidly in the south-eastern coastal region and less in the rest of the country; they were high and varied rapidly on the Yunnan-Guizhou Plateau and in the southern part of the south-eastern hills. They were also generally lower on the Qinghai-Tibet Plateau, onthe Inner Mongolia Plateau, and in the northern part of the Loess Plateau.

Discussion

For terrain indicators, IL-4 reference values showed significant correlations with latitude and elevation, with the largest correlation coefficient for elevation. Altitude medicine shows that the human body's immune, circulatory and psychological systems differ depending on whether people live on mountainous plateaus or in the plains, *e.g.,* changes can include capillary wall damaged by hypoxia, platelet aggregation leading to microcirculatory disorders and disrupted CD4+/CD8+ balance leasing to Th1/Th2 imbalance (Wang, Chen, *et al.*, 2020). As altitude increases, the body requires more red blood cells to co-transport oxygen, the blood is sticky, blood flow slows down, and the number of immune cells cleared by the spleen per unit of time increases, leading to a decrease in the number of immune factors. It has also been suggested that the

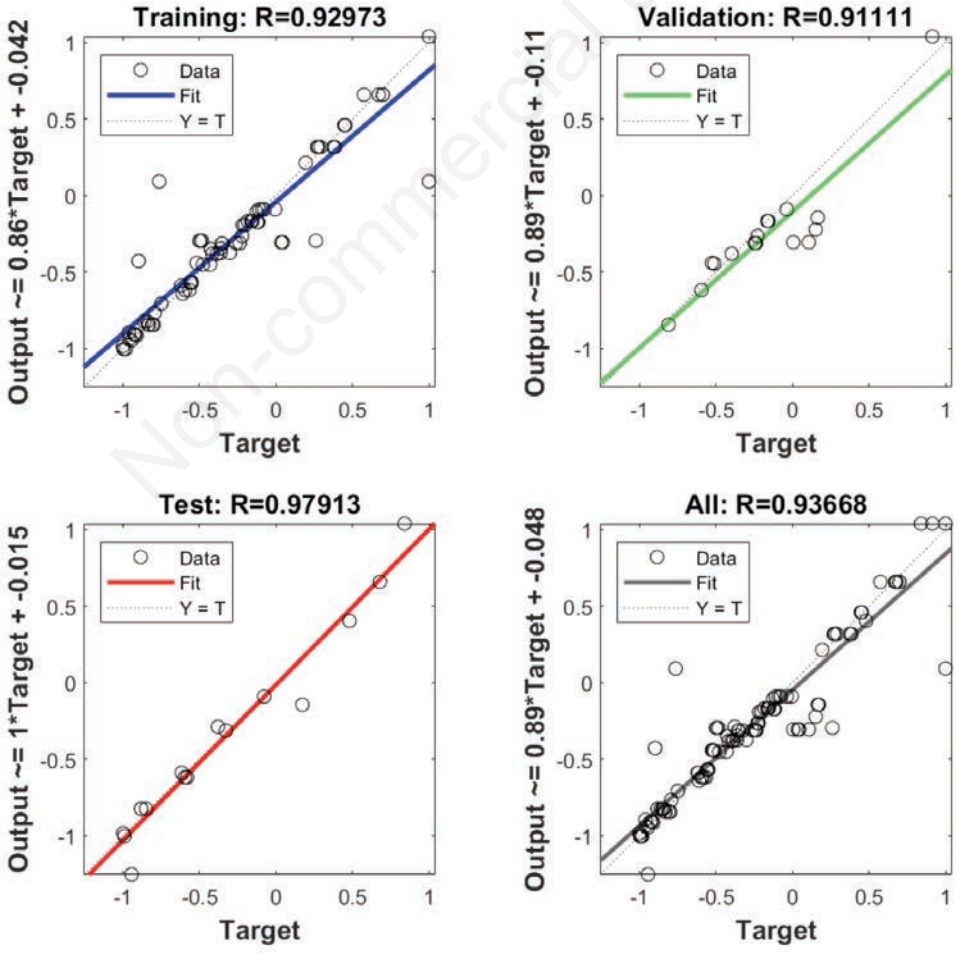

Figure 6. Fitted regression scatter plot of the IL-4 prediction model.

