Prediction of Collapsibility of Loess of Construction Sites in Xining Based on Machine Learning Methods

Qifei Zhao

Xi'an University of Science and Technology

Xiaojun Li (✉ Lxjun@xust.edu.cn)

Xi'an University of Science and Technology

Yunning Cao

State Key Laboratory of Loess and Quaternary Geology

Zhikun Li

Xi'an University of Science and Technology

Jixin Fan

Qinghai Bureau of Coal Geology

Research Article

Keywords: Prediction of Loess Collapsibility, Grey Relational analysis, SVM, RSSVM, RF, NBTree

Posted Date: March 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-307514/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Prediction of collapsibility of loess of construction sites in Xining based on machine learning methods

Qifei Zhao^{1,2},Xiaojun Li^{1,2,*},Yunning Cao²,Zhikun Li¹,Jixin Fan³

¹Xi'an University of Science and Technology, Xi'an, 710054, China

²State Key Laboratory of Loess and Quaternary Geology, Xi'an, 710061, China

³Qinghai Bureau of Coal Geology, Xining, Qinghai, 810029, China

*Correspondence:Lxjun@xust.edu.cn

Abstract: Collapsibility of loess is a significant factor affecting engineering construction in loess area, and testing the collapsibility of loess is costly. In this study, A total of 4,256 loess samples are collected from the north, east, west and middle regions of Xining. 70% of the samples are used to generate training data set, and the rest are used to generate verification data set, so as to construct and validate the machine learning models. The most important six factors are selected from thirteen factors by using Grey Relational analysis and multicollinearity analysis: burial depth、 water content、 specific gravity of soil particles、 void rate、 geostatic stress and plasticity limit. In order to predict the collapsibility of loess, four machine learning methods: Support Vector Machine (SVM), Random Subspace Based Support Vector Machine (RSSVM), Random Forest (RF) and Naïve Bayes Tree (NBTree), are studied and compared. The receiver operating characteristic (ROC) curve indicators, standard error (SD) and 95% confidence interval (CI) are used to verify and compare the models in different research areas. The results show that: RF model is the most efficient in predicting the collapsibility of loess in Xining, and its AUC average is above 80%, which can be used in engineering practice.

Key words : Prediction of Loess Collapsibility, Grey Relational analysis, SVM, RSSVM, RF, NBTree

1. Introduction

Being a special soil, loess retains a variety of special geotechnical engineering properties, such as regionality, large porosity, low bearing capacity, water sensitivity and low strength. Especially, the collapsibility of loess often leads to surface subsidence, deformation or cracking of building foundations, which seriously affects the stability of buildings¹. Therefore, measuring loess collapsibility coefficient is the main index for evaluating the collapsibility of loess site². The collapsibility coefficient and related physical and mechanical indexes are generally obtained by in-situ and laboratory testing, but the regional collapsibility analysis requires a large number of on-site sampling and laboratory tests, which sounds time-consuming and uneconomical. Based on the existing massive data, establishing prediction model of loess collapsibility with physical indexes will be an indirectly method to reduce the cost, speed up the progress and improve the efficiency³. Located in the intersection of northwest inland arid area and Qinghai Tibet Plateau alpine region, Xining is in the western edge of Chinese Loess Plateau. Being a key provincial capital city for the development of western China, it is a general trend to develop industrial economy and promote the construction of industrial clusters in the economic zone. However, due to the particularity of geological environment in Xining, the collapsibility of loess foundation is the most significant geological problem in engineering construction. Recently, with the help of certain mathematical methods, plenty of scholars use the existing data to simplify the laboratory test and improve the efficiency within the acceptable error range by predicting. Some scholars have established ANFIS prediction model by MATLAB. By training and predicting loess samples, the results are quite approximate to the actual loess collapsibility coefficient, which testifies that ANFIS method is an ideal prediction method⁴. Anning⁵ finally has proposed a new method to predict loess collapsibility by extracting the principal component and eliminating the correlation of variables with the help of the principal component analysis (PCA), and establishing adaptive neuro-fuzzy inference system (ANFIS) model. Some researchers have utilized partial correlation analysis to remove some insignificant loess collapsibility influencing factors. The RBF neural network is introduced and the remaining factors are taken as input variables to establish a prediction model of the collapsibility coefficient which can meet the engineering requirements⁶. Zhang Shuhua⁷ has applied to the generalized regression neural network (GRNN) to evaluate loess collapsibility. By adjusting the SPD value appropriately, the prediction accuracy of the model is higher than that of RBF. Some people have made full use of SVM to evaluate loess collapsibility and has proved the feasibility and rationality of utilizing SVM to evaluate loess collapsibility by comparing the predicted data with the measured data⁸. Some domestic scholars have preprocessed the data with PCA,

introduced particle swarm optimization (PSO) to optimize the model parameters, and utilized the least squares support vector machines to build the model to predict loess collapsibility⁹. In conclusion, it is obvious that machine learning method is effective in predicting collapsibility of loess. With the continuous development and optimization of machine learning, some new computing methods have been produced, such as NBTree, which is a mixture of decision tree and Naïve Bayes, and RF based on decision tree, all of which have proved to be powerful tools for predicting environmental problems such as floods¹⁰⁻¹¹, forest fires¹²⁻¹³, landslides¹⁴⁻¹⁶ and groundwater spring potential maps¹⁷. On the other hand, SVM is a popular and effective method in predicting loess collapsibility, yet these methods have not been compared with SVM attempting to predict loess collapsibility.

An excellent integration algorithm, the random subspace algorithm (RS)¹⁸, is introduced in this study. The RS and SVM are coupled and firstly applied to loess collapsibility prediction, and then compared with SVM; RF and NBTree are firstly introduced into the collapsibility prediction of loess foundation. Four models of SVM, RSSVM, RF and NBTree are applied to prediction of loess collapsibility in the north, east, west and middle regions of Xining city. This study compares and analyzes the most effective algorithm model to provide a powerful tool for engineering construction in Xining city.

2. Study site and data

2.1 Description of the study site

The study area is located in the urban area of Xining, Qinghai province, China, with a total area of 460.7 square kilometers, covering 36°28'21"-36°45'29" N north latitude, 101°37'15"-101°54'47"E East longitude. Xining city is located in the middle of a Mesozoic faulted basin, faulted structure and folded structure in Xining basin are relatively weak, and the overall terrain is higher in the south and lower in the north, with the elevation from 2118 meters to 2831 meters. The loess soil layer is widely distributed. The river in the area belongs to the Huangshui Water System(an important tributary of the upper Yellow River), whose main river is Huangshui main stream, and tributaries are Nanchuan River, Beichuan River. The area is a semi-arid plateau continental climate characterized by low precipitation, high evaporation, strong solar radiation and large diurnal temperature differences. According to the perennial observation data of Qinghai Meteorological Observatory, the annual average precipitation is 371.0mm, the precipitation distribution is not uniform, generally concentrated in June to September, accounting for 73% of the total annual precipitation, and has the characteristics of precipitation concentration and heavy rain in July to September.

132 2.2 Data

133 In this study, a total of 5660 samples of 32 projects has been obtained from the north, east, west and middle region
134 of Xining, among which 4256 samples of 28 projects are used to generate modeling data sets. The sample is randomly
135 divided into two parts: 70% of loess samples are used to construct training data sets and 30% of samples are used to
136 establish validation data sets. Finally, 1404 samples of the remaining 4 projects are used for prediction. The exact
137 location of these projects is shown in Figure 1.

138
139 **Figure 1.** Geographic location of the data collection

140 13 influencing factors of loess collapsibility, namely water content (ω), saturability (S_r), dry density (γ_d), natural
141 density (γ), void rate (e), specific gravity of soil particles (G_s), burial depth (L), porosity (n), liquid limit (ω_L), plastic
142 limit (ω_p), liquidity index (I_L), plasticity index (I_p) and geostatic stress (P), among different study areas are shown in
143 following Table 1.

