Geomechanics and Engineering, *Vol. 27, No. 6* (*2021*) *551-560*
https://doi.org/10.12989/gae.2021.27.6.551

Stability evaluation model for loess deposits based on PCA-PNN

Guangkun Li, Maoxin Su, Yiguo Xue*, Qian Song, Daohong Qiu, Kang Fu and Peng Wang

Geotechnical and Structural Engineering Research Center, Shandong University, Ji'nan 250061, Shandong, China

(*Received January 22, 2021, Revised October 14, 2021, Accepted October 15, 2021*)

Abstract. Due to the low strength and high compressibility characteristics, the loess deposits tunnels are prone to large deformations and collapse. An accurate stability evaluation for loess deposits is of considerable significance in deformation control and safety work during tunnel construction. 37 groups of representative data based on real loess deposits cases were adopted to establish the stability evaluation model for the tunnel project in Yan'an, China. Physical and mechanical indices, including water content, cohesion, internal friction angle, elastic modulus, and poisson ratio are selected as index system on the stability level of loess. The data set is randomly divided into 80% as the training set and 20% as the test set. Firstly, principal component analysis (PCA) is used to convert the five index system to three linearly independent principal components X1, X2 and X3. Then, the principal components were used as input vectors for probabilistic neural network (PNN) to map the nonlinear relationship between the index system and stability level of loess. Furthermore, Leave-One-Out cross validation was applied for the training set to find the suitable smoothing factor. At last, the established model with the target smoothing factor 0.04 was applied for the test set, and a 100% prediction accuracy rate was obtained. This intelligent classification method for loess deposits can be easily conducted, which has wide potential applications in evaluating loess deposits.

Keywords: intelligent classification; loess deposits; principal component analysis; probabilistic neural network; stability evaluation

1. Introduction

As a result of the low strength, high compressibility characteristics of the loess deposits, tunnels are prone to large deformations and collapse during excavation (Feda 1988). If the characteristics of loess are ignored, it will often bring severe losses and damage to the project. Being able to assess the stability of loess deposits is essential for deformation control and decision-making. Thus, establishing a stability evaluation model for loess deposits around tunnels based on physical and mechanical indices is essential for guiding construction (Xu *et al.* 2021, Xue *et al.* 2018a, Zhang *et al.* 2019).

Loess refers to the porous yellow silty soil with columnar joints formed under dry weather conditions, formed in the Quaternary period about 2 million years. Once in contact with water, the bonding material of the loess will break and the strength will drop sharply (Graziani and Boldini 2012, Cui *et al*. 2021). Loess is distributed on all continents in the northern hemisphere, with the loess in the north of China being the most typical. The Loess Plateau in northwest China is the largest loess plateau in the world. The nature of the loess is relatively loose and porous, generally light yellow, grey-yellow or yellow-brown, with large visible pores and vertical joints (Tuo *et al.* 2017, Yates *et al.* 2018). Loess is very easy to seep water and has many soluble substances. It is not only prone to be eroded by running water to form valleys, but cause subsidence and collapse. The loess are not tightly bound, and the porosity is generally particles are not tightly bound, and the porosity is generally 40% to 50%. Loess deposits have a metastable structure, composed of loose particles, a large number of macroscopic pores (Ng *et al.* 2012, Ziegel 2005). It is characterized by low shear strength and large compressibility, with relatively small changes in stress (Kruse et al. 2007, Li 2018, Rogers *et al.* 1994). According to the previous studies, water content (w) could affect the mechanical and deformation characteristics of loess as a major influencing factor (Ren *et al.* 2015). Specifically, the higher the water content, the smaller the matrix suction of the unsaturated loess, the greater the force acting on the tunnel support, and the larger the radius of the plastic zone (Sharifzadeh *et al.* 2013, Burland 1990). Moreover, the shear strength of loess is subject to cohesion (c) and internal friction angle (φ) in the way of the stability of surrounding loess (Zhang *et al.* 2019, Ding *et a*l. 2019). Not only these indices, but Elastic modulus (E) and Poisson ratio (μ) could be considered as evaluation indicators for the stability of loess deposits (Feng *et al.* 2015, Grzymala and Ziarko 2000, Xue *et al.* 2018b). The actual stability level of loess can be determined by related studies and evaluation standards (Feng *et al.* 2015, Zhao *et al.* 2011). Engineering experience and engineering analogy are the primary manners for the stability assessment (Fernandez-Delgado *et al.* 2014) which have high randomness and uncertainty. Various factors have a comprehensive influence on the stability of loess with a nonlinear relationship (Jiang *et al*. 2019). Nonlinear methods are widely used in engineering applications (Faradonbeh and Taheri 2019, Fatemi *et al.* 2018, Lu *et al.* 2017, Pawlak 1998, Xue *et al.* 2018b). However, there are

*Corresponding author, Professor
E-mail: xieagle@sdu.edu.cn

ISSN: 2005-307X (Print), 2092-6219 (Online)