apoptosis of CD4+ T-lymphocytes increases under hypoxia, and that their secretion of Th2 cells decreases resulting in decreased IL-4 levels (Gu, 2016). In mountainous areas above 3000 m above the mean sea level, the differentiation and function of CD4+ T cells are suppressed, and the body's overall immunity reduced (Ma *et al.*, 2016), thus affecting the IL-4 level.

China is a vast country with a wide latitude and longitude span and latitude directly affects the distribution of temperature zones and climate types in the North and South (Nguyen, 2016). The country also has a complex and diverse climate, with significant temperature differences between far away areas. If the temperature suddenly drops, the body changes its immune status to adapt to the environmental changes, resulting in a decrease of suppressive CD8+ T cells and an increase of helper CD4+ T cells (Min, 2007). The proliferation and differentiation of immune cells are inhibited and factors such as IL-4 are subsequently reduced. Therefore, in northern areas that have sizeable annual differences in temperature compared to southern areas, with higher average annual temperatures, IL-4 levels are lower, a fact seen in the spatial pattern presented by the geospatial distribution map.

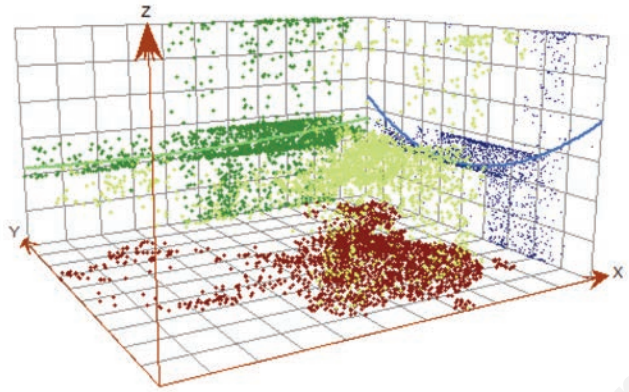

Figure 7. Trend of distribution of the predicted data. In this figure, the X, Y, and Z axes are directed East (longitude) and North (latitude), respectively. The direction of the Z axis is that of increasing IL-4 reference values.

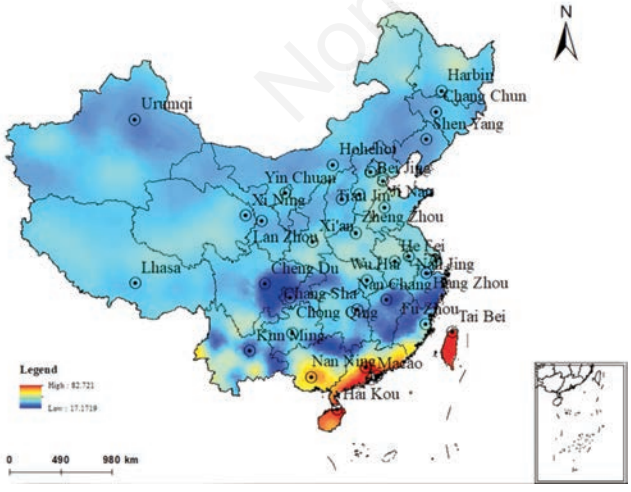

Figure 8. Predicted trend of the geographic distribution of IL-4 reference values in healthy individuals in China. A stronger hue represents a closer range of IL-4 reference values between regions, with shade differences representing more highly significant variations of the reference values between regions.