144 **Table 1.** Influencing factor information of different study areas

The study area	Number of samples	Statistic	L /(m)	ω /(%)	S_r /(%)	γ /(g/cm ³)	γ_d /(g/cm ³)	G_s	e
The north region	1039	Minimum	1.1	4.8	13.1	1.27	1.13	2.69	0.552
		Maximum	23.1	29	93.4	2	1.74	2.72	1.394
		Mean	8.517	14.71	44.949	1.628	1.418	2.694	0.906
		StdDev	6.024	4.689	17.276	0.134	0.085	0.006	0.115
The east region	1019	Minimum	2.1	3	6.6	1.25	1.14	2.69	0.567
		Maximum	26.1	30.1	96.8	2.06	1.74	2.72	1.379
		Mean	8.931	13.362	42.574	1.627	1.433	2.697	0.891

		StdDev	4.944	5.745	21.86	0.169	0.097	0.009	0.122
The west region	1063	Minimum	1.6	5.9	16.7	1.06	0.95	2.69	0.506
		Maximum	16.6	29	95.4	2.02	1.79	2.73	1.845
		Mean	6.649	15.567	49.471	1.677	1.451	2.7	0.869
		StdDev	3.317	4.437	16.35	0.131	0.093	0.008	0.122
The middle region	1135	Minimum	2.1	4.2	10.3	1.32	1.16	2.69	0.512
		Maximum	30.1	27.8	87.9	1.98	1.78	2.72	1.32
		Mean	7.707	14.59	47.164	1.67	1.456	2.699	0.861
		StdDev	4.665	4.328	16.809	0.142	0.097	0.009	0.123
The whole study area	4256	Minimum	1.1	3	6.6	1.06	0.95	2.69	0.506
		Maximum	30.1	30.1	96.8	2.06	1.79	2.73	1.845
		Mean	7.934	14.569	46.101	1.651	1.44	2.698	0.881
		StdDev	4.898	4.875	18.319	0.146	0.094	0.009	0.122

145 Continue table

The study area	Number of samples	Statistic	n /(%)	ω_L /(%)	ω_p /(%)	I_L	I_p	P /(kPa)
The north region	1039	Minimum	35.6	20.9	13.4	-1.27	0.44	18.26
		Maximum	58.2	32.6	25.2	11.7	12.3	404.25
		Mean	47.35	22.802	14.823	-0.023	7.958	132.56
		StdDev	3.124	1.833	1.058	0.726	1.128	94.301
The east region	1019	Minimum	36.2	20.5	1.7	-1.51	6.2	26.25
		Maximum	58	31.8	19.9	1.71	26.6	425.43
		Mean	46.895	23.678	14.973	-0.263	8.705	139.553
		StdDev	3.476	2.68	1.072	0.61	1.814	76.685
The west region	1063	Minimum	33.6	20.9	9.3	-1.12	4.7	24
		Maximum	64.9	31.8	20.3	1.28	16.5	251.98
		Mean	46.277	24.421	15.226	0	9.195	108.668
		StdDev	3.417	2.349	0.878	0.453	1.673	53.875
The middle region	1135	Minimum	33.9	20.7	13.5	-1.32	6.6	28.77
		Maximum	56.9	30.5	17.5	1.11	13	382.46
		Mean	46.048	23.964	15.074	-0.105	8.89	124.33
		StdDev	3.491	2.514	0.871	0.435	1.672	69.376
The whole study area	4256	Minimum	33.6	20.5	1.7	-1.51	0.44	18.26
		Maximum	64.9	32.6	25.2	11.7	26.6	425.43
		Mean	46.626	23.726	15.027	-0.097	8.694	126.072
		StdDev	3.42	2.438	0.982	0.574	1.657	75.589

146 2.3 Calculation of collapsibility coefficient of loess

147 The collapsibility coefficient of loess (δ_s) is the main index to evaluate the collapsibility of loess site. δ_s refers to the
148 ratio of the collapsible deformation of the sample caused by immersing to the initial height of the sample, which can
149 be determined by indoor confined compression experiment. The calculation formula is as follows:

$$\delta_s = \frac{h_p - h'_p}{h_0} \quad (1)$$

Where h_p refers to sample height measured after compression stabilization gradually pressurizing the undisturbed sample to the specified pressure p in the compression instrument. Then it is wet with water, and then measure the height h'_p after the settlement is stable. h_0 refers to the thickness of undisturbed soil sample.

Four different collapsibility grades of loess are determined with different δ_s according to *the Engineering Geology Handbook*¹⁹. They are non collapsibility ($\delta_s < 0.015$), slight collapsibility ($0.015 \leq \delta_s < 0.03$), medium collapsibility ($0.03 \leq \delta_s < 0.07$) and serious collapsibility ($\delta_s \geq 0.07$). The collapsibility grade distribution of loess samples in different regions is shown in Fig. 2.

Figure 2. The collapsibility grade distribution of loess samples in different study areas

3. Research methods

3.1 Selection of Influencing Factors

Choosing the main influencing factors is a key step to improve model accuracy. In this study, the main influencing

163 factors are selected through grey relational analysis, and then the multicollinearity analysis is used to test.

164 3.1.1 Grey Relational analysis

165 The Grey Relational analysis is a multi-factor statistical analysis method proposed by Tan and Deng²⁰, which is based
166 on the sample data of each influencing factor. The grey relational degree is used to represent the strength of the
167 relationship between each factor. If the sample data reflects the change trend of the two factors is basically the same,
168 the relational degree value between them is relatively large. On the contrary, the relational degree value is small. There
169 are totally 13 influencing factors of loess collapsibility, which belongs to 6 categories separately. The representative
170 factors of each category are selected by grey relational analysis.

171 3.1.2 Multicollinearity analysis

172 The tolerance (TOL) and variance inflation factor (VIF) are two common indicators utilized for multicollinearity
173 check. The calculation formula is as follows:

$$\text{TOL} = 1 - R^2 \quad (2)$$

$$\text{VIF} = \frac{1}{\text{TOL}} \quad (3)$$

174 When TOL value is less than 0.1 or VIF value is greater than 10, it indicates that there is a serious multicollinearity
175 problem between factors²¹.

176 3.2 Modeling Approach

177 3.2.1 Support Vector Machine (SVM)

178 SVM is a dichotomous model first proposed by Corinna Cortes and Vapnik²² in 1995, which can be used to solve
179 Machine learning problems such as small samples, nonlinear and high-dimensional models, which aims to find a
180 hyperplane to segment the sample. The principle of segmentation is to maximize the interval and finally transform it
181 into convex quadratic programming problem²³. Among them, Kernel function is the essence of SVM, which can map
182 the original input space to the feature space of higher dimensions, so that the original linearly indivisible samples can
183 become linearly separable in the new Kernel space²⁴. Common Kernel functions include:

$$\text{Polynomial kernel function: } k(x_1, x_2) = (x_1^T x_2 + c)^d \quad (4)$$

$$\text{RBF: } k(x_1, x_2) = e^{-\frac{\|x_1 - x_2\|^2}{2\sigma^2}} \quad (5)$$

$$\text{Sigmoid: } k(x_1, x_2) = \tanh(x_1^T x_2 + c) \quad (6)$$

184 Where d, σ and c are the parameters of the kernel function.

3.2.2 Random Subspace (RS)

RS is a classical ensemble learning algorithm firstly proposed by HO²⁵ in 1998. RS reduces the correlation between classifiers by training each classifier with random partial features. RS can not only randomly select data from training data sets to establish training subsets, but also modify data sets with redundant functions and overfitting problems in feature space²⁶⁻²⁷. RS is used to reduce the feature set of the high-dimensional feature space to the low-dimensional space, and then a classifier is constructed in the low-dimensional space to classify the class. Finally, the final result is obtained by the rule of majority voting²⁸.

3.2.3 Random Forests (RF)

RF is an ensemble learning method based on Bagging²⁹, which can be used for both regression and classification, which will be used widely in the future³⁰. It mainly combines several weak classifiers, and the final result is voted or averaged, thus letting the results of the overall model with high accuracy and generalization performance, and finally achieving good results³¹. The superior performance of RF is mainly attributed to “random” and “forest”, among which one makes it have anti-over-fitting ability and anti-noise ability, the other makes it more accurate. RF regards CART decision tree as a weak learner. Meanwhile, when each tree is generated, the selected features of each tree are only a few randomly selected features. Generally, the square root of the total number of features M is taken by default. RF requires little for data sets which need neither normalization nor feature selection. It is able to process discrete data, continuous data and high-dimensional data simply and quickly, making it an excellent classical integration algorithm and widely used.

3.2.4 Naïve Bayes Tree (NBTree)

NBTree is a hybrid algorithm that Kohavi first integrated C4.5 decision tree and Naïve Bayes in 1996³². Its learning process is based on C4.5 decision tree and leaf node creation is conducted through Naïve Bayes classifier³³.

C4.5 algorithm is a classification decision tree algorithm with ID3 algorithm as the core. By employing information gain rate to select split attributes, it overcomes the deficiency of ID3 algorithm that prefers attributes with multiple selection values when using information gain to select attributes, and the continuous attribute can be discretized.

The core of naïve bayes classification algorithm is bayes formula, namely, $P(\text{category} | \text{feature}) = \frac{P(\text{feature} | \text{category})P(\text{category})}{P(\text{feature})}$. Let D be the sample training set, and each sample X is composed of N attribute values, that is, $X = (x_1, x_2, \dots, x_n)$, the corresponding attribute set is $A_1, A_2, A_3, \dots, A_n$. Supposing there are M class tags,

namely $C_1, C_2, \dots, C_m$. For some element X to be classified, the naïve classifier will put $P(C_i|X)$ ($i=1,2,\dots,m$) the class tag with the largest value C_1 is used as the category. So it aims to find the maximum value of $P(C_1|X)$ ($P(C_1|X) = P(X|C_1) \cdot P(C_1) / P(X)$). If the value of N is particularly large, then the calculation for $P(X|C_1)$ is quite complicated. So naïve bayes algorithm makes the assumption that for each attribute in the sample tuple, since they are conditional independent of each other, $P(X|C_1) = P(X_1|C_1)P(X_2|C_1)\dots P(X_n|C_1)$.