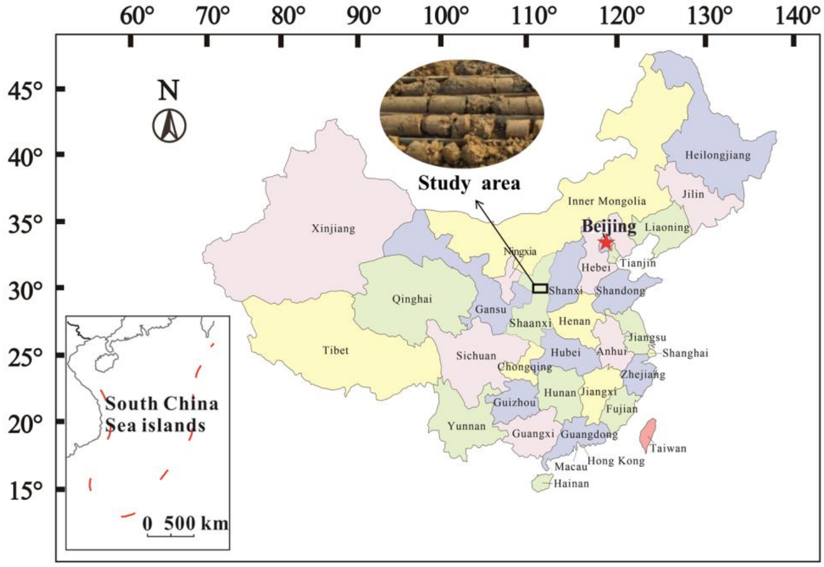

Fig. 1 Photograph showing the study area

relatively few applications in the field of loess deposits evaluation.

A new stability evaluation model for loess deposits around tunnels was proposed in this study based on PCA and PNN. Above all, the PCA was implemented to reduce dimensions of the original impact indexes. Then Leave-One-Out (LOO) cross validation was applied to search for the opportune smoothing factor of PNN, and the model was constructed. Finally, the trained model was verified by the test set and satisfactory results have been achieved.

2. Overview of the project

The Menghua Railway is the only national trunk heavy-duty freight channel railway built by China in the era of high-speed railways. Its total length is about 1800 kilometers, crossing seven provinces and regions, with a designed transportation capacity of 200 million tons per year.

The tunnel project in Yan'an, China (Fig. 1) was examined in both laboratory and field to establish the stability evaluation model for loess deposits in this study. And 37 groups of physical and mechanical indices of loess with corresponding stability level were collected. The Yan'an Tunnel is part of the Menghua Railway located on the Loess Plateau in Shaanxi. The tunnel area is a Mesozoic basin with multiple depressions superimposed after numerous cycles of structural development. The structural trace of this section is relatively weak, folds and faults are not developed, the terrain is high in the East and bottom in the west, and generally inclines to the monoclinic structure in the northwest. The terrain is affected by the uplift of the platform and the erosion of the source of water flow in the Loess Plateau. The undercutting effect is noticeable. The "V" shaped gullies are developed and distributed in a dendritic manner. The terrain is complex, which is a typical erosive gully landform type of Loess Plateau. The topography can be divided into two geomorphic units: loess hills and loess gullies. The height difference is generally 50-200 meters, the ground elevation is 1213-1469 meters, and the natural slope of the slope is 30-70°. The stratum in the area of the tunnel is mainly Quaternary Upper Pleistocene eolian (Q3eol) sandy new loess and clay new loess; Middle Pleistocene pluvial (Q2pl) clay old loess and fine round gravel soil; Jurassic Middle (J2) sandstone and mudstone. The surface water in the tunnel site is mainly atmospheric precipitation and bedrock fissure water, and groundwater is primarily divided into Quaternary pore water and bedrock fissure water. The study area belongs to the cold area, with a standard freezing depth of 1.0 m.

The New Austrian Tunnelling Method (NATM) and the sequential excavation method (SEM) with pipe roof supports, steel arches and hanging nets were adopted to ensure construction quality and safety. Several excavation methods, including the central diaphragm method, three-bench seven-step excavation method, three-bench method and two-bench method were also adopted according to the situation.

3. Establishing a stability evaluation model for loess deposits

3.1 Dataset establishment

The loess tunnel is prone to disasters such as large deformation or collapse as the result of its porous and fragile characteristics, which will adversely affect the tunnel construction. The stability of loess deposits is related to various factors, such as geological conditions and construction operations. Establishing the nonlinear relationship between the stability of loess deposits and different factors is critical to control deformation. 37 groups of representative data were collected to establish the stability evaluation model for loess deposits, including

Table 1 Statistical data from borehole samples in Yan'an tunnel (Zhang *et al.* 2019)

No.	*w* (%)	*c* (kPa)	*φ* (°)	*E* (MPa)	*μ*	Level
1	10.4	51	24	256	0.29	IV_a
2	10.1	53	26	273	0.29	IV_a
3	9.8	49	25	314	0.3	IV_a
4	10.9	48	26	309	0.28	IV_a
5	10.8	12	26	254	0.3	IV_a
6	16.3	47	29	278	0.29	IV_a
7	16.3	32	24	178	0.33	IV_b
8	15.4	40	23	223	0.32	IV_b
9	11.5	33	21	168	0.33	IV_b
10	15.2	35	26	196	0.33	IV_b
11	15.4	48	28	313	0.32	IV_b
12	16.8	34	25	199	0.29	IV_b
13	16.8	23	20	76	0.33	V_a
14	12	22	21	67	0.33	V_a
15	13.1	27	19	54	0.32	V_a
16	17.3	24	22	62	0.32	V_a
17	16.8	22	21	77	0.33	V_a
18	13.1	19	20	72	0.32	V_a
19	18	19	16	89	0.33	V_b
20	17.6	7	16	156	0.33	V_b
21	17.7	14	18	148	0.33	V_b
22	17.9	28	15	149	0.32	V_b
23	17.7	16	18	123	0.33	V_b
24	17.6	11	17	109	0.32	V_b
25	18	26	16	93	0.33	V_b
26	18.2	23	17	129	0.32	V_b
27	18.2	9	14	34	0.42	VI
28	18.3	11	13	42	0.41	VI
29	18.8	13	13	42	0.42	VI
30	18.9	13	13	41	0.36	VI
31#	10.1	50	24	267	0.29	IV_a
32#	14	32	25	180	0.32	IV_b
33#	13.4	37	25	188	0.32	IV_b
34#	17.6	25	20	72	0.3	V_a
35#	17.9	8	17	112	0.32	V_b
36#	17.7	1.41	16	117	0.33	V_b
37#	18.3	14	14	39	0.41	VI