Regarding meteorological factors, there was also a significant correlation between IL-4 levels and the mean annual temperature, and the mean annual relative humidity in the healthy population. In recent years, many climate model studies have shown that the thermoregulatory systems of children, the elderly, and the frail are in "overload" at high temperatures (Oliver, 1990). Indeed, meteorological factors such as temperature, humidity, and other factors are closely related to human health (Ding, 2010) and it has been found in animal experiments that exposure of the airways to cold air stimulates cytokine expression in general but inhibits cell-mediated immune responses as well as IL-4 (Davis, 2007). In addition, Hong *et al.* (2022) concluded that the CD4+/CD8+ ratio in the sera of rats in high humidity environment under normothermic conditions is lower than in rats under normal humidity conditions; in fact, CD4+ T cells was found to promote the proliferation and differentiation of immune cells, while CD8+ T cells suppressed the immune state of the body. When CD4+ is reduced, and CD8+ is increased, the body's immune function becomes imbalanced and produces more IL-4 to suppress pro-inflammatory factors and cellular responses to act as an anti-inflammatory agent (Hong *et al.*, 2022). The finding that the IL-4 level is higher in a high-humidity environment agrees with Meng (2015) who report an increased rise in T helper lymphocytes in a humid and hot environment. Therefore, areas with higher average annual temperatures and relative humidity have higher IL-4 reference values.

Most of the relationships between soil factors, such as percentage of topsoil gravel and percentage of topsoil clay particles, are indirect and do not show a direct relationship with cytokine production and the immune system. However, soil types in China have regional patterns (Ma, 1992) and soil fertility levels have an impact as crops grow differently in different soil textures leading to variations in people's diets in different regions. Unlike people living in arid desert areas, those living in plain areas mostly grow crops such as rice and wheat as staple food. Carbohydrates are converted into sugar in the body and increased glucose levels in the blood can promote non-enzymatic glycosylation reactions of proteins and lipids. This leads to increased production of late end products that activate stress kinases and stimulate the production of cytokines such as IL-4 (Donath, 2013). The inhabitants of arid desert areas are affected by the soil conditions, and their diet mostly lacks nutrients such as dietary fibre. Of particular importance is that dietary fibre can promote intestinal metabolic activity and regulate the development and function of regulatory T cells by altering the composition of the intestinal flora. Therefore, the inhabitants of arid desert areas have a homogeneous diet and low IL-4 values compared to those living in fertile areas.

A limitation of this study was that the exploration of reference values of IL-4 in this study did not include human environment factors such as diet and nutrition, which need to be further explored. Due to the complexity of human and geographical environments, the influence mechanism between the two also needs further research and analysis.

Conclusions

The distribution pattern and characteristics of IL-4 reference values in China were found to be generally high in the South and low in the North, high in the East and low in the West. A significant spatial autocorrelation of IL-4 reference values between regions was also found together with nine geographic factors affecting the results of IL-4 prediction values: latitude, elevation, annual sunshine hours, annual mean temperature, mean relative humidity, annual precipitation, annual difference in temperature, percentage of topsoil gravel, and the percentage of topsoil clay particles.

References

Braun C, Vocanson M, Nicolas JF, Nosbaum A, 2020. Pathophysiology of atopic dermatitis and other atopic diseases: is a global approach possible? Ann Dermatol Venereol 147:11S4-11S11

Chen YG, 2009. Development of spatial autocorrelation theory and methodological improvements based on Moran statistics. Geogr Stud 28:1449-63.

Davis MS, Williams CC, Meinkoth JH, Malayer JR, Royer CM, Williamson KK, McKenzie, EC, 2007. Influx of neutrophils and persistence of cytokine expression in airways of horses after performing exercise while breathing cold air. Am J Vet Res 68:185-9.

Ding XS, 2010. Spatial and temporal distribution characteristics of high temperature heat waves in Jiangsu Province and their impact on human health, Vol. Master. Nanjing University of Information Engineering. Nanjing Univ Inf Eng. pp. 72.

Donath MY, Dalmas É, Sauter NS, Böni-Schnetzler M, 2013. Inflammation in obesity and diabetes: islet dysfunction and therapeutic opportunity. Cell Metab 17:860-72.