3.3 Model evaluation and assessment

ROC curve is a comprehensive indicator reflecting continuous variables of sensitivity and specificity³⁴. The ROC curve is used to evaluate the performance of the loess collapsibility model. For each predicted value of the model, there are cases of correct classification and wrong classification. TP (true positive) and TN (true negative) respectively represents the number of results correctly classified as yes or no, while FP (false positive) and FN (false negative) respectively represents the number of results wrongly classified as yes or no. For a classification algorithm, the evaluation indexes mainly include: Accuracy, Sensitivity, Specificity, etc. The calculation method is as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (9)$$

When comparing model performance, if the ROC curves of different models do not intersect, the closer they are to the curve in the upper left corner, the model they represent has the best performance. On the contrary, if the ROC curve intersects, it is not conducive to judgment. Therefore, the learning performance of loess collapsibility model is evaluated through the area under the ROC curve (AUC). The higher the AUC value, the better the model performance. $AUC > 0.9$ represents nearly perfect, $0.8 < AUC < 0.9$ represents excellent, $0.7 < AUC < 0.8$ represents fine, $0.6 < AUC < 0.7$ represents medium, $AUC < 0.6$ represents no value. When AUC value is equal to 1, it represents the ideal model³⁵. The calculation formula of AUC value is as follows:

$$AUC = \frac{\sum TP + \sum TN}{P + N} \quad (10)$$

Where P is the total number of YES and N is the total number of NO.

Flowchart of the study is shown in Fig3.

Figure 3. Flowchart of the study

4. Results and discussion

4.1 Analysis of representative factors

4.1.1 Determination of influence factors

The 13 influences of loess collapsibility are divided into six categories, namely, water indicators, pore, geostatic stress, density, burial depth and physical characteristics. The most representative factors of each category are found out to be the important influence factors of loess collapsibility through grey relational analysis. The results are represented in FIG. 4. It can be seen from the figure that Pore is proportional to the coefficient of collapsibility, that is, the more pores in the soil, the larger the degree of loess collapsibility. The density is inversely proportional to the coefficient of collapsibility, indicating that the collapsibility of loess decreases with the increase of soil density, which shows a reasonable result³⁶. The burial depth and geostatic stress are inversely proportional to the collapsibility coefficient, which is consistent with the results of previous studies³⁷. Physical characteristics are inversely proportional to the collapsibility coefficient, which may be that the stronger the plasticity of the soil, the lower the collapsibility. Water indicators are inversely proportional to the coefficient of collapsibility, indicating that with the increase of water indicators, the collapsibility of loess weakens and the results are in line with previous studies on Xining loess³⁸. According to the results of grey relational degree analysis, six representative factors are selected from

250 six types of factors, namely, burial depth, water content, specific gravity of soil particles, void rate, geostatic stress
 251 and plasticity limit.
 252

253
 254 **Figure 4.** The correlation degree between the influence factors and δ_s

255 *4.1.2 Correlation Analysis of Influence Factors*

256 The training data set of the entire research area is used to examine the multicollinearity of six factors of Plasticity
 257 depth, Water Content, Specific gravity of soil particles, Void rate, Geostatic stress and Plasticity limit through IBM
 258 SPSS Statistics 26 software, and the results are shown in Table 2:

259

Table 2. Multi-collinearity analysis of Loess collapsibility factors		
Loess collapsibility factors	Statistics	
	TOL	VIF
Burial depth	0.018	56.762
Water content	0.522	0.914
Specific gravity of soil particles	0.379	2.638
Void rate	0.642	1.559
Geostatic stress	0.017	57.726
Plasticity limit	0.411	2.432

260 As can be seen from Table 2, the TOL values of the four factors of Water content, Specific gravity of soil particles,

Void and Plasticity limit are all greater than 0.1 and VIF values are all less than 10, so there is no multicollinearity. The TOL values of both the Burial Depth and Geostatic stress are less than 0.1 and the VIF values are both greater than 10, indicating that the two factors have multicollinearity and one of them should be removed. However, considering the special geological environment of Xining, the loess in the area is mainly hydroformed loess, and the foundation excavation, groundwater level and its changes in the area will have an effect on the additional pressure, pore density and soil structure in the soil ², which will finally affect the change of soil pressure, so Geostatic stress is adopted in this study.

4.2 Performance and validation of models in different study areas

4.2.1 Performance of models

SVM, RSSVM, RF and NBTree models are built with the help of the training data set. In the calculation process, the 10-fold cross-validation method is used to reduce the variability of the results. The calculated results of the training model are shown in Table 3. It can be seen from the table that in the non-collapsible loess model training, RF has the highest accuracy compared with SVM, RSSVM and NBTree, whose values are 88.6% (north of the city), 88.1% (east of the city), 87.8% (west of the city) and 84.9% (middle of the city) respectively. NBTree is second to the RF model. However, Sensitivity refers to the percentage of correctly predicted non-collapsibility loess samples in the actual data set of non-collapsibility loess, and RF is still the highest value in the north, west and middle urban areas, with values of 82.8%, 82.1% and 79.4%, respectively. But in the eastern part of the city, NBTree (83.4%) has the highest value and RF (77.7%) is second to it. RSSVM shows the highest specificity in the north region (93.1%), east region (96.9%) and West Region (96.7%). This indicates that there are 93.1%, 96.9% and 96.7% collapsible loess samples, respectively in the north region, east region and west region, being predicted correctly. Meanwhile, RF shows a good specificity in north region (92.4%), East region (93%) and West Region (90.9%). Even in urban areas, RF (88.5%) was higher than the RSSVM (87%) specificity. The RF had the highest AUC values in the four study areas, which are 94.5% (North Region), 95.2% (East Region), 94.7% (West Region) and 92.1% (Middle Region), indicating that the RF is highly suitable for training data sets.

In the slight collapsible loess model training, RF and NBTree have the same accuracy rate (79.2%) in the north urban area. In the eastern region, NBTree (75.9%) had a better accuracy rate than SVM (75.2%), while RF (74.8%) had a lower accuracy rate than SVM. RF (70.8%) is only second to SVM (71.8%) in the west part of the city, but RF (73.1%) has the highest accuracy in the middle part of the city. As for Sensitivity, the Sensitivity values of north region, east

region, west region and middle region are not ideal. The specificity of RSSVM is highest in north region (97.2%), east region (95.5%), west region (95.4%) and middle region (89.1%). Although the overall specificity of RF in the four models is the lowest, it is also relatively excellent, with 88.4% (north region), 87% (east region), 84% (West Region) and 80.8% (middle region), respectively. The RF has the highest AUC value in the north region (84.5%), east region (75.5%), west region (72.6%) and middle region (80.4%), indicating that the RF model has the best performance in the training process.

In the training of medium collapsible loess model, the RF model has the highest accuracy in the north region (83.1%), the west region (75.9%) and the middle region (81.3%). In the middle region, the accuracy of SVM (81.1%) and RSSVM (81%) is very close to that of RF. In the eastern part of the city, SVM (77.1%) and RF (77%) have the highest accuracy among the four models. Model Sensitivity in the north region, west region and middle region of the city is not ideal, but NBTree (71.7%) in the east region showed good Sensitivity. In terms of specificity, RSSVM performs best in the north region (96.9%), east region (92.3%), west region (94.2%), and middle region (97%). The other three models also have good specificity values in these four study areas. On the whole, RF is the most suitable training data set among the four models, because its AUC value is the highest among north (88.9%), east (82.3%), west (79.9%) and middle (83.5%).

In the training of the loess model of serious collapsibility, all four kinds of models shows quite low Sensitivity and high Specificity in different research areas, and all their accuracy rate is over 90%. However, both SVM and RSSVM models consistently shows poor performance, while RF and NBTree models shows excellent and similar performance, with AUC values between 76% and 92% in different study areas.

By comparing (a), (b), (c) and (d) in Figure 5, it is obvious that during the training process of non-collapsibility, slight collapsibility, medium collapsibility and serious collapsibility of loess model. The AUC value of RF model is almost the highest among the four models, which means that the RF model is most suitable to the training data set in these four different regions, followed by NBTree and then RSSVM models, while the SVM model performs poorly.