Note: w- Water content; c- Cohesion; φ- Internal friction angle; E- Elastic modulus; μ- Poisson ratio; Samples of no.1- No.30 are taken as the original training set, the remaining with # mark are taken as the original test set

water content, cohesion, internal friction angle, elastic modulus, and Poisson ratio (Feng *et al.* 2015, Ren *et al.* 2015, Xue *et al.* 2018b, Zhang *et al.*2019, Zhao *et al.* 2011, Zhou *et al.* 2020). The corresponding stability level of loess is determined as five levels: IV_a, IV_b, V_a, V_b, and VI by the actual situation of surrounding rock revealed by tunnel excavation and previous studies (Feng *et al.* 2015, Zhao *et al.*2011) (Table 1).

3.2 Dimensionality reduction with PCA

The five physical and mechanical indices in Table 1 are proved to affect the stability of loess deposits (Zhang *et al.*2019). The values of correlation coefficients (Joshaghani *et al.* 2018) among the five parameters are calculated, and the degree of linear correlation between variables is visualized with elliptical color patches (Fig. 2). A strong linear correlation exceeding 0.5 or even higher is noted between the parameters. To screen out the independent features from the redundant raw data for the sake of calculation, it is of great importance to reduce the dimension of primal data (Pu *et al.* 2019).

Considering that there are five indices selected in this paper, the PCA method (Hotelling 1933) is used to preprocess the physical and mechanical indices of loess deposits, which can not only eliminate the correlation

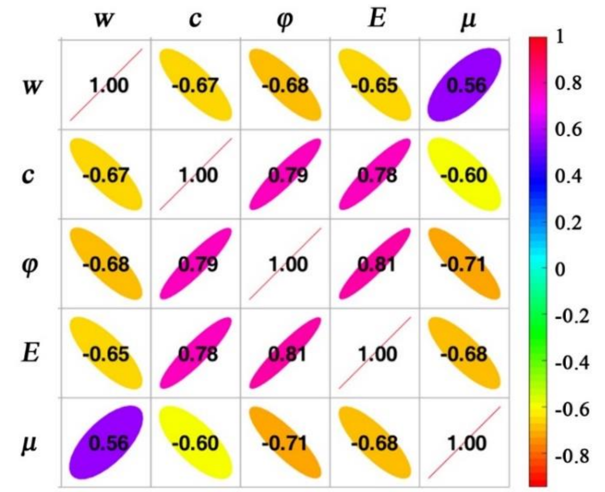

Fig. 2 Correlation coefficient graph among the five indicators

Note: A flatter ellipse indicates that the absolute value of the correlation coefficient between variables is closer to 1, and a rounder one is closer to 0. If the long-axis direction of the ellipse is from the lower left to upper right, the variables are positively correlated and vice versa (Xie 2010)

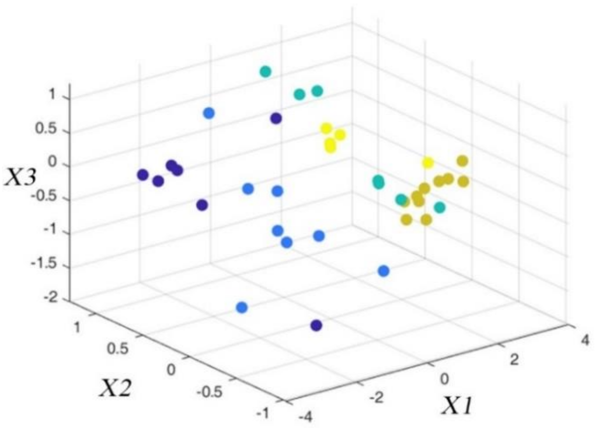

Fig. 3 Data visualization after dimensionality reduction with PCA

between indexes but reduce the dimensionality. The core of the PCA is to transform the original indicators into new, independent, and comprehensive indicators containing most of the original information so as to achieve the purpose of dimensionality reduction (Ren and Yu 2011, Wang *et al.* 2020). The specific process is as follows:

Step 1: Standardize the raw data of indices of loess deposits as X*.

Step 2: Calculate the Pearson correlation coefficient matrix R (Fig. 2) among the five indices.

Step 3: Calculate the eigenvalues λ_s and eigenvectors P_s of the correlation matrix R.

Step 4: Determine the top k principal components, whose cumulative contribution rate reaches 85% to 95%.

Step 5: Extract the first k principal components from the original indicators in the following manner

$$X = X * Ps(s = 1, 2, \ldots k) \tag{1}$$

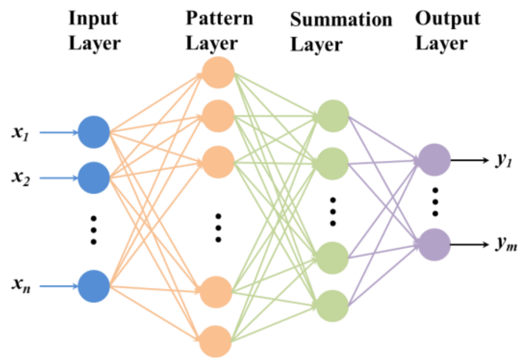

Fig. 4 Structure of the PNN network

In this work, the original five features were converted to the three principal components (X1, X2 and X3) with a cumulative contribution rate of 92.18%. Fig. 3 shows the scatter plot for the three-dimensional data after dimensionality reduction.

