

# PREDICTION OF VEGETATION COVER INDEX ALONG HIGHWAYS IN HIGH-ALTITUDE AREAS USING MACHINE LEARNING

JIA, X. – LIU, F.\* – WANG, D.

*School of Highway, Chang'an University, Xi'an 710064, China*  
(*e-mail: jiaxingli@chd.edu.cn, wangdang@179997689.com; phone: +86-134-6893-1646, +86-187-0090-7083*)

*\*Corresponding author*  
*e-mail: liufabao@1950480311.com; phone: +86-183-0926-4289*

(Received 16<sup>th</sup> May 2019; accepted 16<sup>th</sup> Jul 2019)

**Abstract.** The ecological environment in high-altitude areas are fragile. Improper highway construction or excessive vehicle emission may lead to irreparable damages to the natural environment in such areas. To ensure the sustainable development of the ecological environment in high-altitude areas, it is essential to evaluate the influence of highways construction and operation on the environment in such areas. In this article, using the normalized difference vegetation index (NDVI) to indicate environmental changes in the high-altitude areas road domain. Studying sections from Pengjiazhai Town to DaYagen, DaYagen to Dongxia Town and Dongxia Township to Riyue Tibetan Town along the National Highway G214, with the altitude ranging from 2620~3760 m. Four different machine learning methods including Extreme Learning Machine, Wavelet Neural Network, BP Neural Network and Cubic Smoothing Index are then applied to analyze the NDVI changes in study areas in order to build prediction models. The results of the MAE, WMAE, RMSE and R<sup>2</sup> value of the four different models show that the Wavelet Neural Network model works the best in predicting the NDVI in high-altitude areas. Based on this result, it is suggested that Wavelet Neural Network is more suitable for the intelligent prediction of road vegetation coverage in high-altitude areas.

**Keywords:** *normalized difference vegetation index (NDVI), prediction models; high-altitude highways, road land vegetation cover, machine learning*

## Introduction

Due to the vulnerability and sensitivity of the ecological environment in high-altitude areas, it can be easily destroyed by engineering construction or other human activities compared to normal areas. Moreover, the ecological recovery ability of such areas is relatively weak and the recovery cycle is long. However, during the highways construction and operation process, the surrounding environment will be affected (Tu, 2016). For example, vehicle emission will cause the reduction of the surrounding vegetation coverage, which may further lead to soil degradation. In addition, the related indexes used to characterize the ecological environment of the highway domain are uncertain due to the particularity of the region and the data is usually classified. Therefore, it is essential to explore a certain algorithm that can be used to estimate and predict the environmental influence of highway on the surrounding environment in order to better reveal the environment change in road domain and to keep the sustainable development of the ecological environment in high-altitude areas. About vegetation indexes there are different kinds of vegetation indices that are commonly used, including Normalized Difference Vegetation Index (NDVI), ratio vegetation index, difference vegetation index, vertical vegetation index, environmental vegetation index and so on. Within which the NDVI is the most commonly used index in

vegetation research due to its advantages of detecting vegetation growth state, vegetation coverage and eliminating partial radiation errors. It is the best indicator of plant growth state and vegetation spatial distribution density, and can reflect the influence of plant canopy. Therefore, NDVI is selected to characterize the vegetation environment around the highway region in this study.

As for the research methods, five traditional prediction methods were used in previous research including empirical prediction method, trend extrapolation method, regression prediction method (Abutaqiya et al., 2019; Zhu et al., 2019), time series prediction method (Yang and Chen, 2019; Moews et al., 2019; Zhang et al., 2019), and gray prediction (Guo et al., 2019; Lin et al., 2018; Zhao and Zhou, 2018). The above five methods are more inclined to capture the data distribution characteristics through simple mathematical model. For instance, the regression prediction method and the gray theory fit more to the linear distributed data. Such prediction methods can not describe the characteristics of more complex nonlinear distributed data and can not reflect the accurate change of the dataset. On the other hand, with the continuous development of scientific information technology and the application of big data analysis, artificial intelligence prediction methods were used in environment-related research. For example, neural network (Chen, 2018; Leng et al., 2017), fuzzy logic (Xu et al., 2017) and multiple combination prediction methods (Mishra et al., 2015; Feng et al., 2015; Perez and Gramsch, 2016; Sun et al., 2013) were used to predict the concentration of PM<sub>2.5</sub>. In recent years, different machine learning models are introduced to predict wind, electricity, solar, coal and other resources. Cocchi et al. (2017) applied the time series analysis method, S-ARIMA, S-ARIMAX, extreme learning machine, support vector machine (SVR) and other methods in renewable energy market price bidding. It turned out that the accuracy of Support Vector Machines is superior to other methods. Liu et al. (2018) used wavelet decomposition, empirical mode decomposition and extreme learning machine methods to predict wind speed multi-step. Li et al. (2018) predicted energy consumption based on grey prediction theory and support vector machine that optimize through extreme learning machine. The results showed that the combination of SVR and ELM is more effective than single prediction method. Yi et al. (2019) applied the extreme learning machine to predict the growth of algae in the river. It was concluded that the ELM model had better prediction and generalization performance than the traditional BP neural network and adaptive neuro-fuzzy inference system. Fijani et al. (2019) employed a two-layer analysis method combined with a hybrid model formed by an extreme learning machine to monitor and predict the relevant quality parameters of water in the environment. By using the mixed model analysis, it was found that the results were more accurate than the single model in the prediction of results. Kisi et al. (2018) utilized the new wavelet combination model to establish a reference evaporation prediction model by comparing the wavelet extreme learning machine with the neural network. It was found that wavelet connection models usually had better precision than a single model, while WELM model was more suitable for predicting daily evaporation.

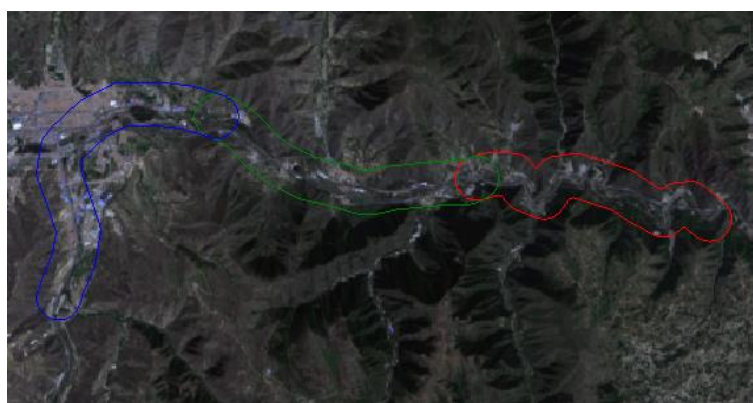
With the economic and infrastructure development in recent years, the demand for highway infrastructure in high-altitude areas is increasing. However, related research methods less attention were paid on the prediction of environmental indicators in highway regions. Considering the vulnerability and sensitivity of the ecological environment in high-altitude areas, it is important to explore the environmental influence of highway construction and operation in such areas. At the same time, it is

necessary to discuss which machine learning method can be applied to the road domain NDVI value in high-altitude areas. In addition to, the diversity and uncertainty of the factors affecting the road ecology asks for prediction models with higher ability and accuracy. Considering the characteristics of the discrete-distributed data, ELM, wavelet neural network, BP neural network and cubic smoothing index are considered to predict the NDVI along road domain. These methods are suitable for capturing the nonlinear characteristics and have good global learning ability. In this paper, spatial analysis of the NDVI along National Highway G214 were conducted using GIS. The extreme learning model, wavelet neural network, BP neural network and cubic smoothing index were applied to predict the vegetation index of high-altitude highway roads. This research explored the possibility of using machine learning method in exploring the ecological environmental impact of the road in the high-altitude areas. The results would help improve the sustainable development of the ecological environment in high-altitude areas.