Table 3. Comparison of performance of models in different study areas

The study area	Parameter	Non collapsibility				Slight collapsibility			
		SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
The north region	True positive	205	168	240	217	97	70	136	114
	True negative	393	407	404	391	458	484	440	462
	False positive	44	30	33	46	40	14	58	36
	False negative	85	122	50	73	132	159	93	115

	Sensitivity	0.707	0.579	0.828	0.748	0.424	0.306	0.594	0.498
	Specificity	0.899	0.931	0.924	0.895	0.92	0.972	0.884	0.928
	Accuracy	0.823	0.791	0.886	0.836	0.763	0.762	0.792	0.792
	AUC	0.803	0.889	0.945	0.898	0.672	0.751	0.845	0.759
The east region	True positive	138	116	178	191	66	42	92	90
	True negative	455	469	450	430	470	484	441	451
	False positive	29	15	34	54	37	23	66	56
	False negative	91	113	51	38	140	164	114	116
	Sensitivity	0.603	0.507	0.777	0.834	0.32	0.204	0.447	0.437
	Specificity	0.94	0.969	0.93	0.888	0.927	0.955	0.87	0.89
	Accuracy	0.832	0.82	0.881	0.871	0.752	0.738	0.748	0.759
	AUC	0.771	0.883	0.952	0.915	0.624	0.675	0.755	0.74
The west region	True positive	158	175	216	205	72	34	103	33
	True negative	448	465	437	439	462	482	424	454
	False positive	33	16	44	42	43	23	81	51
	False negative	105	88	47	58	167	205	136	206
	Sensitivity	0.601	0.665	0.821	0.779	0.301	0.142	0.431	0.138
	Specificity	0.931	0.967	0.909	0.913	0.915	0.954	0.84	0.899
	Accuracy	0.815	0.86	0.878	0.866	0.718	0.694	0.708	0.655
	AUC	0.766	0.912	0.947	0.922	0.608	0.677	0.726	0.613
The middle region	True positive	240	209	246	235	146	108	182	164
	True negative	416	422	429	419	420	440	399	381
	False positive	69	63	56	66	74	54	95	113

313 Continue table

The study area	Parameter	Medium collapsibility				Serious collapsibility			
		SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
The north region	True positive	68	49	113	110	0	0	2	0
	True negative	504	525	491	468	703	704	702	704
	False positive	38	17	51	74	1	0	2	0
	False negative	117	136	72	75	23	23	21	23
	Sensitivity	0.368	0.265	0.611	0.595	0	0	0.087	0
	Specificity	0.93	0.969	0.906	0.863	0.999	1	0.997	1
	Accuracy	0.787	0.79	0.831	0.795	0.967	0.968	0.968	0.968
	AUC	0.649	0.729	0.889	0.84	0.499	0.499	0.798	0.826
The east region	True positive	126	86	147	165	9	2	11	13
	True negative	424	446	402	375	664	665	654	647
	False positive	59	37	81	108	1	0	11	18
	False negative	104	144	83	65	39	46	37	35
	Sensitivity	0.548	0.374	0.639	0.717	0.188	0.042	0.229	0.271
	Specificity	0.878	0.923	0.832	0.776	0.998	1	0.983	0.973
	Accuracy	0.771	0.746	0.77	0.757	0.944	0.935	0.933	0.926
	AUC	0.713	0.789	0.823	0.815	0.593	0.612	0.917	0.861

The west region	True positive	62	31	99	85	0	0	4	0
	True negative	492	505	466	436	709	710	702	710
	False positive	44	31	70	100	1	0	8	0
	False negative	146	177	109	123	34	34	30	34
	Sensitivity	0.298	0.149	0.476	0.409	0	0	0.118	0
	Specificity	0.918	0.942	0.869	0.813	0.999	1	0.989	1
	Accuracy	0.745	0.72	0.759	0.7	0.953	0.954	0.949	0.954
	AUC	0.608	0.707	0.799	0.743	0.499	0.5	0.769	0.811
The middle region	True positive	53	35	76	66	0	0	3	3
	True negative	592	609	570	559	778	778	776	773
	False positive	36	19	58	69	0	0	2	5
	False negative	114	132	91	101	17	17	14	14
	Sensitivity	0.317	0.21	0.455	0.395	0	0	0.176	0.176
	Specificity	0.943	0.97	0.908	0.89	1	1	0.997	0.994
	Accuracy	0.811	0.81	0.813	0.786	0.979	0.979	0.98	0.976
	AUC	0.63	0.692	0.835	0.771	0.5	0.5	0.885	0.784

(a) AUC value of non-collapsible model

(b) AUC value of slightly collapsible model

(c) AUC value of medium collapsibility model

(d) AUC value of serious collapsible model

Figure 5. The performance comparison of different collapsibility training models

Table 4 shows the comparative evaluation results of the loess training models for non-collapsibility, slight collapsibility, medium collapsibility and serious collapsibility in different study areas. Three evaluation statistics, namely, the SD, 95% CI and significance level (P-value), are used. These statistical methods are used to verify and compare the performance of the model. Generally speaking, the smaller SD, the narrower 95% CI and the lower P-value indicate that the model has better performance. It can be found in Table 4 that, compared with SVM, RSSVM and NBTree models, RF models can have the smallest SD, the narrowest 95% CI and the lowest P-value, both in different research areas and in different training of collapsible loess, which indicates that the RF model is most suitable for the training data set of each region in this study.

Table 4. Performance parameters of the training model

The study area	Evaluation statistics	Non collapsibility				Slight collapsibility			
		SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
The north region	SD	0.018	0.013	0.008	0.012	0.023	0.022	0.015	0.02
	95% CI	0.768 to	0.863 to	0.929 to	0.874 to	0.626 to	0.709 to	0.815 to	0.72 to
		0.838	0.915	0.96	0.921	0.717	0.793	0.875	0.798
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
The east region	SD	0.021	0.015	0.008	0.011	0.025	0.024	0.019	0.021
	95% CI	0.73 to	0.854 to	0.937 to	0.892 to	0.575 to	0.628 to	0.717 to	0.699 to
		0.813	0.913	0.968	0.937	0.672	0.721	0.793	0.781
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
The west region	SD	0.02	0.012	0.008	0.01	0.023	0.021	0.019	0.021
	95% CI	0.727 to	0.888 to	0.932 to	0.902 to	0.563 to	0.635 to	0.689 to	0.572 to
		0.805	0.936	0.962	0.942	0.654	0.719	0.763	0.653
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
The middle region	SD	0.017	0.013	0.01	0.014	0.021	0.018	0.015	0.018
	95% CI	0.783 to	0.844 to	0.902 to	0.836 to	0.627 to	0.705 to	0.774 to	0.693 to
		0.848	0.896	0.94	0.891	0.708	0.776	0.834	0.764
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Continue table

The study area	Evaluation statistics	Medium collapsibility				Serious collapsibility			
		SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
The north region	SD	0.026	0.024	0.013	0.017	0.061	0.061	0.049	0.039
	95% CI	0.599 to	0.682 to	0.863 to	0.808 to	0.379 to	0.379 to	0.702 to	0.75 to
		0.699	0.775	0.915	0.872	0.619	0.619	0.894	0.902
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	0.991	0.991	<0.0001	<0.0001
	SD	0.022	0.019	0.016	0.017	0.048	0.048	0.018	0.03

The east region	95% CI	0.669 to 0.756	0.752 to 0.826	0.792 to 0.855	0.782 to 0.847	0.499 to 0.687	0.517 to 0.706	0.881 to 0.953	0.802 to 0.92
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	0.031	0.01	<0.0001	<0.0001
	SD	0.024	0.022	0.016	0.018	0.051	0.051	0.048	0.035
The west region	95% CI	0.56 to 0.656	0.664 to 0.75	0.767 to 0.831	0.708 to 0.778	0.4 to 0.599	0.401 to 0.599	0.675 to 0.862	0.742 to 0.88
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	0.989	0.1	<0.0001	<0.0001
	SD	0.027	0.026	0.016	0.02	0.071	0.071	0.051	0.063
The middle region	95% CI	0.578 to 0.682	0.642 to 0.743	0.803 to 0.866	0.732 to 0.811	0.361 to 0.639	0.361 to 0.639	0.385 to 0.661	0.661 to 0.908
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	1	1	<0.0001	<0.0001
	SD	0.027	0.026	0.016	0.02	0.071	0.071	0.051	0.063

4.2.2 Validation of models

After the training data set is used to build the model, the established model is verified through the validation data set to determine the predictive ability and accuracy of the model. The validation results are shown in Table 5. It can be seen from the table that in the verification of non-collapsible loess model, RF model has the best classification accuracy rate in the eastern part of the city (89.9%), the western part of the city (88.7%) and the middle part of the city (87.6%). In the northern part of the city, the RF (87.8%) classification accuracy rate of the model is second to NBTree's (89.4%). Among the north region (86.7%), the east region (87.8%), and the west region (85.9%), NBTree has the highest Sensitivity, while the SVM model in the middle region (82.1%) has the highest sensitivity. RSSVM shows the highest specificity in north region (94%), east region (95.5%) and West Region (95.8%), and RF (92.3%) has the highest specificity in middle region. By observing the AUC values of the four models in four different regions, it is found that all the four models have good predictive ability, among which the best one is the RF model. The AUC values of north region, east region, west region and middle region are 94.1%, 95.3%, 95.5% and 94% respectively, among which the prediction ability of NBTree model is second to RF.

According to the verification results of the slight collapsible loess model, the accuracy of RF model (82.4% in the north region, 78.4% in the east region, 72.7% in the west region, and 73.5% in the middle region) is better than other models. As for Sensitivity, the four models of SVM, RSSVM, RF and NBTree fails to reach the ideal Sensitivity value in the whole research area. The RSSVM model has the highest specificity values in the north, east, West and middle regions, which are 98.1%, 92.5%, 94.6% and 91.5%, respectively, indicating that 98.1%, 92.5%, 94.6% and 91.5% of non-slight collapsible loess samples are correctly classified in the four regions. The RF model still has the highest AUC value in the northern region (87.8%), eastern region (79.8%), western region (77.4%) and middle region (81.1%), indicating that the RF model has better predictive ability than the other three models.

