Neural Network Analysis Model of Air Environmental Factors for Urticaria Skin Disease Susceptibility

Weihua Wang*
Beijing Jiaotong University, Beijing, 100044 China
*Corresponding Author Email: whwang1@bjtu.edu.cn

Abstract. There is a certain correlation between the occurrence of skin diseases and external inducing factors, among which air quality is one of the important factors. Based on the analysis of outpatient medical records in a certain period of time, taking urticaria as a representative, the degree of skin disease susceptibility in outpatient medical records was divided. Combined with the changes of air quality indicators in the same period, the prediction model for analyzing the correlation between urticaria skin diseases and air quality indicators was established by using artificial neural network method, and a new way of skin disease risk prediction was discussed.

1. Introduction
Due to the different influence of external environmental factors, the risk of various types of skin diseases also varies greatly. In addition, the inducing effect of air environmental factors on skin diseases itself is not easy to accurately measure, and the degree of action of various related factors is not clear. For this kind of complex correlation problems, it is necessary to consider the nonlinear comprehensive analysis method of complex factors. Artificial neural network (ANN) is an artificial intelligence method that abstracts human brain neural network from the perspective of information processing, establishes models, and comprehensively analyzes complex multi factor problems. In recent years, the research work of artificial neural network has been deepened continuously, and great progress has been made. Many modern strategies have been successfully solved in the fields of automatic control, biology, medicine, economy, etc The computer is difficult to solve the practical problems, showing good intelligent characteristics[1-6].

2. Neural Network Method
Artificial neural network is a system that simulates the structure and characteristics of human brain neural network by means of engineering technology. Artificial neural network can be used to construct neural networks with different topological structures. It is a simulation and approximation of biological neural networks. The main connection forms of neural network are feedforward neural network and feedback neural network. The commonly used feedforward types are perceptron neural network and BP neural network, and the commonly used feedback type is Hopfield network. BP neural network, namely error back propagation algorithm. It is a kind of multilayer feedforward network trained by error back propagation algorithm, and is one of the most widely used neural network models. The topological structure of BP neural network model includes input layer, hidden layer and output layer. The hidden layer can be one layer or multi-layer.
Any continuous mapping function from input to output can be realized by a three-layer nonlinear network. BP algorithm consists of two processes: forward calculation of data stream (forward propagation) and back propagation of error signal. In forward propagation, the direction of propagation is input layer $\rightarrow$ hidden layer $\rightarrow$ output layer. The state of neurons in each layer only affects the neurons in the next layer. If the desired output is not obtained at the output layer, the reverse propagation process of the steering error signal is carried out. Through the alternation of these two processes, the gradient descent strategy of error function is implemented in the weight vector space, and a group of weight vectors are searched dynamically to make the network error function reach the minimum value, thus completing the process of information extraction and memory. The process of error back-propagation is the process of weight learning, and the network weights are updated according to different training modes. Online mode and batch mode are commonly used. The training samples in online mode are processed one by one, while all training samples in batch mode are processed in batches. In the process of establishing ANN analysis model, the parameters that need to be determined are: initial weight, threshold, learning rate parameter, momentum parameter, hidden layer node number, etc. The BP learning and training process is as follows: (1) initializing the network, assigning the network parameters and weight coefficients, the weight coefficients are random numbers; (2) input the training samples, through the connection between nodes, the values of each layer are calculated layer by layer, and compared with the real value, the input error of the network is obtained; (3) according to the error back-propagation rule, the weight coefficient between each layer is adjusted according to the gradient descent method, so that the connection weight of the whole neural network is transformed to the direction of error reduction; (4) repeat (2) and (3) until the prediction error meets the conditions or the training times reach the specified number.

The neural network algorithm needs different parameters for different networks, and the selection of parameters needs to be trained and tested according to the known test data. Trial and error of neural network: neural network needs to set more parameters, initial weight, learning rate, momentum coefficient, hidden layer node number need to be set according to specific problems, the more the number of classification, the greater the difficulty of parameter adjustment. In the process of adjustment, it is necessary to observe the change of error and accuracy. No matter which parameter is adjusted, the following principle is followed: if the parameter is reduced, the correct rate will increase, then continue to reduce the parameter to reach a certain critical value. If the accuracy rate of the parameter is reduced, the optimal value of the parameter should be found near the critical value[7-11].