3.3 PNN intelligent response surface

On account of the complexity of the loess deposits deformation mechanism, the intelligent response surface method of PNN was adopted to establish the nonlinear mapping relationship between the three principal components (derived by the five physical and mechanical indices of loess deposits) and the stability level of loess deposits.

PNN has a wide range of applications in the classification process, which was proposed by Dr. Specht in 1989 (Specht 1990). PNN can be regarded as a radial basis function neural network, which combines density function estimation and Bayesian decision theory based on the RBF network. Under some conditions that are easy to be satisfied, and the discriminant boundary realized by PNN is used to gradually approach the optimal Bayesian decision surface (Chen 2013). PNN is a forward propagation network, which generally has four layers: input layer, pattern layer, summation layer and output layer (Fig. 4). The input layer receives the training sample data, and the number of neurons is equal to the length of the input vector. The input vector is weighted and passed to the pattern layer as the formula

$$z_i = xw_i \tag{2}$$

Where x is the input vector, and w_i is the weight vector.

The number of neurons in the pattern layer is the same as the number of input training samples. And the output of the *j-th* neuron of the type *i-th* pattern in the pattern layer can be calculated as

$$\varphi_{ij} = \frac{1}{(2\pi)^{0.5d}\,\delta^{d}} exp[-\frac{(x - x_{ij})^{\mathrm{T}}(x - x_{ij})}{2\delta^{2}} \tag{3}$$

Where d is the dimension of the metric space, and δ is the smoothing factor. x_{ij} is the *j-th* center of the class i sample. φ_{ij} is the output value of the *j-th* neuron of the type *i-th* pattern in the pattern layer.

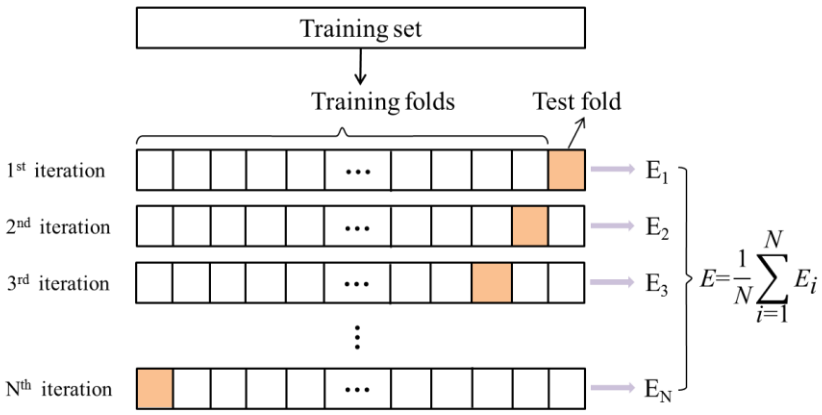

Fig. 5 A sketch map of LOO cross validation

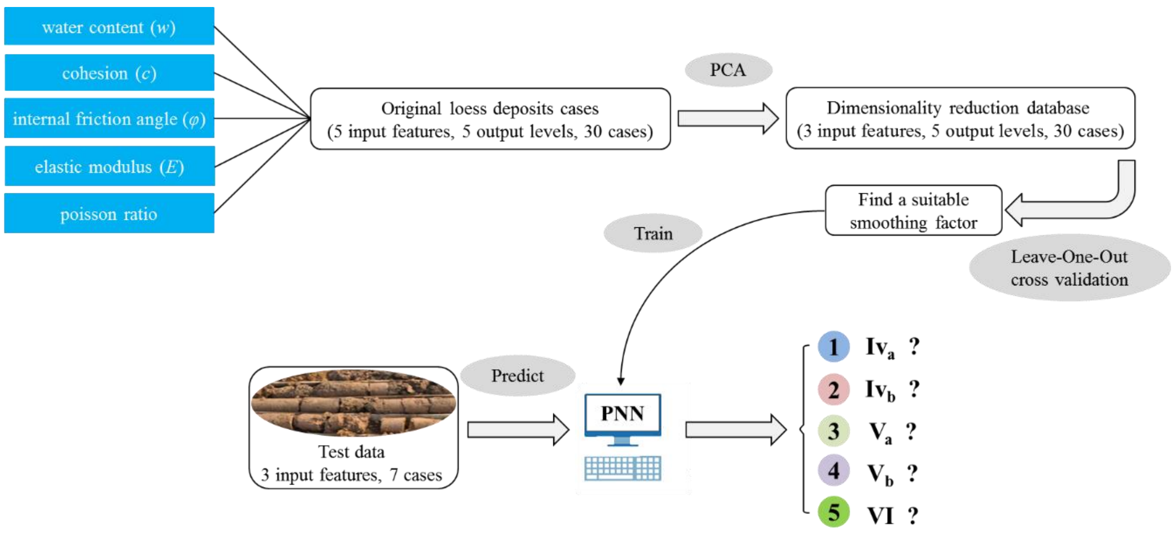

Fig. 6 Flowchart of this study (Pu *et al*. 2019)

The main function of the summation layer is a linear summation and weighted average. The number of neurons in the output layer is the same as the total number of training sample patterns, and each neuron corresponds to a pattern. It is used to distinguish the critical value for the output of the accumulation layer, the output of the neuron with the maximum a posteriori probability density in the output layer is 1, and the rest is 0. The 3-D uncorrelated variables produced by PCA can meet the requirements when the activation function of a Gaussian function is adopted in PNN (Cochrane and Weatherall 1972).

