

Research on and Application of Tunnel Structure Defects Prediction Using Machine Learning Methods

Bo SHI^{a,b,c,d}, Hui SU^{a,c,d}, Xu DU^{a,c,d}, Bao JIAO^{a,c,d}, Lin WANG^{a,c,d,e,1}

^aSGIDI Engineering Consulting (Group) Co., Ltd, Shanghai 200093, China

^bSchool of Computer Science and Technology, Fudan University, Shanghai 2000438, China

^cShanghai Engineering Research Center of Geotechnical Test for Underground Space, Shanghai 200093, China

^dShanghai Professional Technology Service Platform of Geotechnical Engineering, Shanghai 200093, China

^eDepartment of Geotechnical Engineering, College of Civil Engineering, Tongji University, Shanghai 200092, China

Abstract. With the rapid development of underground engineering in China, more metro tunnels are being constructed, the mileage of subway tunnels is increasing, and the corresponding problems of tunnel structure diseases are becoming more prominent. At present, the treatment of tunnel structural diseases mainly relies on manual inspection and identification, and research on defects prediction is still lacking. Because of the complexity of the factors affecting tunnel structure diseases, it is difficult to analyze the causes and development trend of the diseases comprehensively by manual analysis. Fortunately, machine learning methods have gained popularity in classification and regression tasks in recent decades. Many algorithms, such as decision tree algorithms, the random forest algorithm, and XGBoost, have been applied in fields including finance, engineering, and transportation. This study aimed to analyze the prediction effect of machine learning models by feeding 68055 segment lining rings of six subway lines in a city. According to the disease records from 2014 to 2016 and the corresponding convergence and characteristic data, defect conditions in 2017 were predicted and compared with real defect conditions in 2017. The accuracy rates and F1 values of the predicted results were all above 80%. The prediction results can help tunnel maintenance departments and relevant government regulators make auxiliary decisions to control tunnel structure diseases, and can help them focus on the tunnel interval of severe diseases to clarify the development trend of tunnel disease.

Keywords: Tunnel engineering, operational tunnel structure safety, tunnel structure disease prediction, machine learning

1. Introduction

With the rapid development of China's economy and the continuous acceleration of urban construction, underground engineering is in a stage of rapid development, and the

¹Corresponding author: Lin Wang, SGIDI Engineering Consulting (Group) Co., Ltd, Shanghai 200093, China; E-mail: 1410024@tongji.edu.cn.

corresponding tunnel construction is also increasing. Increased operation time, coupled with the influence of tunnel construction conditions, the surrounding tunnel environment, and other factors, have made various tunnel diseases more prominent, especially in soft soil areas, such as Shanghai, Hangzhou, and other cities in China. Owing to the particularity of soft soil, leakage, segment damage, cracks, and other situations are more significant. Serious tunnel structure damage can lead to severe accidents, thereby affecting the safe operation of the subway and seriously threatening the safety of life and property.

However, tunnel structure safety is affected by many factors, such as geology, design, construction, and operation, and the causes of diseases are complex [1]. Yuan et al. defined a quantitative assessment of the service states of the structure in an operational tunnel and proposed a framework of assessment procedure. Previous studies on tunnel assessments mostly used data mining technology to evaluate the state of tunnel structure and put forward corresponding measures [2]. Alimoradi et al. proposed that an appropriately trained neural network could reliably predict weak geological zones in front of a tunnel face accurately [3]. Adoko et al. used Multivariate Adaptive Regression Spline (MARS) and ANN models to predict the diameter convergence of a high-speed railway tunnel in weak rock and concluded that MARS constitutes a reliable alternative to ANN in modeling geo-engineering problems [4]. Mahdevvari et al. developed a dynamic model based on Support Vector Machines (SVMs) for the prediction of convergence in the tunnel, and got good results in situ measured ones [5]. Feng et al. proposed a Bayesian approach to improve time-dependent convergence predictions and the results showed that the proposed method could help improve the accuracy of predictions and reduce their uncertainty [6].

However, previous studies concerned with the deformation prediction of the ring segments such as convergence, and these studies lacked a forecast of the development trend of tunnel structure defects based on a large amount of engineering data. Unlike predicting the deformation in tunnels in previous studies, the paper first provided an idea for predicting defects and defect types using machine learning methods. At present, the tunnel maintenance department in China has accumulated a large amount of disease information according to regular tunnel inspection. Unfortunately, the use of the data is limited to basic inquiry and statistical work, so more intelligent methods are needed to predict the development trend of the disease. Accurately and efficiently predicting possible tunnel diseases has important engineering guiding significance to help tunnel maintenance departments determine the inspection scope and avoid major disasters.

In this study, three machine learning methods, the decision tree, random forest, and XGBoost, were applied to the prediction of tunnel structure diseases to help the tunnel maintenance department understand the development trend of the disease and make timely decisions on prevention and control measures.