Facco M, Zilli C, Siviero M, Ermolao A, Travain, G, Baesso I, Bonamico S, Cabrelle A, Zaccaria M, Agostini C, 2005. Modulation of immune response by the acute and chronic exposure to high altitude. Med Sci Sports Exerc 37:768-74.

Feng J, Chen TH, Dai LF, Tang N, Du W, Ye XG, Lei SG, 2022. Bivariate spatial autocorrelation analysis of influenza vaccination rate and socioeconomic indicators among children aged 6 months-5 years in Guizhou Province during the influenza season of 2020-2021. Vac Im Chin 28:199-203.

Ge Y, Yao S, Pu YX, Jia L,2005. A bivariate spatial autocorrelation analysis of influenza vaccination rates and socioeconomic indicators for children aged 5 months in China. Geographical patterns of agglomeration economic types using spatial autocorrelation analysis. Hum Geogr 03:21-25. [in Chinese]

Gu SQ, 2016. Alterations in lymphocyte subpopulations, platelets and coagulation function in healthy individuals at different altitudes. Chin J Blood Trans 29:700-703.

Guo HY, Wang DX, 2018. Comparison of serum IL-2, IL-4, TNF-α, CRP dynamic levels and recovery time of gastrointestinal function in elderly patients with acute cholecystitis under different surgical approaches. Mark Im Clin 25:94-98.

Hong MY, Yu HW, Zhu Y, JL, Z C, Z YL, Wang WY, 2022. Effects of acupuncture on the ultrastructure of synovial cells and IL-1β, IL-4 and IL-10 in rats with rheumatoid arthritis model. J Gansu Univ Trad Chin Med 39:9-12.

Hou B, 2014. Clinical study of the relationship between serum IL-2 and IL-4 and chronic obstructive pulmonary disease at different altitudes. Qinghai Univ. [in Chinese] (That's all the information I can find on this article, it's a dissertation)

Huang WQ, Li H, Yuan M, Chen SD, 2015. Changes in the levels of peripheral blood T-cell subsets Th1, Th2 and their related cytokines in patients with bronchial asthma and their significance. Shandong Med 55:40-41.

Ji J, Zhang RP, Yan J, Song T, Xu P, 2020. Effect of tacrolimus combined with compound glycopyrrolate tablets on serum Th1/Th2 type cytokine levels in patients with stable vitiligo. Chin Aes Med 29: 48-51. [in Chinese]

Jiang QW, Zhao F, 2011. Application of spatial autocorrelation analysis method in epidemiology. Chin J Epidemiol 32:539-546.

Lan X, Li ZM, Zhang WD, 2018. Determination of the reference range of serum inflammatory cytokines in a normal population in Zhengzhou. J Zhengzhou Univ 53:341-343.

Li JX, Li ZQ, Yin W, 2013. Kriging interpolation method based on ArcGIS and its application. Mapping Bul 09: 87-90. [in Chinese]

Li L, Xu Y, 2021. Building energy consumption prediction based on BP neural network. Com Prog Skills Main 04 :10-12. [in Chinese]

Li L, Zhao W, Wang ZF, 2006. Superiority of ridge regression method over LS method in solving multicollinearity problems. J Bohai Univ (Natural Science Edition) 02: 124-126. [in Chinese]

Li SQ, 2013. Effects of lead and cadmium pollution and p53 gene polymorphisms on the serum levels of related factors in the Sha Ying River area. J Zhengzhou Univ 48:788-791.

Liang JY, Feng CJ, Song P, 2016. A review of big data correlation analysis. J Com Science 39:1-18.

Liu GM, Wang YA, Zhang HR, Wang D, 2011. Several interpolations in spatial analysis. A comparative study of several interpolation methods in spatial analysis. Geog Inf World 9:41-45.

Liu J, Jin YY, Sun BQ, Yang J, 2019. Distribution pattern of TCM body quality in 1336 residents in Kunshan area. Shanxi Trad Chin Med 35:50-52.