## Materials

In order to study and forecast the normalized difference vegetation index of highway in high-altitude region by using machine learning algorithm, G214 highway of Xi'ning City, Qinghai Province of China is selected for research and analysis. In this paper, we apply ArcGIS and ENVI tools to process and analyze the remote sensing image data in the research areas (101°20'~101°57'E; 36°55'~36°65'N). Remote sensing (RS) technology can quickly obtain the regional multi-factor surface information, and through combining the spatial analysis technology of GIS, can quickly get the required information for the target areas. On the satellite selection, the Landsat-8 satellite was chosen. The Landsat-8 satellite was launched on February 11, 2013 and is a follow-on satellite to the Landsat series of US terrestrial satellites. It is equipped with land imager (OLI) and thermal infrared sensor (TIRS). It has important applications in many fields, such as natural resource protection, energy exploration, environmental management, natural disaster monitoring and so on.

The starting point of the research data extraction is G214 Pengjiazhai Town, and the end point is the Riyue Tibetan Township. In order to reflect the particularity of the road environment indicators, the research range is divided into three sections, specifically Pengjiazhai Town~Dayagen; Dayagen~Dongxia Township; Dongxia Township~Riyue Tibetan Township, as shown in *Figure 1*.



**Figure 1.** Three sections of road research domain

After the remote sensing image is processed by ArcGIS and ENVI, the image is clipped by ENVI. The main purpose is to eliminate unnecessary research areas. The range of the research road domain is 400 m on each side of the road of the G214 median line. The vector polygon of the study areas uses ArcGIS to analyze the spatial buffer of the road center line data, and the buffer radius is 400 m. Then the image clipping operation of ENVI software is used to obtain the image data of the research areas. The cropped remote sensing image is shown in *Figure 2*.



*Figure 2. The overall research domain obtained by clipping*

As for the research indicator of vegetation index, also known as spectral vegetation index, refers to the multi-spectral data obtained from remote sensing sensors, which are composed of linear and nonlinear combinations of various values that have certain indicative significance to vegetation (Wang, 2015). Because of the strong absorption of blue and red lights radiation by plant leaf tissue, but the strong reflection of near infrared radiation, the near infrared reflection increases with the increase of vegetation. Therefore, the linear or non-linear combination of red band and near infrared band is usually used to characterize the quantity distribution and quality of surface vegetation. This paper studies the vegetation coverage information by extracting NDVI to characterize the ecological environment in the G214 research areas.

As a remote sensing index reflecting the status of land cover vegetation, NDVI is defined as the quotient of the difference between the reflectivity of the near infrared band and the red band. The formula for calculating the difference between the reflectivity of the near infrared band and the red band is as follows:

$$NDVI = \frac{NIR - R}{NIR + R} = \frac{Band5 - Band4}{Band5 + Band4} \quad (\text{Eq.1})$$

In the formula: Band4 and Band5 are the fourth (red) band luminance value and the fifth (near infrared) band luminance value of Landsat-8 remote sensing satellite OLI image, respectively. The calculation process is accomplished by ENVI remote sensing image processing software, and the NDVI is calculated by reflectivity image.

Specific Landsat-8 remote sensing satellite data selection from 2013~2018, according to the quarterly selection of one to two small cloud cover, small interference factors of Landsat8 remote sensing satellite map to establish a buffer radius of 400 m range of data extraction. The extracted data are shown in *Tables 1–3*.

**Table 1.** Pengjiazhai Town ~ Dayagen NDVI data

Pengjiazhai ~Dayagen	Jan	Feb	Apr.	June	Jul	Aug	Sep	Oct	Nov	Dec
2013	-	-	0.217734	0.579018	-	-	0.438763	-	-	0.117657
2014	-	0.099117	0.173993	-	-	0.491571	0.472169	0.350313	-	0.105011
2015	-	0.134264	0.134264	-	-	0.536533	-	0.326867	-	0.140341
2016	0.144673	0.184593	0.185297	-	-	0.53848	--	-	0.202526	-
2017	0.140669	-	0.19548	-	0.59262	-	-	-	0.147942	-
2018	0.137455	-	0.227371	-	-	-	-	-	-	-

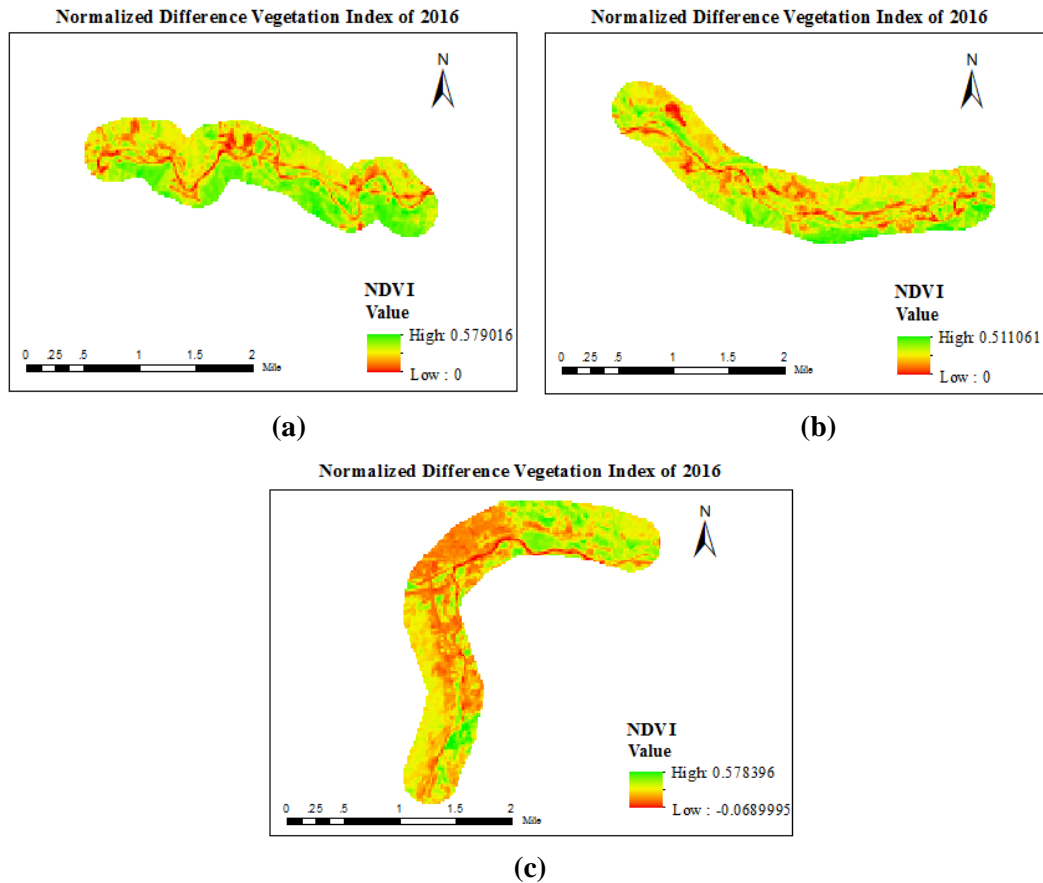
**Table 2.** Dayaegen ~ Dongxia Township NDVI data