In the verification of the medium collapsible loess model, the accuracy of the four models in the north region and the middle region of the city is quite high, between 80% and 85%, which is higher than that of the east region and the west region on the whole. In the east region, RF (77.8%) model is only second to NBTree (78.4%) model in accuracy. RSSVM (78.1%) has the highest accuracy rate in the west region, with RF and NBTree having the same accuracy rate of 76.5%. For Sensitivity, except for NBTree model, which shows good performance in the eastern part of the city (78.2%), the Sensitivity of the model is lower in other areas. The four models shows excellent specificity in different research areas, among which RF model is the best, which is 97.9%, 94.1%, 96.7% and 96.1% respectively in the north, east, West and middle regions. By observing the AUC values of different models in different research areas, it can be found that the RF model has the best predictive ability on the whole. The AUC values of north region, east Region, West region and middle region are 89.2%, 84.6%, 82.2% and 84.9% respectively. The prediction ability of NBTree model is the closest to RF, with AUC values of 85.5%, 84.1% and 81% in the north, east and middle regions of the city, respectively. Even in the west region, NBTree (AUC value is 83%) has a higher prediction ability than RF model.

The validation results of serious collapsible loess model finds that all four models has quite low sensitivity, quite high specificity and accuracy, just as the training results. In the north, east and west regions of the city, only the AUC values of RF and NBTree models reach the excellent level, both at about 85%. However, only the NBTree model shows good predictive ability in the middle area, with an AUC value of 86.3%.

Figure 6 shows the predictive performance of the model through the AUC value of each model. By comparing (a), (b), (c) and (d), it can be clearly found that: Overall, the AUC value of the RF model is in the highest position, which indicates that the predictive ability of the RF model is superior to the other three models in the prediction of different collapsible loess models in the four study areas.

371

Table 5. Comparison of validation of models in different study areas

The study area	Parameter	Non collapsibility				Slight collapsibility			
		SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
The north region	True positive	101	84	105	111	50	45	59	52
	True negative	168	173	169	168	197	206	198	199
	False positive	16	11	15	16	13	4	12	11
	False negative	27	44	23	17	52	57	43	50
	Sensitivity	0.789	0.656	0.82	0.867	0.49	0.441	0.578	0.51
	Specificity	0.913	0.94	0.918	0.913	0.938	0.981	0.943	0.948
	Accuracy	0.862	0.824	0.878	0.894	0.792	0.804	0.824	0.804

	AUC	0.851	0.905	0.941	0.926	0.714	0.794	0.878	0.819
The east region	True positive	54	49	68	72	28	25	51	38
	True negative	206	214	207	194	192	198	189	185
	False positive	18	10	17	30	22	16	25	29
	False negative	28	33	14	10	64	67	41	54
	Sensitivity	0.659	0.598	0.829	0.878	0.304	0.272	0.554	0.413
	Specificity	0.92	0.955	0.924	0.866	0.897	0.925	0.883	0.864
	Accuracy	0.85	0.859	0.899	0.869	0.719	0.729	0.784	0.729
	AUC	0.789	0.865	0.953	0.898	0.601	0.7	0.798	0.722
The west region	True positive	77	83	106	110	29	17	43	25
	True negative	180	183	177	170	201	212	189	179
	False positive	11	8	14	21	23	12	35	45
	False negative	51	45	22	18	66	78	52	70
	Sensitivity	0.602	0.648	0.828	0.859	0.305	0.179	0.453	0.263
	Specificity	0.942	0.958	0.927	0.89	0.897	0.946	0.844	0.799
	Accuracy	0.806	0.834	0.887	0.878	0.721	0.718	0.727	0.639
	AUC	0.772	0.912	0.955	0.941	0.601	0.717	0.774	0.63
The middle region	True positive	119	97	118	113	63	45	77	76
	True negative	164	172	180	175	186	193	173	170
	False positive	31	23	15	20	25	18	38	41
	False negative	26	48	27	32	66	84	52	53
	Sensitivity	0.821	0.669	0.814	0.779	0.488	0.349	0.597	0.589
	Specificity	0.841	0.882	0.923	0.897	0.882	0.915	0.82	0.806
	Accuracy	0.832	0.791	0.876	0.847	0.732	0.7	0.735	0.724
	AUC	0.831	0.86	0.94	0.896	0.685	0.715	0.811	0.766

372 Continue table

The study area	Parameter	Medium collapsibility				Serious collapsibility			
		SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
The north region	True positive	29	20	44	36	0	0	0	0
	True negative	227	232	214	216	305	305	304	305
	False positive	10	5	23	21	0	0	1	0
	False negative	46	55	31	39	7	7	7	7
	Sensitivity	0.387	0.267	0.587	0.48	0	0	0	0
	Specificity	0.958	0.979	0.903	0.911	1	1	0.997	1
	Accuracy	0.821	0.808	0.827	0.808	0.978	0.978	0.974	0.978
	AUC	0.672	0.778	0.892	0.855	0.5	0.5	0.844	0.838
The east region	True positive	55	30	74	93	1	0	3	1
	True negative	168	176	164	147	290	293	288	290
	False positive	19	11	23	40	3	0	5	3
	False negative	64	89	45	26	12	13	10	12
	Sensitivity	0.462	0.252	0.622	0.782	0.077	0	0.231	0.077
	Specificity	0.898	0.941	0.877	0.786	0.99	1	0.983	0.99

	Accuracy	0.729	0.673	0.778	0.784	0.951	0.958	0.951	0.951
	AUC	0.68	0.709	0.846	0.841	0.533	0.569	0.864	0.86
The west region	True positive	19	13	34	49	0	0	3	0
	True negative	220	236	210	195	297	298	295	298
	False positive	24	8	34	49	1	0	3	0
	False negative	56	62	41	26	21	21	18	21
	Sensitivity	0.253	0.173	0.453	0.653	0	0	0.143	0
	Specificity	0.902	0.967	0.861	0.799	0.997	1	0.99	1
	Accuracy	0.749	0.781	0.765	0.765	0.931	0.934	0.934	0.934
	AUC	0.577	0.73	0.822	0.83	0.498	0.5	0.847	0.835
The middle region	True positive	25	18	29	24	1	0	3	3
	True negative	261	269	257	252	333	334	333	333
	False positive	19	11	23	28	1	0	1	1
	False negative	35	42	31	36	5	6	3	3
	Sensitivity	0.417	0.3	0.483	0.4	0.167	0	0.5	0.5
	Specificity	0.932	0.961	0.918	0.9	0.997	1	0.997	0.997
	Accuracy	0.841	0.844	0.841	0.812	0.982	0.982	0.988	0.988
	AUC	0.674	0.701	0.849	0.81	0.582	0.583	0.698	0.863

(a) AUC value of non-collapsible model

(b) AUC value of slightly collapsible model

(c) AUC value of medium collapsibility model

(d) AUC value of serious collapsible model

Figure 6. The performance comparison of different collapsibility validation models

Similarly, three evaluation statistics, the SD, 95% CI and P-value, are used to evaluate and compare the validation models of collapsible loess in each study area, and the results are shown in Table 6. It can be found that the overall SD, 95% CI and P-value of the RF model are superior to other validation models. Therefore, the RF model is the loess collapsibility validation model with the best prediction ability in each research area.

Table 6. Performance parameters of the validation model

The study area	Evaluation statistics	Non collapsibility				Slight collapsibility			
		SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
The north region	SD	0.024	0.019	0.014	0.016	0.034	0.031	0.021	0.025
	95% CI	0.803	0.868	0.915	0.894	0.648	0.734	0.838	0.769
		to	to	to	to	to	to	to	to
	P-value	0.899	0.941	0.968	0.958	0.781	0.854	0.918	0.868
The east region	SD	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	95% CI	0.033	0.027	0.011	0.019	0.037	0.035	0.027	0.031
		0.723 to	0.811 to	0.931 to	0.86 to	0.528 to	0.632 to	0.744 to	0.661 to
	P-value	0.855	0.918	0.975	0.935	0.673	0.768	0.852	0.782
The west region	SD	<0.0001	<0.0001	<0.0001	<0.0001	0.005	<0.0001	<0.0001	<0.0001
	95% CI	0.029	0.019	0.01	0.012	0.036	0.032	0.026	0.031
		0.715 to	0.876 to	0.934 to	0.917 to	0.53 to	0.654 to	0.722 to	0.569 to
	P-value	0.829	0.948	0.975	0.965	0.673	0.78	0.825	0.691
The middle region	SD	<0.0001	<0.0001	<0.0001	<0.0001	0.004	<0.0001	<0.0001	<0.0001
	95% CI	0.024	0.021	0.013	0.019	0.031	0.03	0.023	0.026
		0.784 to	0.818 to	0.914 to	0.858 to	0.624 to	0.656 to	0.766 to	0.714 to
		0.878	0.902	0.966	0.934	0.746	0.773	0.857	0.818