3.4 Construction of a stability evaluation model for loess deposits

Judging the stability level of loess deposits is a nonlinear classification problem (Zhang *et al.* 2019) according to the physical and mechanical indices including water content, cohesion, internal friction angle, elastic modulus, and poisson ratio, as well as the three principal components derived by the five physical and mechanical indices of loess deposits. PNN has been widely used in different pattern recognition fields in recent years because of its simple learning process and convergence to Bayesian optimal solution (Chen *et al.* 2019, Din *et al.* 2020, Zhang *et al.* 2020).

In this work, the PNN intelligent response surface was adopted to establish the indirect relationship between the physical and mechanical indices and the stability level of loess. The original data is randomly disrupted and divided into 80% for the training set and the remaining 20% for the test set (Table 1). To simplify the calculation process, the dimensionality reduction is implemented for the original indices with PCA. X1, X2 and X3 form the input vectors of the PNN network after dimensionality reduction from the original indices. In addition, the output vector "1" represents expected stability level IV_a of loess deposits; the output vector "2" represents expected stability level IV_b of loess deposits; the output vector "3" represents expected stability level V_a of loess deposits; the output vector "4" represents expected stability level V_b of loess deposits; the output vector "5" represents expected stability level VI of loess deposits. The smaller the output vector, the more stable the loess deposits (Zhang *et al.* 2019).

In a PNN model, the selection of smoothing factor δ plays a crucial role in network performance. If the model learner works “too well” for the training set, it is more likely that some features of the training samples are regarded as the general nature of the model, which will cause the model to over-fit on the training samples and reduce the learner's generalization performance (Xue *et al.* 2020). To find a suitable smoothing factor "δ", 100 groups of smoothing factor values uniformly distributed in the interval of [0.01, 1] are selected to train the network. Considering the limited representative data in this article, LOO cross validation was adopted to find a model with better generalization ability and test the model accuracy. In LOO cross validation, assuming that there are n samples, each sample is taken as a test sample, and the other N-1 samples are taken as training samples. In this way, the model performs N times of training and testing to obtain N test results. The average prediction accuracy of these N accuracy test results is employed to measure the performance of the model. The flowchart for LOO cross validation is shown in Fig. 5. Finally, the trained model with a suitable smoothing factor δ was employed to predict the stability level of loess of the test set. Fig. 6 provides a flowchart for this study. And the platform adopted to develop the stability evaluation model for loess deposits with the method of PCA-PNN was MATLAB software.

4. Results

The original data in Table 1 are implemented for dimensionality reduction, and the first 30 sets of data were used to train the model. The vectors X1, X2 and X3 after dimension reduction constitute the input vectors of the PNN model. Before training the model, LOO cross validation was employed with every smoothing factor δ to find the suitable value of δ. The training set has 30 samples, then the 30 sets of results of predicted accuracy were averaged to produce a single accuracy for the model with each smoothing factor δ. Obviously, the higher the predicted accuracy is, the better performance of the model shows. 100 groups uniform value of δ were verified from 0.01 to 1. The effect of the smoothing factor δ on the prediction accuracy of the cross-validated training set, the cross-validated test set, and the entire training set is displayed in Fig. 7.

It can be seen from Fig. 7 that in the interval of [0.01, 1], the prediction accuracy of the model generally decreases as the value of δ increases. In some ranges, the prediction accuracy of the cross-validated training set, the cross-validated test set, and the entire training set does not change with δ, indicating that the PNN is not sensitive to the δ value changes in these ranges. When the value of δ is 0.02 or 0.04, the prediction accuracy of the cross-validated training set, the cross-validated test set, and the entire training set reaches the maximum at the same time, which is 100%, 87% and 100%, respectively. Considering that the δ value should not be too small (Adams 1971), the δ value 0.04 is selected as the optimal smoothing factor for the model. Therefore, the stability evaluation model for loess deposits was built.

The 31st to 37th samples in Table 1 were used to test the trained model after the same PCA dimensionality reduction. MATLAB software was still adopted to run the model to evaluate the stability for loess deposits around tunnels.

Prediction results of training samples and test samples are exhibited in Figs. 8 and 9. It can be seen that the established model with an optimal smoothing factor 0.04 possesses a 100% accuracy for the training samples (Fig. 8). And the predicted results of test samples (Fig. 9) exactly match the actual levels of loess deposits with a high-resolution. The convergence speed of PNN is fast with an ideal classification effect for stability evaluation for loess

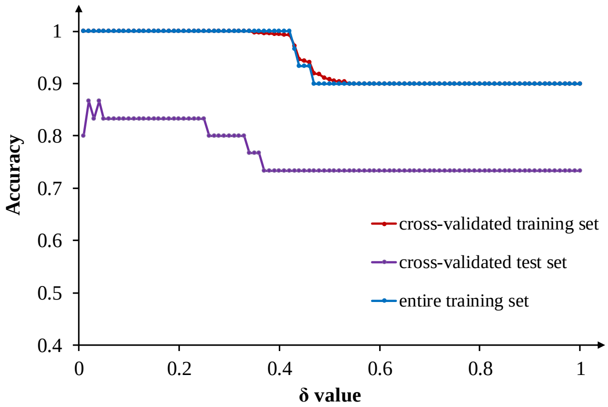

Fig. 7 Model prediction accuracy varies with smoothing factor "δ"