Dayaegen-Dongxia Township	Jan	Feb	Apr.	June	Jul	Aug	Sep	Oct	Nov	Dec
2013	-	-	0.202384	0.511061	-	-	0.391358	-	-	0.114336
2014	-	0.096136	0.188883	-	-	0.353186	0.457984	0.351201	-	0.240173
2015	-	0.192425	0.200327	-	-	0.526266	-	0.343435	-	0.254257
2016	0.247882	0.306538	0.364527	-	-	0.557571	--	-	0.266581	-
2017	0.152848	-	0.365258	-	0.56718	-	-	-	0.208416	-
2018	0.184353	-	0.316285	-	-	-	-	-	-	-

**Table 3.** Dongxia Township-Riyue Tibetan Township NDVI data

Dongxia Township-Riyue Tibetan Township	Jan	Feb	Apr	June	Jul	Aug	Sep	Oct	Nov	Dec
2013	-	-	0.22077	0.578396	-	-	0.444658	-	-	0.168276
2014	-	0.119978	0.218701	-	-	0.383716	0.538561	0.367532	-	0.240173
2015	-	0.192425	0.200327	-	-	0.567645	-	0.408466	-	0.254257
2016	0.247882	0.306538	0.364527	-	-	0.580294	-	-	0.271642	-
2017	0.152848	-	0.365258	-	0.566043	-	-	-	0.208416	-
2018	0.18822	-	0.316285	-	-	-	-	-	-	-

It can be known from the above tables that normalized difference vegetation index is extracted according to the range of road domain segments, and the results of NDVI extraction are limited to [-1,1], which avoids the inconvenience caused by too large or too small data. The negative value indicates that the surface cover is cloud, water, snow and so on, reflecting the height of visible light; zero means there are rocks or bare soil, Band4 and Band5 are approximately equal; positive value indicates vegetation cover, which increases with the increase of coverage. In general, there is a strong positive correlation between vegetation index and vegetation coverage. The higher the positive value of NDVI, the higher the vegetation coverage rate is, and the better the corresponding ecological background environmental quality is. In this study, taking the remote sensing satellite map of June in the second quarter of 2013 as an example, NDVI data extracted from ArcGIS are used to show the results of the data. Just showing in *Figure 3* of (a) is Pengjiazhai Town to Dayagen; (b) is Dayagen to Dongxia Township; (c) is Dongxia Township to Riyue Tibetan Township.



**Figure 3.** (a) is NDVI value of Pengjiazhai Town to Dayagen; (b) is Dayagen to Dongxia Township; (c) is Dongxia Township to Riyue Tibetan Township

## Four types of methods

### ELM

ELM is a new training algorithm for single hidden layer feed forward networks, (SLFNs) (Huang et al., 2006). Before the training begins, ELM algorithm selects the excitation function of hidden layer, and randomly selects the number of hidden layer nodes in the network, and inputs the weights. Let  $N$  different random data sample data are  $G = \{(x_1, t_1), (x_2, t_2), \dots, (x_N, t_N)\}$ ,  $x_i = [x_{i1}, x_{i2}, \dots, x_{in}] \in R^n$ ,  $t_i = [t_{i1}, t_{i2}, \dots, t_{im}] \in R^m$ ,  $i = 1, 2, \dots, N$ . Single hidden layer number is  $L$ , and the excitation function  $g(x)$  that standard SLFNs model just show as:

$$\sum_{i=1}^L \beta_i g_i(x_j) = \sum_{i=1}^L \beta_i g(a_i \cdot x_j + b_i) = t_j, j = 1, 2, \dots, n \quad (\text{Eq.2})$$

In the formula:  $a_i = [a_{i1}, a_{i2}, \dots, a_{in}]^T$  is the input weight vector that connects the node of the  $i$  hidden layer;  $b_i$  is the bias for  $i$  hidden layer nodes;  $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$  is an output weight vector connecting  $i$  hidden layer nodes;  $a_i \cdot x_j$  represent the inner product of  $a_i$  and  $x_j$ .  $g(x)$  contains  $L$  hidden layer nodes. Its standard SLFNs can approach  $L$

training sample data infinitely, and there exists  $\beta$ ,  $a_i$  and  $x_j$  can make it possible to build a formula like this.

Then we can set up the following extreme learning machines:

$$\sum_{i=1}^L \beta_i g(\alpha_i \cdot x_j + b_i) = t_j, j = 1, 2, \dots, n \quad (\text{Eq.3})$$

The above formula is transformed to form the following matrix:

$$H\beta = T \quad (\text{Eq.4})$$

H is the hidden layer output matrix of the definite matrix network, Column  $i$  representation and input  $x_1, x_2, \dots, x_N$  as the output vector of the associated  $i$  hidden layer node, Line  $j$  represents the implicit layer output vector associated with the input.

In the formula:

$$H(\alpha_1, \alpha_2, \dots, \alpha_L, b_1, b_2, \dots, b_L, x_1, x_2, \dots, x_N) = \begin{matrix} g(\alpha_1 - x_1 + b_1) & g(\alpha_2 - x_1 + b_2) & \dots & g(\alpha_L - x_1 + b_L) & \beta_1^T & t_1^T \\ g(\alpha_1 - x_2 + b_1) & g(\alpha_2 - x_2 + b_2) & \dots & g(\alpha_L - x_2 + b_L) & \beta_2^T & t_2^T \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ g(\alpha_1 - x_N + b_1) & g(\alpha_2 - x_N + b_2) & \dots & g(\alpha_L - x_N + b_L) & \beta_L^T & t_L^T \end{matrix}_{N \times L} \quad (\text{Eq.5})$$

Training SLFNs is equivalent to finding out least square solution of  $\beta$  for the  $H\beta = T$  just like this:  $\beta = \min_{\beta} \|T(a_1, \dots, a_N, b_1, \dots, b_N)\beta - T\|$ .

And the Moore-Penrose generalized inverse matrix of the hidden layer output matrix  $H^+$  can be expressed as follows:  $H^+ \beta = T$ .

The solution is used as the weight parameter on the right side of the network. Overcoming the traditional artificial neural network prediction network prediction model some disadvantages such as the training time is long, and it is easy to produce over-fitting. The method mainly includes input layer, hidden layer and output layer, so when we are doing algorithmic calculations, it only needs to determine the number of neurons in the hidden layer to carry out the next operation, which overcomes the difficulty of determining multiple computational parameters and the complicated programming process of the traditional neural network system.

### 3.2. Wavelet neural network

Wavelet neural network (WNN) is a new type of neural network based on wavelet theory (Nian et al., 2001), which replaces the usual linear or nonlinear excitation function with wavelet basis which can scale and translate. The time-frequency local characteristic of wavelet transform is combined with the self-learning of neural network. Compared with the standard neural network, the structure of wavelet transform is simpler, the convergence is faster, and the blindness is effectively avoided.

In WNN, the number of input layer nodes, hidden layer wavelet elements and output layer nodes are M; J; N; respectively. The expansion coefficient of the j-th wavelet

element in the wavelet layer is  $a_j$ , translation coefficient is  $b_j$ , The weights of the  $m$  neurons in the input layer to the  $j$  wavelet elements in the hidden layer are  $\omega_{mj}$ . The connection weights of the  $j$ -th wavelet element to the  $n$ -th output neuron is  $\omega_{jn}$ . Output layer  $n$ -th node ideal output value is  $y_n^*$ . The wavelet layer uses non-orthogonal Morlet wavelet functions. The classical Tansig function is used in the output layer. WNN is a compact combination of wavelet analysis and BP neural network. Based on the structure of BP neural network, wavelet basis function is used to replace the traditional Sigmoid function as the transfer function of neuron nodes.