3. Neural Network Analysis Model
The basic idea of evaluating and forecasting is to use the data of specific time periods as typical units, and quantify the meteorological factors that may affect the incidence rate as input values of the input nodes. According to the different degree of risk, the possibility of disease is divided into different grades and quantized as the expected output of the nodes at the output level. The network is trained with these known samples until the total error of the network reaches the accuracy requirement, that is to say, the network has mastered the relationship between the elements and the expected output. Then,
the prediction results can be directly output through the associative memory function of neural network to evaluate and predict the disease risk in other periods.

3.1 Typical Sample Data
A total of 1600 newly diagnosed dermatological outpatients were selected as the research objects. There were 900 males and 700 females. The average age was 19 years. Dermatologists fill in the form after physical examination. The contents include age, gender, place of birth, health habits, dietary preference, skin properties, skin diseases, and diagnostic criteria are based on modern dermatology and Venereology. A total of 1600 cases were investigated, and 45 cases and 2321 cases of skin diseases were found. Some cases had two or more than two kinds of skin diseases. According to the classification of clinical dermatology, the incidence of common skin diseases was 25.5%. The incidence rate of urticaria is the highest and the number of patients is the largest in each disease case. Some kinds of skin diseases show the characteristics of regular changes with time and seasons. According to the outpatient service of newly diagnosed college students, investigation and statistical analysis, dermatomycosis. Facial acne. Allergic dermatitis and rose rash have certain seasonality. Especially, superficial mycosis has obvious seasonality. It occurs frequently in spring and summer. Among the above diseases, acne, allergic dermatitis and seborrheic dermatitis are related to diet. Acne and seborrheic dermatitis are also related to students’ stress and eating habits. Urticaria, allergic dermatitis, and pityriasis rosea are common in spring. Neurodermatitis has little to do with seasons, but it has obvious relationship with mental stress and tension. Psoriasis, vitiligo. It is not obvious that corns, contact dermatitis and wart diseases are related to seasons. Taking urticaria as an example, the correlation between disease occurrence and ambient air quality is analyzed. The time distribution of urticaria cases during the investigation period is shown in Table 1 and table 2.

| time     | 9-17 | 10-17 | 11-17 | 12-17 | 1-18 | 2-18 | 3-18 | 4-18 | 5-18 | 6-18 | 7-18 | 8-18 | 9-18 | 10-18 |
|----------|------|-------|-------|-------|------|------|------|------|------|------|------|------|------|-------|
| Number of cases | 12   | 11    | 3     | 2     | 1    | 0    | 2    | 13   | 10   | 7    | 3    | 2    | 14   | 13    |
| Ratio %  | 7.0  | 6.4   | 1.8   | 1.2   | 0.6  | 0    | 1.2  | 7.6  | 5.8  | 4.1  | 1.8  | 1.2  | 8.2  | 7.6   |

| time     | 11-1 | 12-1 | 1-19 | 1-19 | 2-19 | 3-19 | 4-19 | 5-19 | 6-19 | 7-19 | 8-19 | 9-19 | 10-1 | 9     | 11-1 | 9     | 12-1 | 9     |
|----------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Number of cases | 7    | 1    | 3    | 2    | 7    | 15   | 10   | 7    | 5    | 2    | 12   | 3    | 2    | 2    |
| Ratio %  | 4.1  | 0.6  | 1.8  | 1.2  | 4.1  | 8.8  | 5.8  | 4.1  | 2.9  | 1.2  | 7.0  | 1.8  | 1.2  | 1.2  |

3.2 Air Quality Data over the Same Period
According to the data released by China air quality online monitoring and analysis platform of China Meteorological Observatory, the air quality index data of Beijing from September 2017 to December 2019 are shown in Table 3.
Table 3. Air quality monitoring data in Beijing