P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
---------	---------	---------	---------	---------	---------	---------	---------	---------

384 Continue table

The study area	Evaluation statistics	Medium collapsibility				Serious collapsibility			
		SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
The north region	SD	0.04	0.034	0.018	0.023	0.111	0.111	0.844	0.838
	95% CI	0.594	0.712	0.856	0.811	0.283	0.283	0.778	0.751
		to	to	to	to	to	to	to	to
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.002	0.002
The east region	SD	0.033	0.032	0.023	0.023	0.086	0.09	0.053	0.068
	95% CI	0.616 to	0.646 to	0.802 to	0.796 to	0.364 to	0.393 to	0.76 to	0.726 to
		0.745	0.771	0.891	0.885	0.702	0.744	0.968	0.994
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	0.684	0.4	<0.0001	<0.0001
The west region	SD	0.04	0.034	0.023	0.023	0.065	0.065	0.047	0.028
	95% CI	0.5 to	0.663 to	0.776 to	0.784 to	0.371 to	0.372 to	0.755 to	0.781 to
		0.655	0.797	0.868	0.876	0.626	0.628	0.938	0.889
	P-value	0.42	<0.0001	<0.0001	<0.0001	0.979	1	<0.0001	<0.0001
The middle region	SD	0.043	0.043	0.025	0.032	0.132	0.133	0.14	0.094
	95% CI	0.589 to	0.617 to	0.8 to	0.748 to	0.323 to	0.324 to	0.424 to	0.679 to
		0.759	0.785	0.898	0.872	0.841	0.843	0.973	1
	P-value	<0.0001	<0.0001	<0.0001	<0.0001	0.492	0.484	0.096	0.002

385 4.3 Performance and validation of models in the whole study area

386 4.3.1 Performance of models

387 The training data sets of north region, east region, west region and middle region are integrated into the training data
388 sets of Xining city region, and SVM, RSSVM, RF and NBTree models are constructed. Also, the 10-fold cross-
389 validation method is used to reduce the variability of the results. The results of the training model are shown in Table
390 7, and (a) (b) (c) (d) in Figure 7 are the ROC curves of the four collapsibility models. It can be seen that in the training
391 of non-collapsible loess model, RF model has the highest accuracy with a value of 84.8%, followed by NBTree (82%),
392 SVM (81.9%) is very close to NBTree, and RSSVM with an accuracy of 79.7% is still excellent. As for Sensitivity,
393 RF (75.7%) and NBTree (73.3%) models are more excellent. The RSSVM model has the highest specificity value
394 (92.8%), which indicates that 92.8% of the collapsible loess samples are correctly classified, closely followed by RF
395 (90.1%), SVM (89.7%) and NBTree (87.2%). The AUC values of SVM, RSSVM, RF and NBTree models are 79.1%,
396 85.5%, 91.2% and 87.2% respectively, indicating that these four models are all suitable for training data sets, among
397 which RF model is highly suitable.

398 The RF model has the highest accuracy (69.7%) in the training of slight collapsible loess model. All the four models
 399 fails to show good Sensitivity in the whole study area. RSSVM has the highest specificity value (97.3%), indicating
 400 that 97.3% of non-slight collapsible loess has been correctly classified, and NBTree (95%) has a specificity value
 401 second to RSSVM. Using the training data set of the whole study area to train these four models, only the RF model
 402 showed good adaptability, with an AUC value of 73.6%.
 403 In the training of medium collapsible loess model, the accuracy of the four models in the whole study area is similar,
 404 and the data from large to small were RF (77.4%), SVM (76%), NBTree (75.7%) and RSSVM (74.6%).The
 405 Sensitivity indexes of all four models are not satisfactory. As for specificity, all four models performs well, with data
 406 from RSSVM (96.2%), SVM (91.1%), RF (88.4%), and NBTree (82.2%), in descending order.The AUC value of the
 407 RF model reaches 81.8%, indicating that it is well adapted to the training data set of the whole study area, and the
 408 adaptability of the NBTree model (79.8%) is second to the RF.
 409 In the training of serious collapsible loess model, it is found that all four models had extremely low sensitivity,
 410 extremely high specificity and accuracy. This is consistent with the training results of a single regional model. Only
 411 the RF and NBTree models have excellent adaptability to all-region training data, and their AUC values are about
 412 85%.

413 **Table 7.** Model performance of the whole study area

Parameter	Non collapsibility				Slight collapsibility			
	SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
True positive	755	633	833	807	280	87	425	92
True negative	1684	1742	1693	1637	1725	1947	1650	1900
False positive	194	136	185	241	276	54	351	101
False negative	346	468	268	294	698	891	553	886
Sensitivity	0.686	0.575	0.757	0.733	0.286	0.089	0.435	0.094
Specificity	0.897	0.928	0.901	0.872	0.862	0.973	0.825	0.95
Accuracy	0.819	0.797	0.848	0.82	0.673	0.683	0.697	0.669
AUC	0.791	0.855	0.912	0.872	0.574	0.598	0.736	0.665

414 Continue table

Parameter	Medium collapsibility				Serious collapsibility			
	SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
True positive	258	104	361	446	11	2	16	18
True negative	2006	2118	1946	1810	2855	2856	2846	2853
False positive	196	84	256	392	1	0	10	3
False negative	519	673	416	331	112	121	107	105

Sensitivity	0.332	0.134	0.465	0.574	0.089	0.016	0.13	0.146
Specificity	0.911	0.962	0.884	0.822	1	1	0.996	0.999
Accuracy	0.76	0.746	0.774	0.757	0.962	0.959	0.961	0.964
AUC	0.622	0.685	0.818	0.798	0.545	0.544	0.845	0.855

Figure 7. ROC curve of training model

According to the three evaluation statistics in Table 8, the performance of the loess models of non-collapsibility, slight collapsibility, medium collapsibility and serious collapsibility are evaluated with the training data set of the whole study area. The results are shown in the table 8. The RF model is most suitable for data in the training of non-collapsible and slightly collapsible loess, because it has the lowest SD (0.005 and 0.009, respectively), the narrowest

95% CI (0.902 to 0.923 and 0.718 to 0.754, respectively) and the lower P-value ($P < 0.0001$). In the training of medium collapsible loess, the performance of RF and NBTree models are not significantly different, the SD is 0.008, and the 95% CI is 0.802 to 0.833 and 0.782 to 0.815, respectively, and the P-value is $P < 0.0001$. In the training of seriously collapsible loess, NBTree model has better performance, with a SD of 0.017, a 95% CI of 0.822 to 0.887, and a $P < 0.0001$.

Table 8. Performance parameters of the training model

Evaluation statistics	Non collapsibility				Slight collapsibility			
	SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
SD	0.009	0.008	0.005	0.007	0.011	0.011	0.009	0.01
95% CI	0.773 to	0.84 to	0.902 to	0.859 to	0.552 to	0.575 to	0.718 to	0.645 to
	0.81	0.87	0.923	0.885	0.597	0.62	0.754	0.685
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Continue table

Evaluation statistics	Medium collapsibility				Serious collapsibility			
	SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
SD	0.012	0.012	0.008	0.008	0.028	0.028	0.019	0.017
95% CI	0.597 to	0.662 to	0.802 to	0.782 to	0.489 to	0.489 to	0.807 to	0.822 to
	0.646	0.708	0.833	0.815	0.6	0.6	0.882	0.887
p-value	<0.0001	<0.0001	<0.0001	<0.0001	0.094	0.097	<0.0001	<0.0001

4.3.2 Validation of models

The verification data sets of north region, east region, west region and middle region of the city are integrated into the verification data sets of Xining City region. The verification data sets are used to verify the established SVM, RSSVM, RF and NBTree models, aiming to judge the excellence of the model prediction ability. The ROC curves of the four validation models for collapsibility are shown in Figure 8 (a) (b) (c) (d). According to the prediction results of the models in Table 9, in the verification of the non-collapsible loess model, RF (85.8%), NBTree (84.5%), SVM (82.4%) and RSSVM (81%) are of higher accuracy. The Sensitivity of RF model is the highest, with a value of 75.9%, indicating that 75.9% of non-collapsible loess samples are correctly predicted, followed only by the NBTree model (74.5%). For specificity, each model is excellent, ranking from highest to lowest as RSSVM (94.6%), RF (91.7%), SVM (90.5%), and NBTree (90.4%). In terms of overall performance, RF (AUC value is 92.1%) model has the best predictive ability, NBTree (AUC value is 89.1%) model is second, and SVM model has an AUC value of 79.6%. After RS integration algorithm optimization, The AUC value is increased to 87.6%.