The input of the  $j$ -th wavelet element of the hidden layer is:

$$r_j = \sum_{m=1}^M \omega_{mj} X_m \quad (\text{Eq.6})$$

The output of the  $j$ -th wavelet element is as follows:

$$c_j = \psi_{a_j, b_j}(r_j) = \cos(1.75 \frac{r_j - b_j}{a_j}) \cdot \exp(-0.5(\frac{r_j - b_j}{a_j})^2) \quad (\text{Eq.7})$$

The output layer's  $n$ -th node outputs as follows:

$$y_i = 1 / (1 + \exp(-\sum_{n=1}^N \omega_{jn} c_j)) \quad (\text{Eq.8})$$

The network output error function is:

$$E = \frac{1}{N} \sum_{n=1}^N (y_n^* - y_n)^2 \quad (\text{Eq.9})$$

The training of WNN is mainly aimed at  $\omega_{mj}$ ,  $a_j$ ,  $b_j$ ,  $\omega_{jn}$  to optimized and selected. WNN has the advantage of faster convergence than the forward neural network. Firstly, the basic elements and the whole structure of the WNN are determined according to the wavelet analysis theory, which can avoid the blindness in the design of the structure such as BP neural network. Secondly, WNN has stronger learning ability and higher precision.

### ***BP neural network***

BP neural network is a multi-layer feedforward neural network trained according to the error reverse propagation algorithm. It was proposed by scientists led by Rumelhart and Mc Clelland in 1986 (Rumelhart et al., 1986). As one of the widely used prediction models, BP neural network is also called back propagation neural network. Its main model structure is composed of input layer, hidden layer and output layer. The number of hidden layers can be adjusted according to the actual needs. One or more layers can be used to realize all connections between the neurons between the adjacent upper and



lower layers, but there is no connection between the neurons in each layer. The BP neural network has better description ability for complex nonlinear problems, that is, good nonlinear mapping ability, function approximation ability and fault tolerance.

### ***Cubic smoothing exponential model***

Exponential smoothing method is proposed by Brown (2013). Brown thinks that the situation of time series is stable or regular, so the time series can be postponed reasonably. He believes that the recent past situation, to a certain extent, will continue into the future, thus putting a greater weight on the recent data. In the calculation of exponential smoothing method, the key is the value of  $\alpha$ , but the value of  $\alpha$  is easy to be subject to subjective influence, so how to reasonably determine the value of  $\alpha$  is very important. Generally speaking, if the data fluctuates greatly, the value of  $\alpha$  should be larger. It can increase the impact of recent data on the prediction results. If the data fluctuates steadily, the  $\alpha$  value should be smaller. When the time series fluctuates, but the long-term trend changes little, we can choose a slightly larger  $\alpha$  value, which is usually between 0.1~0.4. When the time series fluctuates greatly, the long-term trend changes greatly. In order to make the prediction model more sensitive and keep up with the change of data, it is advisable to choose a larger  $\alpha$  value when the trend is obvious and rapid rise or fall, for example, the value can be chosen between 0.6~0.8, so as to make the prediction model more sensitive and able to keep up with the change of the data. The trial algorithm is usually used to determine the  $\alpha$  value in the model. First of all, according to the specific time series analysis, referring to the empirical judgment method, to roughly determine the rated value range, and then take different  $\alpha$  value for trial calculation, compare the prediction standard error under different  $\alpha$  value. The minimum error  $\alpha$  of prediction standard is chosen as the reasonable smoothing coefficient of the model. The basic formula for calculating the smoothing index is as follows.

$$Y_{t+1} = X_t + (1-\alpha)Y_t \quad (\text{Eq.10})$$

In the formula,  $X_t$  is the actual value of period  $t$ ,  $Y_t$  is the predicted value of phase  $t$ ;  $\alpha$  is an exponential smoothing coefficient.

The formula for calculating the cubic smoothing index is:

$$S_t^{(1)} = \alpha X_t + (1-\alpha)S_{t-1}^{(1)} \quad (\text{Eq.11})$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)} \quad (\text{Eq.12})$$

$$S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha)S_{t-1}^{(3)} \quad (\text{Eq.13})$$

$$a_t = 3S_t^{(1)} - 3S_t^{(2)} + 3S_t^{(3)} \quad (\text{Eq.14})$$

$$b_t = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}] \quad (\text{Eq.15})$$

$$c_t = \frac{\alpha}{2(1-\alpha)^2} (S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}) \quad (\text{Eq.16})$$

When the forecast period is  $T$  year and the base year is the  $t$  year. The prediction model of  $Y_{t+T}$  cubic exponential smoothing method is as follows:

$$Y_{t+T} = a_t + b_t T + c_t T^2 \quad (\text{Eq.17})$$

In the formula  $S_t^{(1)}$  is the single exponential smoothing value;  $S_t^{(2)}$  is the quadratic exponential smoothing value;  $S_t^{(3)}$  is the cubic exponential smoothing values,  $a_t$ ,  $b_t$ ,  $c_t$  are the smoothing factor. The cubic smoothing index prediction model, it has the advantage of requiring a small amount of data to predict the desired results, and its main theoretical basis is a kind of time developed on the basis of moving average method. The method of inter-sequence analysis and prediction is compatible with the whole average and moving average without abandoning the data of the past.

### ***Optimum architecture of model***

#### ***(1) ELM model***

Considering the output of ELM is influenced by weight and matrix, the prediction results are fluctuating. In order to make the prediction result more accurate, it is necessary to determine the optimal number of hidden layer nodes and neurons. Therefore, the normal mean square error (NMSE) is used as the criterion to select the optimal number of nodes and the number of neurons. When the smaller the value of NMSE, the higher the prediction accuracy of the model. Therefore, the ELM model need to determine number of neurons and input nodes to construct model. By repeated tests the trail number of neurons and the number of input nodes were 5; 10; 15; 20; 25; 30 and 2; 3; 4; 5; 6; 7; 8, respectively. The NMSE values of the output results of each experiment are calculated to compare the parameters, so as to obtain the most suitable number of neurons and input nodes. Its mathematical expression is:

$$NMSE = \frac{\sum_{j=1}^N (\hat{p}_j - p_j)^2}{\sum_{j=1}^N (\hat{p}_j - p_j)^2} \quad (\text{Eq.18})$$

In the formula, the  $\hat{p}_j$  represents predictive NDVI value; the  $p_j$  represents the actual NDVI value; N for valid number  $j = 1, 2, \dots, N$ .

According to *Table 4*, when the number of neurons is 25 and the number of input nodes is 6, the NMSE value can reach the minimum, which indicates that the prediction precision of ELM model is the best.

#### ***(2) WNN model***

In the process of WNN training, the number of hidden layers and the number of iterations are mainly constructed. The standard variance is used as the evaluation index

to determine the number of the best hidden layers and the number of iterations, and the predicted value is compared with the actual standard variance. When the error between the two is the smallest, the prediction model is the best. The number of neurons used in this paper by the actual tests chose 5; 10; 15; 20; 25; 30, and the number of iterations is 10; 30; 50; 70; 90 just showing in *Table 5*.