| Month / year | PM2.5 | PM10 | SO₂ | CO   | NO₂ | O₃   |
|--------------|-------|------|-----|------|-----|------|
| Sep-17       | 58    | 100  | 4   | 0.913| 49  | 123  |
| Oct-17       | 57    | 63   | 3   | 0.923| 46  | 47   |
| Nov-17       | 46    | 73   | 5   | 0.903| 49  | 44   |
| Dec-17       | 44    | 68   | 8   | 1.09 | 49  | 25   |
| Jan-18       | 34    | 65   | 7   | 0.768| 39  | 48   |
| Feb-18       | 50    | 72   | 9   | 0.786| 31  | 70   |
| Mar-18       | 82    | 104  | 9   | 1.003| 53  | 81   |
| Apr-18       | 59    | 108  | 6   | 0.673| 38  | 115  |
| May-18       | 45    | 85   | 4   | 0.648| 36  | 134  |
| Jun-18       | 43    | 67   | 4   | 0.7  | 32  | 168  |
| Jul-18       | 44    | 48   | 3   | 0.787| 24  | 135  |
| Aug-18       | 31    | 46   | 3   | 0.745| 26  | 133  |
| Sep-18       | 28    | 45   | 3   | 0.543| 34  | 93   |
| Oct-18       | 42    | 59   | 4   | 0.668| 45  | 70   |
| Nov-18       | 71    | 104  | 7   | 1.07 | 60  | 37   |
| Dec-18       | 38    | 73   | 7   | 0.835| 46  | 40   |
| Jan-19       | 51    | 80   | 9   | 0.997| 50  | 44   |
| Feb-19       | 53    | 72   | 6   | 0.807| 35  | 65   |
| Mar-19       | 52    | 80   | 5   | 0.635| 39  | 86   |
| Apr-19       | 48    | 91   | 4   | 0.547| 34  | 103  |
| May-19       | 37    | 78   | 4   | 0.503| 34  | 143  |
| Jun-19       | 39    | 60   | 4   | 0.713| 27  | 179  |
| Jul-19       | 37    | 50   | 3   | 0.661| 27  | 159  |
| Aug-19       | 23    | 37   | 3   | 0.571| 26  | 118  |
| Sep-19       | 36    | 58   | 3   | 0.67 | 33  | 151  |
| Oct-19       | 40    | 66   | 3   | 0.681| 41  | 42   |
| Nov-19       | 44    | 78   | 5   | 0.797| 48  | 32   |
| Dec-19       | 44    | 60   | 4   | 0.845| 46  | 37   |

(Note: data released by China air quality online monitoring and analysis platform)

3.3 Prediction Model

PM2.5, PM10, SO₂, CO, NO₂ and O₃ were selected as the control indicators of external air quality factors affecting the occurrence of skin diseases, and the possible impact on the incidence of skin diseases was analyzed. PM2.5, SO₂, CO and NO₂ were selected as the main influencing factors. The basic idea of establishing the prediction model is as follows: PM2.5, SO₂, CO, NO₂, which may affect the occurrence of urticaria skin disease, are taken as the input values of each node in the network input layer. According to the investigation and analysis time interval, the incidence rate of urticaria of key skin diseases, based on the total number of cases in the survey interval, and according to the total number of cases in 28 months, the occurrence of urticaria is divided into 4 grades: high incidence, high incidence, medium incidence and low incidence (Table4). It is regarded as the expected output of each node in the output layer. These known cases are used as known samples to train the network until the network has mastered the complex nonlinear relationship between the data. Then the air quality index data of the time period to be predicted is used as the prediction sample and input into the network which has been learned well, and the network can output the prediction result directly through its associative memory function.
Table 4. Urticaria skin disease susceptibility

| Susceptibility Level | Cases / Month |
|----------------------|--------------|
| High susceptibility  | More than 25 |
|                      | cases / month|
| Relatively high      | 15-25 cases  |
|                      | / month      |
| Moderate susceptibility| 5-15 cases   |
|                      | / month      |
| Low susceptibility    | Less than 5  |
|                      | cases / month|

Two value method was used to determine the value of output layer variables. The grade with the closest degree of skin disease incidence was taken as 1, and the others were taken as 0, i.e. low incidence was (1,0,0,0), moderate incidence was (0,1,0,0), high incidence was (0,0,1,0), and high incidence was (0,0,0,1). Five nodes are selected in the middle hidden layer to improve the association ability of the network. Using the above network, the air quality data example data given in Table 4 is used to train the network. After 10000 iterations, the accuracy meets the requirements, and the sum of squares of the error between the actual output and the expected output is less than $1.0 \times 10^{-4}$. The network training was successful, and the connection weights and node thresholds of each node (Table 5 and table6) were obtained. Thus, a neural network model was established to determine the susceptibility of urticaria skin diseases based on the actual air quality data.