Liu M, 2013. Resolution of multicollinearity: a new criterion for eliminating variables. Stat Dec Making 05: 82-83. [in Chinese]

Liu T, Yao ML, Huang JG, Huang SQ, Chen HY, Yang WW, Cai J, Wu R, 2021. Comparison of the prediction effectiveness of BPNN neural network model and SARIMA model in the incidence number of class B infectious diseases in Jingzhou City. Chin J Soc Med 38:109-113.

Liu YY, Li JC, 2016. Comparison of spatial interpolation methods for atmospheric PM_(2.5) concentration distribution in Changsha City. Env Mon Man Tech 28:14-18.

Luo W, ZJ, WZ, 2021. Circulating levels of IL-2, IL-4, TNF-α, IFN-γ, and C-reactive protein are not associated with severity of COVID-19 symptoms. J Med Virol 93:89-91.

Ma JH, 1992. Re-understanding of soil distribution laws. J Henan Univ (Natural Science Edition) 1: 1-5. [in Chinese]

Ma YP, Huang JF, Li CX, Qiu EC, Yu HB, Qi XL, Li HL, 2016. Analysis of injury and disease occurrence during plateau training in a department in Xinjiang. J Pla Prev Med 34:401-403.

Meng Y, Weng XQ, Yuan Y, 2015. Effects of acute exercise of different intensities on body T lymphocytes and their subpopulations in humid and hot environment, 2015 10th Nat Sports Science Conf, Hangzhou, Zhejiang, China. pp. 2.