Through the experiments showing in *Table 5*, we can see that when the number of neurons is 10 and the number of iterations is 50, the WNN can achieve the best prediction effect.

**Table 4.** ELM experiment

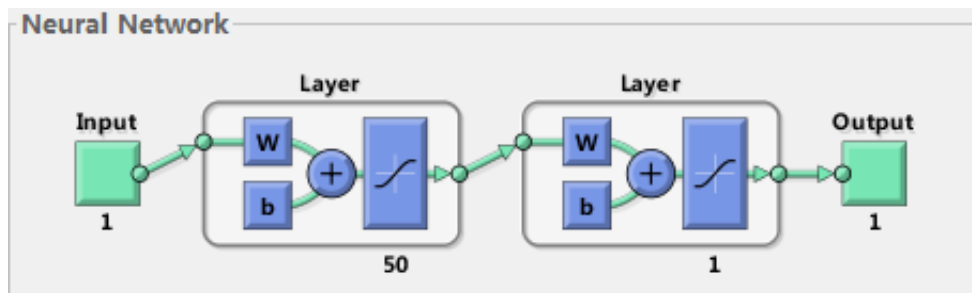
Trail	Cell	Input	NMSE	Trail	Cell	Input	NMSE
1	5	2	0.6613	22	20	5	0.4937
2	10	2	0.7067	23	25	5	0.6473
3	15	2	0.7528	24	30	5	0.6346
4	20	2	0.7476	25	5	6	0.5766
5	25	2	0.7759	26	10	6	0.8067
6	30	2	0.9280	27	15	6	0.3656
7	5	3	0.7740	28	20	6	0.3592
8	10	3	0.7723	29	25	6	0.3077
9	15	3	0.6767	30	30	6	0.1720
10	20	3	0.6107	31	5	7	1.0726
11	25	3	0.8229	32	10	7	0.5019
12	30	3	0.4532	33	15	7	0.5276
13	5	4	0.9184	34	20	7	0.4566
14	10	4	0.7812	35	25	7	0.486
15	15	4	0.5598	36	30	7	0.3515
16	20	4	0.7221	37	5	8	0.7423
17	25	4	1.0046	38	10	8	0.4724
18	30	4	1.3483	39	15	8	0.5642
19	5	5	0.9524	40	20	8	0.5394
20	10	5	0.5766	41	25	8	0.4017
21	15	5	0.6338	42	30	8	0.7695

### (3) BP neural network

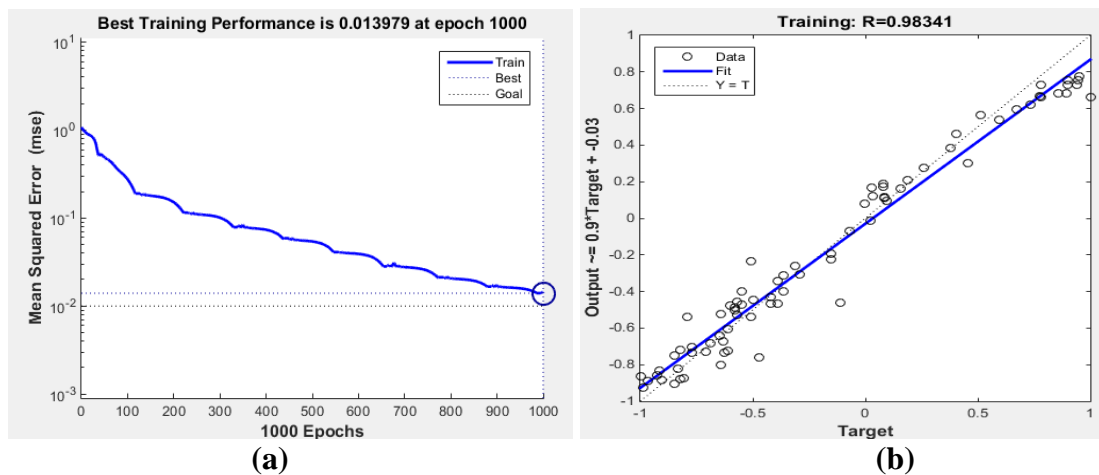
In the research process of this paper, when the double hidden layers algorithm model is constructed, the key parameters of the model need to determine the learning rate parameters, the number of training times, and the number of neurons in the double hidden layer. By repeating the test when the test parameters of learning rate parameter is set to 0.005, the precision requirement is set to 0.01, and the training times are 1000 meet the accuracy requirements. According to the characteristics of the neural network, the data type of this study belongs to the single-layer input and output type. Respectively, the input layer and output layer set to 1. In the experiment, when the double hidden layer numbers are set to 50 and 1 shown in *Figure 4*, the MSE value and the variance value shown in *Figure 5* are obtained, which meets the requirements of the test.

**Table 5.** WNN experiment

Trail	Cell	Iterations	$R^2$	Trail	Cell	Iterations	$R^2$
1	5	10	0.08028	19	5	70	0.1521
2	10	10	0.1669	20	10	70	0.1551
3	15	10	0.1648	21	15	70	0.1666
4	20	10	0.1631	22	20	70	0.1553
5	25	10	0.1599	23	25	70	0.1561
6	30	10	0.2094	24	30	70	0.1693
7	5	30	0.1455	25	5	90	0.1321
8	10	30	0.1431	26	10	90	0.1605
9	15	30	0.1562	27	15	90	0.1578
10	20	30	0.1649	28	20	90	0.1731
11	25	30	0.1627	29	25	90	0.1664
12	30	30	0.154	30	30	90	0.1621
13	5	50	0.1338	31	5	110	0.1596
14	10	50	0.156	32	10	110	0.152
15	15	50	0.1579	33	15	110	0.1564
16	20	50	0.1529	34	20	110	0.1577
17	25	50	0.1519	35	25	110	0.1541
18	30	50	0.1524	36	30	110	0.1671



**Figure 4.** Double implicit layers parameter



**Figure 5.** Training result diagram

(4) *Cubic smoothing exponential model*

In this paper, after the construction of the cubic smoothing index algorithm, the results of the model can be obtained by determining the smoothing index of  $\alpha$ . The extracted data show that the road domain NDVI values are fluctuating and the range of fluctuation is not very large. According to the characteristics of the algorithm, when the data series fluctuates, but the long-term trend changes little, we can choose a slightly larger  $\alpha$  value, which is usually between 0.1~0.4. Therefore, the smoothing coefficients  $\alpha$  are 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, respectively, and the corresponding NMSE values are calculated. The details are shown in *Table 6*.

When the smoothing coefficient  $\alpha$  is 0.35, the NMSE is 0.1597723, and the corresponding NMSE is larger when the smoothing coefficient is small or the smoothing coefficient is larger. Therefore, 0.35 is chosen as the best smoothing coefficient of the model.

**Table 6.** *Cubic smoothing index experiment*

Trail	$\alpha$	NMSE
1	0.05	0.866579159
2	0.1	0.688243365
3	0.15	0.523380124
4	0.2	0.379448459
5	0.25	0.263479288
6	0.3	0.18579078
7	0.35	0.159723238
8	0.4	0.199460299
9	0.45	0.318450584
10	0.5	0.52758426

**Application results of model**

The optimal prediction models are obtained after the correlation parameters are determined by the above four kinds of prediction models. There are different indicators used to evaluate the prediction effect of the models, the MAE; WMAE; RMSE and  $R^2$  (Goodness of Fit) of the test values and the real values are compared and analyzed to evaluate the prediction effect (Sun et al., 2013; Feng et al., 2015; Li et al., 2018; Yi et al., 2018).