443 In the verification of slight collapsible loess model, RF has the highest accuracy (72.6%), and RSSVM has only 2.4%
444 lower accuracy than RF. All the four models shows unsatisfactory Sensitivity, but their specificity is good. The highest
445 Sensitivity is RSSVM (97%), SVM and RF are 87.1% and 85.4%, respectively, and NBTree is the lowest (74.7%).
446 Only the RF (75.8%) model has an AUC of more than 70%, indicating that RF is the best predictor of the four models.
447 In the verification of the loess model with medium collapsibility, the accuracy of the four models is close, from large
448 to small, which are SVM (76.7%), RF (75%), RSSVM (74.8%) and NBTree (73.8%).The sensitivity of the four
449 models is still not qualified, but their specificity is excellent. RSSVM (97.1%) has the highest specificity, follows by
450 SVM (91%), RF (87.4%), and NBTree (79.4%). For predictive power, RF (AUC: 80.6%) and NBTree (AUC: 79.8%)
451 models are similar, and RSSVM (AUC: 71.9%) models are acceptable.
452 In the validation of serious collapsible loess model, RF and NBTree models have excellent predictive ability.
453 Although the sensitivity of the model are extremely low, it has a very high specificity and accuracy rate, with an AUC
454 value of 87.8% and 86.5%, respectively, indicating that the two models owns a similar predictive ability.

455 **Table 9.** Model validation of the whole study area

Parameter	Non collapsibility				Slight collapsibility			
	SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
True positive	325	274	360	353	122	43	176	187
True negative	727	760	736	726	766	853	751	657
False positive	76	43	67	77	113	26	128	222
False negative	149	200	114	121	276	355	222	211
Sensitivity	0.686	0.578	0.759	0.745	0.307	0.108	0.442	0.47
Specificity	0.905	0.946	0.917	0.904	0.871	0.97	0.854	0.747
Accuracy	0.824	0.81	0.858	0.845	0.695	0.702	0.726	0.661
AUC	0.796	0.876	0.921	0.891	0.589	0.631	0.758	0.669

456 Continue table

Parameter	Medium collapsibility				Serious collapsibility			
	SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
True positive	141	60	152	211	7	0	9	10
True negative	839	895	806	732	1221	1227	1213	1219
False positive	83	27	116	190	6	0	14	8
False negative	214	295	203	144	43	50	41	40
Sensitivity	0.397	0.169	0.428	0.594	0.14	0	0.18	0.2
Specificity	0.91	0.971	0.874	0.794	0.995	1	0.989	0.993
Accuracy	0.767	0.748	0.75	0.738	0.962	0.961	0.957	0.962
AUC	0.654	0.719	0.806	0.798	0.568	0.598	0.878	0.865

(a) ROC curve of non-collapsibility

(b) ROC curve of slight collapsibility

(c) ROC curve of medium collapsibility

(d) ROC curve of serious collapsibility

Figure 8. ROC curve of validation model

Using the validation data set in the whole study area, the prediction ability of the models are analyzed and compared through the SD, 95% CI and P-value. The results are shown in Table 10. The RF model still has the best performance in the non-collapsible and slightly collapsible loess verification model, with SD of 0.008 and 0.014, 95% CI of 0.906 to 0.936 and 0.731 to 0.785, respectively, and P-value of $P < 0.0001$. In the verification model of medium collapsible loess, RF and NBTree models have almost the same predictive ability, with SD of 0.013 and CI of 0.781 to 0.831 and 0.773 to 0.822, respectively, and $P < 0.0001$. In the validation model of serious collapse-loess, both RF and NBTree model have excellent predictive ability, among which NBTree model has the smallest SD, the narrowest 95% CI and

P value <0.0001. Compared with the training model, the performance of the validation model is slightly weaker than that of the training model.

Table 10. Performance parameters of the validation model

Evaluation statistics	Non collapsibility				Slight collapsibility			
	SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
SD	0.014	0.011	0.008	0.01	0.018	0.018	0.014	0.016
95% CI	0.768 to	0.855 to	0.906 to	0.872 to	0.554 to	0.597 to	0.731 to	0.638 to
	0.823	0.897	0.936	0.91	0.624	0.666	0.785	0.7
P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Continue table

Evaluation statistics	Medium collapsibility				Serious collapsibility			
	SVM	RSSVM	RF	NBTree	SVM	RSSVM	RF	NBTree
SD	0.018	0.017	0.013	0.13	0.046	0.047	0.027	0.024
95% CI	0.617 to	0.686 to	0.781 to	0.773 to	0.478 to	0.507 to	0.826 to	0.817 to
	0.69	0.753	0.831	0.822	0.657	0.69	0.931	0.912
P-value	<0.0001	<0.0001	<0.0001	<0.0001	0.105	0.018	<0.0001	<0.0001

4.4 Prediction of loess collapsibility in Xining city

By comparing and analyzing the training and verification results of the above models, it can be found that the RF model as a whole can have almost the highest AUC value, no matter in the north and east of the city, or in the west and middle of the city, or even in the whole research area. In other words, the RF model shows almost the strongest predictive ability in the four models. Therefore, in this study, RF model is adopted to predict one project in each research area, and the predicted results are compared with the measured results. The comparison results are shown in Table 11. A total of 392 loess samples are predicted in the north part of the city, among which 112 non-collapsible loess samples are predicted, 76 slight collapsible loess samples are predicted, 73 loess samples are predicted as medium collapsible loess, and only 25 loess samples are predicted as serious collapsible loess. In the eastern part of the city, there are 340 predicted samples, among which 51 are non-collapsible loess, 91 are predicted to be slightly collapsible loess, 134 to be medium collapsible loess and 7 to be seriously collapsible loess. Among the 384 predicted samples in the west part of the city, 111 loess samples are classified as non-collapsible loess, 127 loess samples are slightly collapsible, 92 loess samples are predicted as medium collapsible, and 2 loess samples are seriously collapsible. Finally, among the 288 predicted samples in the middle region, the predicted non-collapsible loess accounts for the majority, with 194 and 58 slightly collapsible loess samples. Only 28 loess samples are classified as

488 medium collapsible, and there is no serious collapsible loess.

489 **Table 11.** Comparison of measured and predicted results of the RF model

The study area	Number of prediction data sets	Non	Slight	Medium	Serious
		collapsibility	collapsibility	collapsibility	collapsibility
		Predicted / Measured	Predicted / Measured	Predicted / Measured	Predicted / Measured
The north region	392	112/114	76/93	73/126	25/59
The east region	340	51/65	91/119	134/142	7/14
The west region	384	111/222	127/82	92/77	2/3
The middle region	288	194/203	58/59	28/26	0/0
The whole study area	1404	427/604	390/353	362/371	33/76

490 The predicted soil samples of the four study areas are integrated, and the RF model is used to predict the collapsibility
 491 of loess for the four projects in the study area. It can be seen from Table 12 that, among the total 1404 loess samples,
 492 427 non-collapsible loess samples are predicted, while 468 non-collapsible loess samples are predicted in four
 493 separate regions, with similar results in comparison. Similarly, 390 pieces of slightly collapsible loess are predicted,
 494 while 352 pieces of slightly collapsible loess are predicted in the four small study areas. It can be seen that the number
 495 of predicted results in small areas is closer to the number of measured results. In the prediction of medium collapsible
 496 loess, 362 are predicted from the whole study area, and 327 are predicted from the single area, 44 less than the
 497 measured results. In the prediction of serious collapsible loess, except for the western and middle parts of the city,
 498 the prediction results of serious collapsible loess in other areas are only about half of the measured results, which
 499 may be caused by too few data sets of the serious collapsible loess model. Therefore, more data of seriously
 500 collapsible loess will be collected in the next study for further exploration.

501 As can be seen from Table 12, in the prediction of non-collapsible loess, the RF model performance in a single region
 502 (average AUC is 94.7%) differs little from that of the RF model in the whole study area (AUC value is 92.1%). The
 503 average AUC of the RF model of slight collapsibility in a single region is 81.5%, and that of the RF model of medium
 504 collapsibility loess is 85.2%. However, the AUC value of the RF model of slight collapsibility in the whole study
 505 area is 75.8%, and that of the RF model of medium collapsibility loess is 80.6%. It can be seen that in the prediction
 506 of slight collapsible loess and medium collapsible loess, the RF model in a single region is significantly better than
 507 the RF model in the whole study area. However, in the prediction of serious collapsible loess, the predictive ability of
 508 RF models in the whole study area (AUC value is 87.8%) is obviously better than that of RF models in a single area
 509 (AUC mean value is 81.3%).