Miao ZL, Zhong XM, Cui R, Li PP, Wang YX, Zhong WY, Zhao

WZ, 2020. Effect of specimen storage time and temperature on Th1/Th2 intracellular cytokine assay by stimulation method. Guangdong Med 41:2522-2525. [in Chinese]
Min JW, Park S-M, Rhim TY, Park S-W, Jang A-S, Uh S-T, Park C-S, Chung IY, 2007. Effect and mechanism of lipopolyeaccharide on allergen-induced interleukin-5 and eotaxins produciton by whole blood cultures of atopic asthmaitcs. Clin Exp Immunol 147:440-8.
Monteiro MF, Casati MZ, Sallum EA, Silvério KG, Nociti-Jr FH, Casarin RCV, 2022. The familial trend of the local inflammatory response in periodontal disease. Oral Dis 28:202-209.
Nguyen JL, Yang W, Ito K, Matte TD, Shaman J, Kinney PL, 2016. Seasonal Influenza Infections and Cardiovascular Disease Mortality. JAMA Cardiol 1:274-281.
Nie Y, Luo Y, Yu J, Chen F, 2013. Spatial autocorrelation-based spatial and temporal evolution characteristics of arable land pressure in Hubei Province. Geogr Res Dev 32:112-116.
Niu HQ, Xu MH, Wang RB, Li XF, 2022. Progress on the regulation of intestinal flora and immune function by diet and its role in rheumatoid arthritis. Chin J Rheumatol 26:262-266.
Oliver MA, Webster R, 1990. Kriging: a method of interpolation for geographical information systems. Intl J Geogr Inf Syst 4:313-332.
Pang YH, Zheng CQ, Wang YC, Li XD, Li F, 2005. Expression of IL-4 and IL-13 in ulcerative colitis. J Gastroenterol Hepatol 143:410-412.
te Velde AA, Huijbens RJ, Heije K, de Vries JE, Figdor CG, 1990. Interleukin-4 (IL-4) inhibits secretion of IL-1 beta, tumor necrosis factor alpha, and IL-6 by human monocytes. Blood 76:1392-7.
Tomar S, Ganesan V, Sharma A, Zeng C, Waggoner L, Smith A, Chang H Kim CH et al., 2021. lL-4-BATF signaling directly modulates Il-9 producing mucosal mast cell (MMC9) function in experimental food allergy. J Allergy Clin Immunol 147:280-95.
Valero M, Li F, Clemente J, Song W, 2019. Distributed and Communication-Efficient Spatial Auto-Correlation Subsurface Imaging in Sensor Networks. Sensors (Basel) 19:2427.
Wang CC, Wu JH, ZhouY, Peng YB, Cao YZ, Song Y, Zhang XY, Yuan X, Wang Q, Wang GL, 2018. A study on the risk prediction of spontaneous hemorrhagic transformation in patients with acute cerebral infarction based on BP neural network. Chin Fam Med 21:1413-1418.
Wang G, Li J, Huang S, 2020. Immunofluorescence localization of CD14, IL-4 and TNF-α in tissues of healthy subjects and patients with chronic periapical disease. Chin J Pathophysiol 36: 539-546. [in Chinese]
Wang HL, Chen WJ, He J, Li CL, Liu JW, 2020. Expression and significance of T-cell subsets in patients with common psoriasis in a plateau environment in Xinjiang. Xinjiang Med 50:1256-1259.
Wang T, Chen PJ, Gao BH, 2006. Changes in peripheral blood leukocyte gamma-interferon, interleukin-4 and perforin gene expression before and after hypoxic training in female rowers. Chin J Sports Med 25: 196-199. [in Chinese]
Wei HY, 2019. Diagnosis and treatment of multicollinearity in regression analysis. J Zhoukou Norm Coll 36:11-15.
Wei J, Shang T, Wang DZ, 2003. Imbalance of Th1 and Th2 cytokine production by peripheral blood mononuclear cells in patients with gestational hypertension syndrome. Chin J Per Med 4: 4-7. [in Chinese]
Yang J, Duan XY, Huang T, He JB, Jia YC, Guo H, Peng X, Zheng ERD, 2020. Spatial and temporal epidemiological characteristics of hand, foot and mouth disease in Yunnan Province, 2014-2018. Chin J Dis Contr 24:290-296.
Yang Q, Tang QH, Zhang YY, 2019. Spatial and temporal characteristics of water quality in the Huaihe River Basin (Henan section) and its relationship with land use types. Env Science Res 32:1519-1530.
Yang X, Xu G, Li Q, Guo Y, Zhang M, 2017. Authorship attribution of source code by using back propagation neural network based on particle swarm optimization. PLoS ONE 12:e187204.
Yue WZ, Xu JH, Xu LH, 2005. Spatial interpolation of climate elements based on geostatistical methods. Highl Met 06: 974-980. [in Chinese]
Zahedi A, Hassanvand MS, Jaafarzadeh N, Ghadiri A, Shamsipour AM, Dehcheshmeh MG, 2021. Effect of ambient air PM2. 5-bound heavy metals on blood metal (loid) s and children's asthma and allergy pro-inflammatory (IgE, IL-4 and IL-13) biomarkers. J Trace Elem Med Bio 68:126826.
Zhang HX, Niu SHW, Qi JH, Ye LQ, Li N, 2016. Geostatistical analysis of population distribution in Henan Province based on township scale. Geogr Stud 35:325-336.
Zhang J, Liang X, Liu YF, Zhang X, Sun LQ, Zhao F, Fu PY, 2021. Spatial distribution prediction of arsenic in groundwater based on principal component co-kriging method. Earth Science 1-22. [in Chinese]
Zhou ZS, Ji ZQ, Zhu XJ, Wang YQ, Xue WL, Tao HW, Jiang WQ, Liang WH, 2014. Correlation of rs3024608, rs1110470, and rs3024685 polymorphisms of IL-4R gene with metacarpal atopic signs in asthmatic patients with large fissure. Chin J Trad Chin Med 29:1076-1079.
Zhu HB, He LJ, 2009. A study on the conditions of consistency test for normal distribution etc. using single sample K-S test in SPSS. J Cap Inst Phy Educ 21:466-470.
Zhu QA, Zhang WC, Yu JH, 2004. A study of GIS-based spatial interpolation method. J Jiangxi Norm Univ (Natural Science Edition) 02: 183-188. [in Chinese]