MAE: it is defined as the average error of road domain NDVI prediction, and a smaller MAE value indicates that the prediction is more accurate. Its expression is as follows.

$$MAE = \frac{\sum_{j=1}^N | \hat{p}_j - p_j |}{N} \tag{Eq.19}$$

WMAE: it is the average relative error of road domain prediction, and the smaller average relative error represents more accurate prediction. Its expression is shown below.

$$WMAE = \frac{\sum_{j=1}^N |p_j^{\wedge} - p_j|}{\sum_{j=1}^N p_j} \quad (\text{Eq.20})$$

RMSE: indicates the fluctuation range between the predicted values and the true values error. The smaller RMSE error indicates that the model is more suitable for this kind of prediction. Its expression is shown below.

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (p_j^{\wedge} - p_j)^2}{N}} \quad (\text{Eq.21})$$

$R^2$  (Goodness of Fit): it is for the predicted value of road field, it is shown that the predicted values are in good agreement with the actual situation, and the expression is as follows.

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{j=1}^N |p_j^{\wedge} - p_j|^2}{\sum_{j=1}^N |p_j - \bar{p}|^2} \quad (\text{Eq.22})$$

Above the formula:  $\hat{p}_j$  represents predictive NDVI values;  $p_j$  represents the actual NDVI values;  $\bar{p}$  represents the average of the actual values, N for valid number  $j = 1, 2, \dots, N$ .

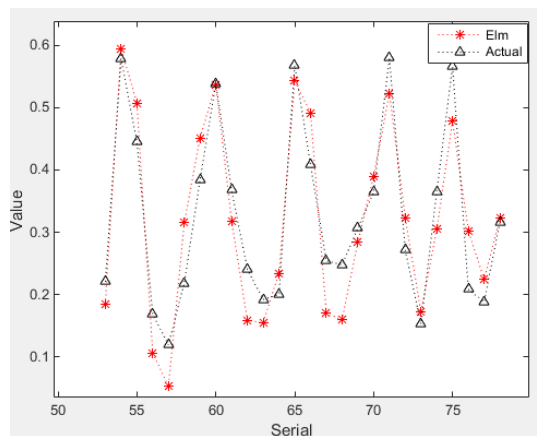
## Results

In this paper, the predicted and actual values of four machine learning models, such as ELM, WNN, double hidden layers BP neural network and cubic smoothing index, are programmed by MATLAB software. Using the above-mentioned extracted Pengjiazhai Town~Dayagen, Dayyagen~Dongxia Township, a total of 52 sets of data for machine learning training, another road Dongxia Township~Riyue Tibetan Township a total of 26 sets of data for prediction simulation test of the effect. Through training the output predictive values were outputted and compared with that of the real values, then the MAE; WMAE; RMSE;  $R^2$  are calculated and analyzed respectively. In addition, the analysis compares the prediction effects of the four machine models and which model is more suitable for the prediction of the normalized difference vegetation index at high-altitude.

### (1) Analysis of the results of ELM

After determining the optimal numbers of neurons in the ELM and the input nodes, apply the MAE, WMAE, RMSE;  $R^2$  shown above and four indexes are used to describe the prediction effect of it. The comparison and analysis of the actual values and the

predicted values are shown in *Figure 6*, and the concrete values of the calculation results are shown in *Table 7*.



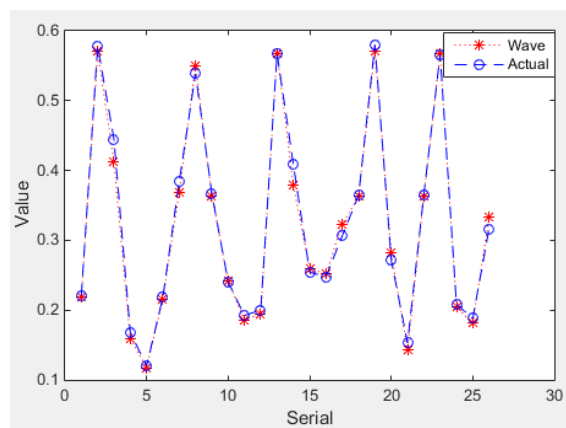
**Figure 6.** Numerical comparison diagram of ELM prediction

**Table 7.** ELM specific indicator numerical table

Cell	Input	MAE	WMAE	RMSE	R <sup>2</sup>
30	6	0.05196807	0.159489669	0.001743194	0.809441907

**(2) Analysis of the result of WNN**

After determining the optimal number of neurons and the number of iterations, the WNN uses two sets of data to learn it. At the same time, in order to prevent the data overfitting, the minimum error is set to 10<sup>-3</sup>. The prediction effect of WNN is compared as shown in *Figure 7*, and the calculation results are shown in *Table 8*.



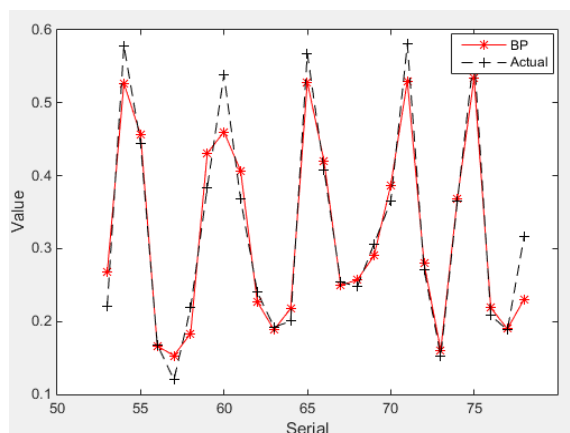
**Figure 7.** Numerical comparison diagram of wavelet neural network prediction

**Table 8.** Numerical table of specific indexes of WNN

Cell	Iterations	MAE	WMAE	RMSE	R <sup>2</sup>
10	50	0.010790846	0.033117032	0.000139242	0.844069851

### (3) Analysis of the results of BP neural network

Through the data training in the above process, the BP neural network determines the number of cycles and the hidden layer, and gets the comparison of the prediction effect of it as shown in *Figure 8*, and the concrete value of the calculation result is shown in *Table 9*.



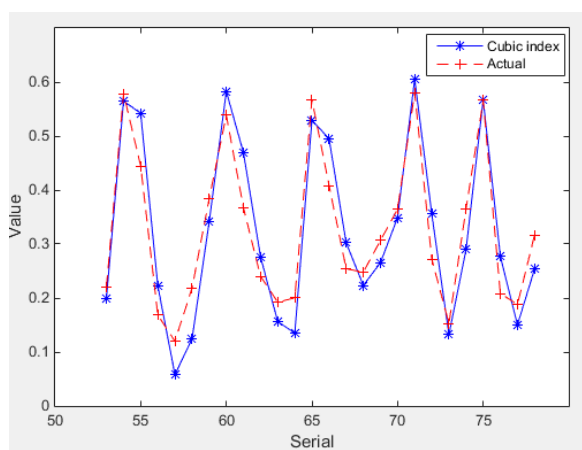
**Figure 8.** Numerical comparison diagram of BP neural network prediction

**Table 9.** Numerical table of specific indexes of BP neural network

The first cell	The second cell	Iterations	MAE	WMAE	RMSE	R <sup>2</sup>
50	1	1000	0.0332575	0.108618413	0.001927419	0.8728

### (4) Cubic smoothing index

It is known that the prediction effect can reach the best when the smoothing coefficient  $\alpha = 0.35$  of the cubic smoothing index, so the corresponding index between the predicted value and the actual value can be calculated. The prediction effects of it compared as shown in *Figure 9*, and the calculated results are shown in *Table 10*.