Table 5. Connection weights and hidden layer thresholds between input layer and hidden layer nodes

| Input layer node | Hidden layer node 1 | Hidden layer node 2 | Hidden layer node 3 | Hidden layer node 4 | Hidden layer node 5 |
|------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| 1                | -9.21               | 2.88                | -15.79              | 10.65               | 3.39                |
| 2                | 2.17                | -15.21              | 8.27                | -13.69              | 10.59               |
| 3                | 8.11                | 7.61                | 7.92                | 1.55                | -22.44              |
| 4                | -30.44              | 13.18               | 5.32                | -5.31               | 4.19                |
| Hidden layer threshold | 3.55               | 2.99                | -21.11              | 16.18               | -0.21               |

Table 6. Connection weights and output layer thresholds between hidden layer and output layer nodes

| Hidden layer node | Output layer node 1 | Output layer node 2 | Output layer node 3 | Output layer node 4 |
|-------------------|---------------------|---------------------|---------------------|---------------------|
| 1                 | 7.68                | -15.45              | 17.89               | -9.79               |
| 2                 | -9.29               | -9.44               | 11.69               | -8.23               |
| 3                 | -1.22               | -6.19               | -9.33               | 14.19               |
| 4                 | 14.31               | -8.33               | -21.09              | 19.51               |
| 5                 | -0.98               | 15.71               | -14.15              | -13.92              |
| Output layer threshold | -12.44             | -5.81               | 5.82                | -4.22               |
3.4 Model Application

According to the air quality samples of the target period given in Table 7, the air quality of the new date was used to analyze the risk of urticaria skin diseases. It can better deal with the complex nonlinear relationship between the main factors and the susceptibility of urticaria skin diseases, which has a reference value for the determination of the susceptibility of urticaria skin diseases affected by air quality.

| Date   | PM2.5 | SO2 | CO   | NO2 | Decision output | Error value | Degree of prediction | Risk degree of susceptibility |
|--------|-------|-----|------|-----|-----------------|-------------|---------------------|-----------------------------|
| 2020-04| 35    | 4   | 0.403| 22  | 0.000           | 0.051       | 0.053               | Low susceptibility           |
| 2020-05| 31    | 4   | 0.603| 23  | 0.000           | 0.048       | 0.045               | Low susceptibility           |

4. Main Conclusions

Considering the air quality indexes that may have some external influence on the occurrence of skin diseases, a neural network prediction model is established to predict the susceptibility of skin diseases affected by air quality. Through learning and training a large number of actual observation data, the network can master the mapping relationship between skin disease susceptibility and various influencing factors, and the more comprehensive the influencing factors are considered, the more samples are learned, the more characteristics of network memory will be, which makes the prediction accuracy of the network continuously improved. The establishment and application of neural network model can provide a new reference for skin disease prevention.

5. Reference

[1] Bao Zhiqiang, Zhao zhichao, Wang yuting. fast convolution neural network algorithm and its application [J]. computer engineering and design, Vil.41, No.8, 2020:2213-2217
[2] Li Lifeng, Yang Hua, Zhang Juan, Liu Qifang. Study on regional landslide prediction based on artificial neural network [J]. Meteorological and Environmental Sciences.2020, Vol. 43 No.3: 65-70
[3] Qu yajing, Wei haiying, ma Jin. prediction of PAHs content in Beijing urban parks based on BP neural network [J].2020,Vol.43 No.3:2-5
[4] An Xiaoyu, Lu Kuihao, Cui Guangzhao, Zhao Fuxing. Prediction of residual chlorine in water based on improved sine and cosine optimization neural network [J]. Yellow River. Vol.42, No.5, 2020: 110-115
[5] Lin Yuan, Guan Zhidong. Prediction of composite autoclave molding uniformity by BP neural network [J]. Journal of Beijing University of Aeronautics and Astronautics. Vol.46, No.6, 2020: 2345-2350
[6] Liu Y Tong, Shi chunrui, Dong Jiyuan, et al. Study on the correlation between PM2.5 concentration in air and daily outpatient visits of dermatitis and eczema [J]. Journal of clinical dermatology, Vol 47, no.92018: 559-563
[7] Chen ya, Xian Liu Wei. Environmental air pollution and skin health [J]. Chinese Journal of Dermatology, Vol 48, No. 12015: 67 ~ 70
[8] Zhou Kaili, Tang Yaohong. Neural network model and its matlab simulation program design [M]. Beijing: Tsinghua University Press, 2005
[9] Han Liqun. Design and application of artificial neural network theory [M] 2. Beijing: Chemical Industry Press, 2007
[10] Gong Jun, Du Chao, Zhong Xiaogang, Xiang Tianyu, Wang Huilai. Study on the risk of primary hypertension complicated with coronary heart disease based on machine learning algorithm [J]. Chinese PLA medical journal, 2020, 7 online edition
[11] Cao Yun, Lu Yi, Chen Jianxin, et al. Application of intelligent syndrome differentiation model of gastroesophageal reflux disease based on machine learning [J]. Journal of Beijing University of traditional Chinese medicine, Vol 42, no.102019: 869-874