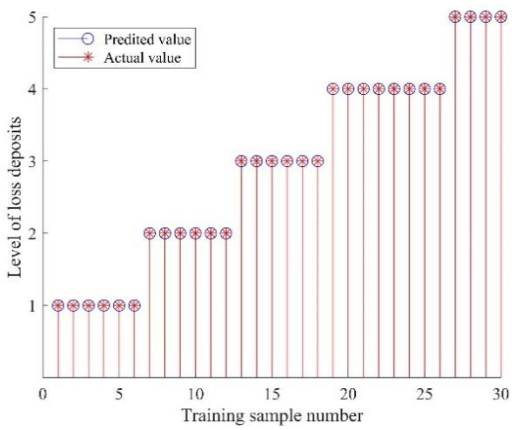

Fig. 8 Prediction results of training samples

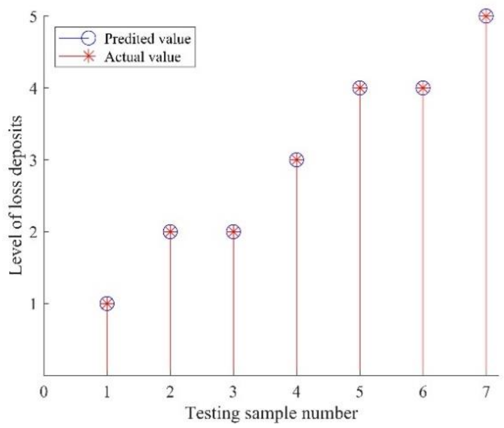

Fig. 9 Prediction results of training samples

deposits. The above results show that the stability evaluation model for loess deposits based on PCA-PNN proposed in this paper is reasonable and feasible. This prediction model is easy to implement, and does not entirely depend on the specific mechanism of loess deformation.

5. Discussion

Based on previous studies, physical and mechanical indices, including water content, cohesion, internal friction angle, elastic modulus, and poisson ratio are selected as influencing factors on the stability level of loess deposits. PCA uses the idea of dimensionality reduction to convert multiple indices into several comprehensive indices under the premise of losing little information. Each principal component is a linear combination of the original variables, and the principal components are not related to each other, making the principal components have some superior performance than the original variables. PNN model achieved satisfactory results in predicting stability level of loess deposits through the dimensionality reduction indices from the original five physical and mechanical indices. The prediction results of the model are consistent with the real cases, and the performance of the model is similar to that of the previous study (Zhang *et al.* 2019) with a more convenient calculation procedure. Although the PNN model has obtained a satisfactory prediction effect with PCA dimensionality reduction, some improvements should be emphasized in future researches.

The stability of loess deposits is affected by many other influencing factors in addition to the five indicators selected in this article, such as excavation face form, excavation method, support conditions, construction arrangement and so on. The selection and collection of more indicators are conducive to a more accurate prediction of loess stability. And measuring the weight of the physical and mechanical indices on loess stability is more helpful for people to understand the effect of different factors on loess stability. For the PNN network, setting off the smoothing factor parameter directly affects the learning effect. In this work, only 30 groups of representative data have been adopted to train for the nonlinear model, which is relatively small compared to other machine learning methods. The smoothing factor δ is determined as 0.04 for the evaluation model, which is applied to the test set and a relatively good evaluation result is obtained. It can be seen from Fig. 7 that when the value of δ is 0.02 or 0.04, the model can achieve high enough prediction accuracy of the cross-validated training set, the cross-validated test set, and the entire training set. There is also a polyline mutation between 0.02 and 0.04, explaining that the model is more sensitive to the δ value in this interval. More representative sample data can contribute to improving the generalization ability of the model with a more stable and optimal smoothing factor.

6. Conclusions

A new stability evaluation model for loess deposits based on PCA-PNN was proposed. 37 groups of representative data of borehole samples were collected from the tunnel project in Yan'an, China, to study the stability of loess deposits.

PNN model was adopted for establishing the intelligent response surface between the stability level and the selected influencing factors of loess deposits due to its simple operation. Considering the correlation between indicators and the PNN's requirements for the independence of indicators, PCA was used to implement dimensionality reduction before training. Then three linearly independent principal components X1, X2 and X3 were obtained and used as input vectors for PNN. Approximately 80% of the datasets after dimensionality reduction was taken to train the PNN model. The optimal smoothing factor in PNN model is determined by LOO cross validation as 0.04. Then the model is built with good generalization ability for the collected datasets. The predicted results for the test set could completely match real stability levels of loess deposits, which proved the feasibility of this model.

This article proposed a new model evaluation method for loess stability, which is easy to operate and has a significant reference value for evaluating the stability of loess deposits in similar projects. And the evaluation system should be continuously improved and checked with more engineering cases.

Acknowledgments

Research in this paper is supported by the National Natural Science Foundation of China (grant numbers 41877239, 51379112, 51422904, 40902084 and 41772298), and Fundamental Research Fund for the Central Universities (grant number 2018JC044), and Shandong Provincial Natural Science Foundation (grant number JQ201513).

References

Adams, M. (1971), "The single woman in today's society: A reappraisal", *Am. J. Orthopsychiatry*, **41**, 776-786. https://doi.org /10.1111/j.1939-0025.1971.tb00741.x.

Burland, J.B. (1990), "On the compressibility and shear strength of natural clay, Rankine lecture", *Geotechnique*, **40**(3), 329-378. https://doi.org/10.1680/geot.1990.40.3.329.