**Figure 9.** Numerical comparison diagram of cubic smoothing index prediction



**Table 10.** Numerical table of specific indexes of cubic smoothing index

$\alpha$	MAE	WMAE	RMSE	$R^2$
0.35	0.045676	0.149187191	0.00139507	0.79231912

When evaluating the prediction quality and applicability of machine learning models, the MAE, WMAE, RMSE, and  $R^2$  calculated above are usually employed. MAE and WMAE are utilized as the average error and the average relative error of the predicted value. The smaller the value, the higher the accuracy of the prediction. As the root mean square error of the predicted value, RMSE can reflect the fluctuation range between the predicted value and the actual value, and the smaller the value, the more suitable this kind of machine learning model is for this kind of research.  $R^2$  can reflect the agreement between the predicted value and the real value in the overall effect, and the larger the value, the better fit the prediction of the whole range. The results of the four indexes are shown in *Table 11*, and the comparison between the true values and the predicted values of the four models are shown in *Figure 10*, the concrete values of the prediction results of the four models are shown in *Table 12*.

## Discussion

Compared with the other three models, the MAE predicted by the ELM with a single hidden layer is the largest. The comparison from *Table 11* and the line chart shows that the average error of the model is 79.24% higher than that of the WMAE model. It is 36.0% higher than the BP neural network model and 12.11% higher than the cubic smoothing index model. In general, the superiority of ELM has been demonstrated in the previous literatures (Liu et al., 2018; Yi et al., 2018; Fijani et al., 2019). The comparison and analysis show that the average error of WNN is relatively small, which indicates that the accuracy of WNN is better. As for the mean relative error (WMAE) of the WNN, it is 79.24% higher than that of the ELM model, 69.51% higher than BP neural network, and 77.80% higher than cubic smoothing index. Compared with the WNN, the RMSE value of the ELM model and BP neural network model are significantly larger than it. The fluctuation range effect of WNN is 92.21% higher than that of ELM model, 92.78% higher than that of BP neural network, and 90.02% higher than cubic smoothing index. This shows that the fluctuation range between the predicted value and the actual value of the WNN is small, and there is no sudden change of the predicted value compared with the other three models, from the predicted comparison chart of the broken line fully proves it. According to some researches using wavelet transform to predict the results is more accurate (Feng et al., 2015; Kisi et al., 2018). In terms of the fit of the prediction effect, extreme learning machine's  $R^2$  reaches 0.809441907; BP neural network is 0.8728, WNN is 0.844069851; the cubic smoothing index is 0.79231912, and the overall fit effect of the BP neural network is 7.26% higher than that of the ELM model, 3.29% higher than the WNN, and 9.22% higher than the cubic smoothing index. Although overall fit of BP neural network is superior to that of the other three types, the improvement effect is not very obvious, which shows that the four kinds of models all have better prediction results in terms of predicted fit degree.

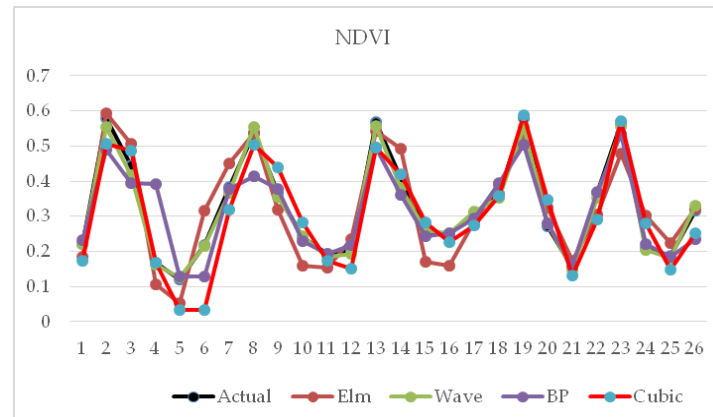


Figure 10. Comparison chart between predicted values and actual values

Table 11. Comparison of four models for predicting effect indexes

	ELM		Wavelet neural network		BP neural network		Cubic smoothing index
	Cell = 25	Input = 6	Cell = 10	Iterations = 50	Cell (50, 1)	Iterations = 1000	$\alpha = 0.35$
MAE	0.051968077		0.010790846		0.0332575		0.045676
WMAE	0.159489669		0.033117032		0.108618413		0.1491872
RMSE	0.001743194		0.000139242		0.001927419		0.0013951
$R^2$	0.809441907		0.844069851		0.8728		0.7923191

Table 12. Four models of predicting values

Actual	0.22077	0.578396	0.444658	0.168276	0.119978	0.218701	0.383716	0.538561	0.367532	0.240173	0.192425	0.200327	0.567645	0.408466	0.254257	0.247882	0.306538	0.364527	0.580294	0.271642	0.152848	0.365258	0.566043	0.208416	0.18822	0.316285
Elm	0.1847	0.5942	0.5071	0.1066	0.05276	0.3151	0.4514	0.537	0.3182	0.1587	0.1549	0.2342	0.5427	0.4916	0.1707	0.1597	0.2843	0.3899	0.5228	0.3234	0.1719	0.3049	0.478	0.3026	0.2242	0.3221
Wave	0.2208	0.5545	0.418	0.1607	0.1264	0.216	0.3658	0.5542	0.3551	0.2424	0.1853	0.1942	0.5567	0.381	0.2595	0.2518	0.3137	0.3533	0.5541	0.2796	0.1474	0.3537	0.557	0.2036	0.1807	0.3296
BP	0.2307	0.4896	0.3949	0.3914	0.1278	0.1293	0.3789	0.4138	0.379	0.2285	0.1917	0.2178	0.4953	0.3613	0.2445	0.2528	0.2928	0.3952	0.5034	0.2792	0.1601	0.3692	0.5334	0.2201	0.1858	0.2336
Cubic	0.1733	0.5063	0.4862	0.1666	0.0344	0.0344	0.3187	0.5047	0.4405	0.2826	0.1721	0.1495	0.4962	0.4194	0.2837	0.2268	0.2747	0.3576	0.5876	0.3465	0.132	0.2918	0.5704	0.2798	0.1467	0.252

## Conclusions

There is little research on predicting the normalized difference vegetation index (NDVI) of high-altitude highways by using relevant machine learning algorithm, but in the process of highways construction, especially, the highways operation and construction in the ecologically fragile areas at high-altitude will affect the surrounding ecological environment. Therefore, in this paper, four machine learning algorithms of ELM, WNN, BP neural network and cubic smoothing index are introduced to predict the normalized difference vegetation index (NDVI) of high-altitude highways to promote the sustainable development of the ecological fragile areas of the high-altitude environment and road construction.