Table 12. AUC value for RF prediction model

The study area	Non collapsibility	Slight collapsibility	Medium collapsibility	Serious collapsibility
The north region	0.941	0.878	0.892	0.844
The east region	0.953	0.798	0.846	0.864
The west region	0.955	0.774	0.822	0.847
The middle region	0.94	0.811	0.849	0.698
Average value	0.947	0.815	0.852	0.813
The whole study area	0.921	0.758	0.806	0.878

5. Conclusion

SVM is a popular and effective machine learning method in loess collapsibility modeling. Accordingly, the random subspace (RS) and SVM are coupled to explore the performance of the integration model. RF and NBTree are relatively new machine learning methods, which have not been used to predict loess collapsibility in the past. This study compares the predictive performance of SVM, RSSVM, RF and NBTree in loess collapsibility in Xining city. The main conclusions are as follows:

- (1) Based on grey relational degree analysis and multicollinearity analysis, the main influences of loess collapsibility in the urban area of Xining are burial depth, water content, specific gravity of soil particles, void rate, geostatic stress and plasticity limit.
- (2) RF model is more efficient in predicting the collapsibility of loess in Xining.
- (3) In predicting non-collapsible loess, slight and medium collapsible loess, RF models in a single region is better than the RF model in the whole study area. However, for predicting serious collapsible loess, the RF model in the whole study area is better than that in a single area.
- (4) The performance of SVM, RSSVM, RF and NBTree in different research areas, non-collapsible loess models are better than collapsible loess model.
- (5) The integrated model (RSSVM) is better than the benchmark model (SVM) in loess collapsibility prediction.

Author Contributions

Qifei Zhao, Xiaojun Li, Yunning Cao, Zhikun Li and Jixin Fan contributed to the work. Xiaojun Li, Yunning Cao and Jixin Fan collected field data. Qifei Zhao conducted the analysis. Zhikun Li provided assistance. Qifei Zhao wrote and edited the manuscript. All the authors discussed the results and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

534 This study is jointly supported by Key Laboratory for Geo-hazards in Loess area, MLR (KLGLAMLR201504) and
535 State key Laboratory of Loess and Quaternary Geology (SKLLQG1612).

536 **Acknowledgments**

537 The authors specially wish to thank Wei Chen for the technical guidance.

538 **Conflicts of Interest**

539 The authors declare no conflicts of interest.

540 **References**

- 541 1. S, L.Y., 1998. Evaluation of collapsible loess foundation. *Journal of Geotechnical Engineering*.
- 542 2. Shuyan, Q., 2011. Study on engineering geological characteristics of collapsible loess in qinghai. *Qinghai science and technology*,
543 3.
- 544 3. Jing Yanlin, P.Y., CUI Zhongxing, 2006. Intelligent evaluation of collaptibility of loess.
- 545 4. Han Xiaomeng, W.J., WANG Yu, 2008. ANFIS method in the evaluation of loess collapsibility of high-speed railway foundation.
546 *Geographic Sciences*, 28, 833-837.
- 547 5. Anning, 2011. Prediction of loess Collapsibility Coefficient based on PCA and ANFIS. *Journal of Shenzhen Vocational and*
548 *Technical College*, 10, 28-31.
- 549 6. Ma Yan, W.J., PENG Shujun, 2016. Relationship between loess collapsibility and soil property index and its prediction model.
550 *Water and Soil Conservation Bulletin*, 36, 120-128.
- 551 7. Zhang, S.-h., 2011. Assessment of loess collapsibility with GRNN, *Communication Systems and Information Technology*.
552 Springer, pp. 745-752.
- 553 8. Li, Y., Zhong, T., 2010. The application of support vector machine in evaluation of loess collapsibility, 2010 International
554 Conference on Artificial Intelligence and Education (ICAIE). IEEE, pp. 203-205.
- 555 9. Jing Yanlin, B.Y., Lin Du Jun, 2010. Prediction mining of collapbility of loess based on Least-squares Support Vector Machine.
556 *Rock and Soil Mechanics*, 31, 1865-1870.
- 557 10. Lai Chengguang, C.X., ZHAO Shiwei, 2015. Flood risk Assessment Model based on Random Forest and its Application. *Chinese*
558 *Journal of Water Resources*, 46, 58-66.
- 559 11. Chen, W. et al., 2020. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree,
560 and random forest methods. 701, 134979.
- 561 12. Ge Jike, C.D., WANG Wenhe, 2019. Fire classification based on improved Naive Bayesian Classification algorithm. *Journal of*
562 *Safety and Environment*, 4.
- 563 13. Li De, C.W., Le Zhangyan, 2020. Prediction of Dangshansu pear flowering date based on random forest algorithm and
564 meteorological factors. *Transactions of the Chinese Society of Agricultural Engineering*, 36.
- 565 14. Chen, W. et al., 2017a. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree
566 classifiers for a landslide susceptibility assessment in Langao County, China. 8, 1955-1977.
- 567 15. Chen, W. et al., 2017b. GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression,
568 Naïve-Bayes tree, and alternating decision tree models. 8, 950-973.
- 569 16. Chen, W. et al., 2017c. A comparative study of logistic model tree, random forest, and classification and regression tree models
570 for spatial prediction of landslide susceptibility. 151, 147-160.
- 571 17. Chen, W., Tsangaratos, P., Ilia, I., Duan, Z., Chen, X.J.S.o.T.T.E., 2019c. Groundwater spring potential mapping using population-
572 based evolutionary algorithms and data mining methods. 684, 31-49.

- 573 18. Li, Y., Chen, W.J.W., 2020. Landslide susceptibility evaluation using hybrid integration of evidential belief function and machine
574 learning techniques. 12, 113.
- 575 19. Board, E.G.H.E., 2007. Handbook of engineering geology (fourth edition), Beijing: China Architecture & Building Press.
- 576 20. Ju Long, D., 1982. Control problems of grey systems. Systems control letters, 1, 288-294.
- 577 21. O'brien, R.M.J.Q., quantity, 2007. A caution regarding rules of thumb for variance inflation factors. 41, 673-690.
- 578 22. Cherkassky, V., Mulier, F.M., 2007. Learning from data: concepts, theory, and methods. John Wiley & Sons.
- 579 23. Hong, H., Naghibi, S.A., Pourghasemi, H.R., Pradhan, B.J.A.J.o.G., 2016. GIS-based landslide spatial modeling in Ganzhou City,
580 China. 9, 112.
- 581 24. Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., 2012. Landslide susceptibility assessment in vietnam using support vector
582 machines, decision tree, and Naive Bayes Models. Mathematical problems in Engineering, 2012.
- 583 25. Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE transactions on pattern analysis machine
584 intelligence, 20, 832-844.
- 585 26. Chandra, S., Maheshkar, S.J.M.T., Applications, 2017. Verification of static signature pattern based on random subspace, REP
586 tree and bagging. 76, 19139-19171.
- 587 27. Chen, W. et al., 2019a. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging
588 and random subspace ensembles. 575, 864-873.
- 589 28. Xia, J. et al., 2015. Random subspace ensembles for hyperspectral image classification with extended morphological attribute
590 profiles. 53, 4768-4786.
- 591 29. Breiman, L., 2001. Random forests. Machine learning, 45, 5-32.
- 592 30. Liaw, A., Wiener, M.J.R.n., 2002. Classification and regression by randomForest. 2, 18-22.
- 593 31. Kohestani, V., Hassanlourad, M., Ardakani, A.J.N.H., 2015. Evaluation of liquefaction potential based on CPT data using random
594 forest. 79, 1079-1089.
- 595 32. Kohavi, R., 1996. Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid, Kdd, pp. 202-207.
- 596 33. Quinlan, J.R., 2014. C4. 5: programs for machine learning. Elsevier.
- 597 34. Chen, W. et al., 2019b. Spatial prediction of groundwater potentiality using ANFIS ensembled with teaching-learning-based and
598 biogeography-based optimization. 572, 435-448.
- 599 35. Bui, D.T. et al., 2017. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent
600 approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides Arabian
601 Journal of Geosciences, 14, 447-458.
- 602 36. Li Ping, L.T., 2007. The relationship between physical properties and collapsibility of loess and its engineering significance.
603 Journal of Engineering Geology, 15, 506-512.
- 604 37. Cui Jing-yu, Z.Y.-y., XIE Banglong, 2019. Research on the collapsibility of undisturbed loess at different depths in xining area.
605 Chinese journal of geotechnical engineering, 41, 249-252.
- 606 38. Ying, G., Yanxia, M., Wuyu, Z., Jiaqing, G., 2019. Analysis of humidifying deformation characteristics and microstructure of
607 loess in Xining area. Journal of Engineering Geology, 27, 803-810.

Figures

Figure 1

Geographic location of the data collection Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 2

The collapsibility grade distribution of loess samples in different study areas

Figure 3

Flowchart of the study

Figure 4

please see the manuscript file for the full caption

(a) AUC value of non-collapsible model

(b) AUC value of slightly collapsible model

(c) AUC value of medium collapsibility model

(d) AUC value of serious collapsible model

Figure 5

The performance comparison of different collapsibility training models

(a) AUC value of non-collapsible model

(b) AUC value of slightly collapsible model

(c) AUC value of medium collapsibility model

(d) AUC value of serious collapsible model

Figure 6

The performance comparison of different collapsibility validation models

(a) ROC curve of non-collapsibility

(b) ROC curve of slight collapsibility

(c) ROC curve of medium collapsibility

(d) ROC curve of serious collapsibility

Figure 7

ROC curve of training model

(a) ROC curve of non-collapsibility

(b) ROC curve of slight collapsibility

(c) ROC curve of medium collapsibility

(d) ROC curve of serious collapsibility

Figure 8

ROC curve of validation model