Chen, M. (2013), "Matlab neural network principles and detailed examples", *Tsinghua University Press*, Beijing, China

Chen, N., Sun, F.C., Ding, L.G. and Wang, H.Q. (2019), "An adaptive pnn-ds approach to classification using multi-sensor information fusion", *Neural Comput. Appl.*, **31**, 693-705. https://doi.org/10.1007/s00521-008-0221-3.

Cochrane, P. and Weatherall, D.J. (1972), "The variation of neutrophil alkaline phosphatase during the menstrual cycle", *J. Obstet. Gynaecol. Br. Commonw.*, **79**, 1002-1008, https://doi.org/ 10.1111/j.1471-0528.1972.tb11878.x.

Cui, G.Y., Ma, J.F. and Wang, D.Y. (2021), "A large 3D laboratory test on the deformation characteristic of shallow loess tunnel under different plastic states", *Bull. Eng. Geology Environ.*, **80**, 7577-7590

Din, N., Zhang, H. and Saeed, W. (2020), "Porosity prediction from model-based seismic inversion by using probabilistic neural network (pnn) in mehar block, pakistan", *Episodes*, **43**(4), 935-946. https://doi.org/10.18814/epiiugs/2020/020055.

Ding, H.H., Wu, Q., Zhao, D.K., Mu W.P. and Yu, S. (2019), "Risk assessment of karst collapse using an integrated fuzzy analytic hierarchy process and grey relational analysis model", *Geomech. Eng.*, **18**(5), 515-525, http://doi.org/10.12989/gae.2019.18.5.515.

Faradonbeh, R.S. and Taheri, A. (2019), "Long-term prediction of rockburst hazard in deep underground openings using three robust data mining techniques", *Eng. with Comput.*, **35**, 659-675, https://doi.org/ 10.1007/s00366-018-0624-4.

Fatemi, S.A., Ahmadi, M. and Rostami, J. (2018), "Evaluation of tbm performance prediction models and sensitivity analysis of input parameters", *Bull. Eng. Geology Environ.*, **77**, 501-513. https://doi.org/ 10.1007/s10064-016-0967-2.

Feda, J. (1988), "Collapse of loess upon wetting", *Eng. Geology*, **25**, 263-269. https://doi.org/ 10.1016/0013-7952(88)90031-2.

Feng, S.J., Du, F.L., Shi, Z.M., Shui, W.H. and Tan, K. (2015), "Field study on the reinforcement of collapsible loess using dynamic compaction", *Eng. Geology*, **185**, 105-115, https://doi.org/ 10.1016/j.enggeo.2014.12.006.

Fernandez-Delgado, M., Cernadas, E., Barro, S. and Amorim, D. (2014), "Do we need hundreds of classifiers to solve real world classification problems?", *J. Machine Learning Res.*, **15**, 3133-3181

Graziani, A. and Boldini, D. (2012), "Remarks on axisymmetric modeling of deep tunnels in argillaceous formations. I: Plastic clays", *Tunn. Undergr. Sp. Tech.*, **28**, 70-79. https://doi.org/10.1016/j.tust.2011.09.006.

Grzymala-Busse, J.W. and Ziarko, W. (2000), "Data mining and rough set theory", *Commun. ACM*, **43**(4), 108-109. https://doi.org/10.1145/332051.332082.

Hotelling, H. (1933), "Analysis of a complex of statistical variables into principal components", *J. Educational Psychology*, **24**, 417-441. https://doi.org/ 10.1037/h0071325.

Jiang, C., Zhang, H.Y., Zhang, Z.D. and Wang, D.W. (2019), "Model-based assessment soil loss by wind and water erosion in China's Loess Plateau: Dynamic change, conservation effectiveness, and strategies for sustainable restoration", *Global Planetary Change*, **172**, 396-413. https://doi.org/10.1016/j.gloplacha.2018.11.002.

Joshaghani, A., Balapour, M. and Ramezanianpour, A.A. (2018), "Effect of controlled environmental conditions on mechanical, microstructural and durability properties of cement mortar", *Constr. Build. Mater.*, **164**, 134-149. https://doi.org/10.1016/j.conbuildmat.2017.12.206.

Kruse, G., Dijkstra, T. and Schokking, F. (2007), "Effects of soil structure on soil behaviour: Illustrated with loess, glacially loaded clay and simulated flaser bedding examples", *Eng. Geology*, **91**, 34-45. https://doi.org/10.1016/j.enggeo.2006.12.011.

Li, Y.R. (2018), "A review of shear and tensile strengths of the malan loess in china", *Eng. Geology*, **236**, 4-10. https://doi.org/10.1016/j.enggeo.2017.02.023.

Lu, Q., Xiao, Z.P., Ji, J., Zheng, J. and Shang, Y.Q. (2017), "Moving least squares method for reliability assessment of rock tunnel excavation considering ground-support interaction", *Comput. Geotech.*, **84**, 88-100, https://doi.org/ 10.1016/j.compgeo.2016.11.019.

Ng, C.W.W., Hong, Y., Liu, G.B. and Liu, T. (2012), "Ground deformations and soil-structure interaction of a multi-propped excavation in shanghai soft clays", *Geotechnique*, **62**, 907-921, https://doi.org/10.1680/geot.10.P.072.

Pu, Y.Y., Apel, D.B. and Xu, H.W. (2019), "Rockburst prediction in kimberlite with unsupervised learning method and support vector classifier", *Tunn. Undergr. Sp. Tech.*, **90**, 12-18. https://doi.org/10.1016/j.tust.2019.04.019.

Pawlak, Z. (1998), "Rough set theory and its applications to data analysis", *Cybernetics Syst.*, **29**(7), 661-688. https://doi.org/10.1080/019697298125470.