Through research, it is found that WNN is more complex than ELM and cubic smoothing index in algorithm models, but compared with BP neural network, it is relatively simple and easy, an because of its multiple iterations and the continuous replacement of algorithm weight, the WNN is more complex than the ELM and the cubic smoothing index. Considering the weight effect of the whole data, the prediction precision is high. The prediction time of BP neural network is relatively longer when the condition of double hidden layer is adopted and the number of iterations is larger, so the efficiency of the neural network is greatly reduced. Under the condition of considering the weight of the data, the cubic smoothing index tries to have different smoothing coefficients, but it fails to solve the sudden change which is easy to produce a large value. The ELM as a single hidden layer is relatively simple, and its prediction effect is poor, the precision is not high, and it is easy to produce abrupt values. In the high-altitude areas, the road normalized difference vegetation index is more easily affected by the external factors, and the climate factors at high-altitude are more complex and changeable than in ordinary areas, and the data are characterized by fluctuating variability. Therefore, the model selection requires high precision, not easy to occur mutation, and can pay attention to the global factors and other characteristics. In this paper, compared with the other three models, WNN has higher overall prediction accuracy, closer to the actual value and no large deviation, thus showing a higher prediction level. Therefore, the WNN is more suitable for the prediction of road normalized difference vegetation index (NDVI) in high-altitude areas, which will helpful to the coordinated and sustainable development of ecological environment and highway construction in high-altitude areas. In the future research, we can pay attention to the method of comparison and selection is combined with the analysis and screening of the factors affecting the vegetation and the extent of roads, so as to find out the overall trend of the influence of highway on vegetation cover in road areas.

**Data availability.** The research data in this paper mainly downloaded the Landsat8 satellite data through the geospatial data cloud <http://www.gscloud.cn/>, and selected the available satellite data in the region where G214 is located through the geographic condition information, then made use of ENVI and GIS software to extract the data for this research as described in the article. The specific data extracted is shown in *Tables 1–3* in this paper.

**Funding.** The Project was Supported by Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2017JQ5122 and No. 2018JQ4009), and the Science and Technology Project of Transportation Department of Qinghai Province (Program No. 2016-03).

**Conflict of interests.** The authors declare no conflict of interests.

## REFERENCES

- [1] Abutaqiya, M. I. L., Sisco, C. J., Vargas, F. M. (2019): A linear extrapolation of normalized cohesive energy (LENCE) for fast and accurate prediction of the asphaltene onset pressure. – *Journal of Fluid Phase Equilib* 483: 52-69.
- [2] Brown, R. G. (2013): Exponential smoothing. – *Journal of Encyclopedia of Operations Research & Management Science* 11: 537-539.
- [3] Chen, Y. G. (2018): Prediction algorithm of PM2.5 mass concentration based on adaptive BP neural network. – *Journal of Computing* 100: 825-838.
- [4] Cocchi, G., Galli, L., Galvan, G., Sciandrone, M., Cantu, M., Tomaselli, G. (2018): Machine learning methods for short-term bid forecasting in the renewable energy market: A case study in Italy. – *Journal of Wind Energy* 21: 357-371.
- [5] Feng, X., Li, Q., Zhu, Y. J., Hou, J. X., Jin, L. Y., Wang, J. J. (2015): Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. – *Journal of Atmos Environ* 107: 118-128.
- [6] Fijani, E., Barzegar, R., Deo, R., Tziritis, E., Skordas, K. (2019): Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters. – *Journal of Sci. Total Environ* 648: 839-853.
- [7] Guo, X. J., Liu, S. F., Yang, Y. J. (2019): A prediction method for plasma concentration by using a nonlinear grey Bernoulli combined model based on a self-memory algorithm. – *Journal of Comput. Biol. Med* 105: 81-91.
- [8] Huang, G., Zhu, Q., Siew, C. (2006): Extreme learning machine: Theory and applications. – *Journal of Science Direct* 70: 489-501.
- [9] Kisi, O., Alizamir, M. (2018): Modelling reference evapotranspiration using a new wavelet conjunction heuristic method: Wavelet extreme learning machine vs wavelet neural networks. – *Journal of Agric for Meteorol* 263: 41-48.
- [10] Leng, X. Z., Wang, J. H., Ji, H. B., Wang, Q. G., Li, H. M., Qian, X., Li, F. Y., Yang, M. (2017): Prediction of size-fractionated airborne particle-bound metals using MLR, BP-ANN and SVM analyses. – *Journal of Chemosphere* 180: 513-522.
- [11] Li, M. L., Wang, W., De, G., Ji, X. H., Tan, Z. F. (2018): Forecasting carbon emissions related to energy consumption in Beijing-Tianjin-Hebei region based on grey prediction theory and extreme learning machine optimized by support vector machine algorithm. – *Journal of Energies* 11: 2475.
- [12] Lin, C. C., He, R. X., Liu, W. Y. (2018): Considering multiple factors to forecast CO2 emissions: a hybrid multivariable grey forecasting and genetic programming approach. – *Journal of Energies* 11: 3432.
- [13] Liu, H., Mi, X. W., Li, Y. F. (2018): An experimental investigation of three new hybrid wind speed forecasting models using multi-decomposing strategy and ELM algorithm. – *Journal of Renew. Energy* 123: 694-705.
- [14] Mishra, D., Goyal, P., Upadhyay, A. (2015): Artificial intelligence based approach to forecast PM2.5 during haze episodes: a case study of Delhi, India. – *Journal of Atmos Environ* 102: 239-248.
- [15] Moews, B., Herrmann, J. M., Ibikunle, G. (2019): Lagged correlation-based deep learning for directional trend change prediction in financial time series. – *Journal of Expert Syst* 120: 197-206.
- [16] Nian, C., Hu, K., Li, S. (2001): Wavelet neural network and its application. – *Journal of Chinese Stereology and Image Analysis* 6: 239-245.
- [17] Perez, P., Gramsch, E. (2016): Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes. – *Journal of Atmos Environ* 124: 22-27.
- [18] Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986): Learning representations by back-propagating errors. – *Journal of Nature* 323: 533-536.

- [19] Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W. D., Liu, S. W. (2013): Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in Northern California. – *Journal of Sci. Total Environ* 443: 93-103.
- [20] Tu, M. (2016): Comprehensive Evaluation of the Impact of Highway Construction on Ecological Environment in Southern Forest Region Based on GIS. – University Of Agriculture and Forestry in Fujian, China.
- [21] Wang, G. F. (2015): Dynamic Monitoring of vegetation coverage based on remote Sensing Technology. – *Journal of Shanxi agricultural science* 43: 592-595.
- [22] Xu, Y. Z., Du, P., Wang, J. Z. (2017): Research and application of a hybrid model based on dynamic fuzzy synthetic evaluation for establishing air quality forecasting and early warning system: a case study in China. – *Journal of Environ Pollu* 223: 435-448.
- [23] Yang, H. F., Chen, Y. P. P. (2019): Hybrid deep learning and empirical mode decomposition model for time series applications. – *Journal of Expert Syst* 120: 128-138.
- [24] Yi, H. S., Park, S., An, K. G., Kwak, K. C. (2018): Algal bloom prediction using extreme learning machine models at artificial weirs in the Nakdong River, Korea. – *Journal of Int. J. Environ. Res. Public Health* 15: 2078.
- [25] Zhang, M. Q., Jiang, X., Fang, Z., Chen, H. et al. (2019): High-order hidden Markov Model for trend prediction in financial time series. – *Journal of Physica A* 517: 1-12.
- [26] Zhao, L. M., Zhou, X. Y. (2018): Forecasting electricity demand using a new grey prediction model with smoothness operator. – *Journal of Symmetry-Basel* 10: 693.
- [27] Zhu, M. T., Shi, T., Chen, Y. et al. (2019): Prediction of fatty acid composition in camellia oil by 1H NMR combined with PLS regression. – *Journal of Food Chem* 279: 339-346.