Ren, X. and Yu, X. (2011), "Multivariate statistical analysis (Version 2)", *China Statistics Press*, Beijing, China

Ren, X.C., Lai, Y.M., Zhang, F.Y. and Hu, K. (2015), "Test method for determination of optimum moisture content of soil and maximum dry density", *Ksce J. Civil Eng.*, **19**, 2061-2066. https://doi.org/10.1007/s12205-015-0163-0.

Rogers, CDF., Dijkstra, T.A. and Smalley, I.J. (1994), "Particle packing from an earth-science viewpoint", *Earth-Sci. Reviews*, **36**, 59-82. https://doi.org/10.1016/0012-8252(94)90008-6.

Sharifzadeh, M., Kolivand, F., Ghorbani, M. and Yasrobi, S. (2013), "Design of sequential excavation method for large span urban tunnels in soft ground—Niayesh tunnel", *Tunn. Undergr. Sp. Tech.*, **35**, 178-188. https://doi.org/10.1016/j.tust.2013.01.002

Specht, D.F. (1990), "Probabilistic neural networks and the polynomial adaline as complementary techniques for classification", *IEEE T. Neural Networ.*, **1**, 111-121, https://doi.org/10.1109/72.80210

Tuo, D.F., Xu, M.X., Li, Q. and Liu, S.H. (2017), "Soil aggregate stability and associated structure affected by long-term fertilization for a loessial soil on the loess plateau of china", *Polish J. Environ. Studies*, **26**, 827-835. https://doi.org/10.15244/pjoes/66716.

Wang, C.L., Li, C.F., Chen, Z., Liao, Z.F., Zhao, G.M., Shi, F. and Yu, W.J. (2020), "Experimental investigation on multi-parameter classification predicting degradation model for rock failure using bayesian method", *Geomech. Eng.*, **20**(2), 113-120. https: //doi.org/10. 12989/gae.2020.20.2.113

Witten, I.H. and Frank, E. (2002), "Data mining: Practical machine learning tools and techniques with Java implementations", *ACM SIGMOD Record*, **31**(1), 76-77, https://doi.org/10.1145/507338.507355.

Xie, Z. (2010), "Matlab statistical analysis and application: 40 case studies", *Beihang University Press*, Beijing,China

Xu, Z.G., Cai, N.G., Li, X.F., Xian, M.T. and Dong, T.W. (2021), "Risk assessment of loess tunnel collapse during construction based on an attribute recognition model", *Bull. Eng. Geology Environ.*, **80**, 6205-6220. https://doi.org/10.1007/s10064-021-02300-8.

Xue, Y., Zhang, X., Li, S., Qiu, D., Su, M., Li, L., Li, Z. and Tao, Y. (2018a), "Analysis of factors influencing tunnel deformation in loess deposits by data mining: A deformation prediction model", *Eng. Geology*, **232**, 94-103. https://doi.org/10.1016/j.enggeo.2017.11.014

Xue, Y.G., Bai, C.H., Qiu, D.H., Kong, F.M. and Li, Z.Q. (2020) "Predicting rockburst with database using particle swarm optimization and extreme learning machine", *Tunn. Undergr. Sp. Tech.*, **98**, https://doi.org/10. 1016/ j.tust.2020.103287.

Xue, Y.G., Zhang, X.L., Li, S.C., Qiu, D.H., Su, M.X., Li, L.P., Li, Z.Q. and Tao, Y.F. (2018b), "Analysis of factors influencing tunnel deformation in loess deposits by data mining: A deformation prediction model", *Eng. Geology*, **232**, 94-103. https://doi.org/ 10.1016/j.enggeo.2017.11.014.

Yates, K., Fenton, C.H. and Bell, D.H. (2018), "A review of the geotechnical characteristics of loess and loess-derived soils from canterbury, south island, new Zealand", *Eng. Geology*, **236**, 11-21, https://doi.org/10.1016/j.enggeo.2017.08.001.

Zhang, C.F., Peng, K.X. and Dong, J. (2020), "An incipient fault detection and self-learning identification method based on robust svdd and rbm-pnn", *J. Process Control*, **85**, 173-183, https://doi.org/10.1016/j.jprocont.2019.12.002.

Zhang, X.L., Xue, Y.G., Qiu, D.H., Yang, W.M., Su, M.X., Li, Z.Q. and Zhou, B.H. (2019), "Multi-index classification model for loess deposits based on rough set and bp neural network", *Polish J. Environ. Studies*, **28**, 953-963, https://doi.org/ 10.15244/pjoes/85303.

Zhao, Y., Li, G. and Yu, Y. (2011), "Loess tunnel engineering",

China Railway Press, Beijing, China.
Ziegel, E.R. (2005), "Discovering knowledge in data/next generation of data-mining applications", *Technometrics*, **47**(4), 528-529.
Zhou, B.H., Xue, Y.G., Li, S., Qiu, D., Tao, Y., Zhang, K. and Xia, T. (2020), "Probabilistic analysis of tunnel collapse: Bayesian method for detecting change points", *Geomech. Eng.*, **22**(4), 291-303. https: //doi. org/10.12989/gae. 2020.22.4.291.

GC

Nomenclature

Abbreviations and symbols

PCA	principal component analysis
PNN	probabilistic neural network
LOO	Leave-One-Out
NATM	New Austrian Tunnelling Method
SEM	Sequential excavation method
w	Water content
c	Cohesion
φ	Internal friction angle
E	Elastic modulus
μ	Poisson ratio