2. Methodology

2.1. Decision Tree

Decision tree is a machine learning prediction model that represents a mapping relationship between features and target values. It is a decision analysis method that is used to obtain the expected probability so as to evaluate risk and judge its feasibility by forming a decision tree based on the known probability of various situations. There are

several decision tree algorithms, including Iterative Dichotomiser 3 (ID3), C4.5 and Classification and Regression Tree (CART) [7-9]. Its basic principle is as follows.

For data feature vectors $x_i \in R^n, i=1, \dots, l$ and a label vector $y \in R^l$, a decision tree recursively partitions the feature space, such that the samples with the same labels or similar target values are grouped together. A node, m , can be represented by Q_m with N_m samples and Q_m is the subset of the whole dataset. For each candidate split, $\theta = (j, t_m)$, consisting of a feature, j , and threshold, t_m , partition the data into $Q_m^{left}(\theta)$ and $Q_m^{right}(\theta)$ subsets, which are calculated as follows:

$$Q_m^{left}(\theta) = \{(x, y) | x_j \leq t_m\} \tag{1}$$

$$Q_m^{right}(\theta) = \{(x, y) | x_j > t_m\} \tag{2}$$

If a target is a classification outcome taking on values $0, 1, \dots, K-1$ for node m , we can calculate p_{mk} (equation (3)) as the proportion of class k observations in node m . Specially, function $I(y=k)$ in equation (3) is called the Indicator function, which can be calculated as in equation (4).

$$p_{mk} = 1/N_m \sum_{y \in Q_m} I(y=k) \tag{3}$$

$$I(y=k) = \begin{cases} 1 & \text{if } y = k \\ 0 & \text{if } y \neq k \end{cases} \tag{4}$$

Subsequently, common measures of impurity (loss function), including ‘Gini index’ in equation (5) and ‘Entropy’ in equation (6) [10], are calculated as follows:

$$H(Q_m) = \sum_k p_{mk} (1 - p_{mk}) \tag{5}$$

$$H(Q_m) = -\sum_k p_{mk} \log(p_{mk}) \tag{6}$$

Then, the quality of a candidate split of node m can be computed using the loss function (equation (7)) [10].

$$G(Q_m, \theta) = \frac{N_m^{left}}{N_m} H(Q_m^{left}(\theta)) + \frac{N_m^{right}}{N_m} H(Q_m^{right}(\theta)) \tag{7}$$

Finally, choose proper parameters to minimize the loss function in equation (8), and recurse for subsets $Q_m^{left}(\theta^*)$ and $Q_m^{right}(\theta^*)$ until the node cannot be split or the given maximum depth is reached.

$$\theta^* = \arg \min_{\theta} G(Q_m, \theta) \quad (8)$$

The sketch map of a decision tree is shown in Figure 1. For a given dataset, it can be recursively split into 'tree' as the sketch map.

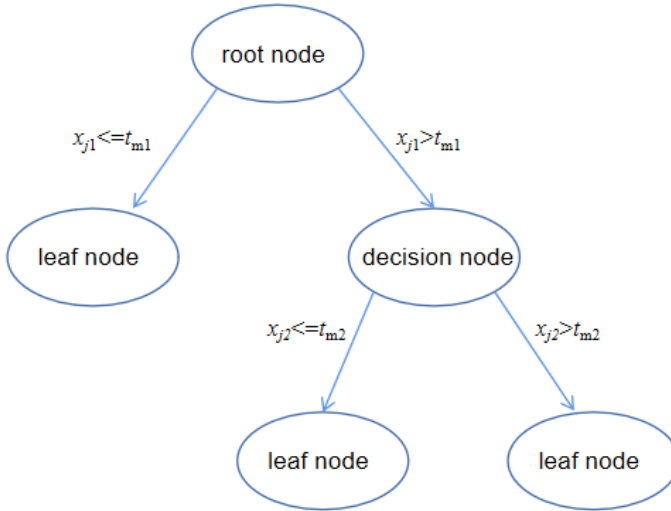


Figure 1. Sketch map of a decision tree.

2.2. Random Forest

Random forest [11,12] is an ensemble method that combines several base estimators to improve accuracy and stability over a single estimator. Furthermore, the random forest algorithm builds several estimators independently and averages their predictions, which is called the bagging method and can reduce the variance of a base estimator by introducing randomization into its construction procedure [13,14,15]. In this study, CART was used as the base estimator, and "Gini" was used as the loss function.

The process of the random forest algorithm is as follows. Firstly, extract m training samples from the original data set using the Bootstrap method; n sub training sets are obtained by carrying out put-back sampling n times. Secondly, select p attributes from m attributes on each training set to generate n CART decision trees. Thirdly, n CART decision trees are formed into a random forest. For the classification problem, the majority voting method is adopted, and the final classification result is determined by voting according to the classifier of n trees.

2.3. XGBoost

XGBoost is also an ensemble method that combines several base estimators [16]. However, unlike the random forest algorithm and bagging methods, XGBoost is a boosting method, and the base estimators are built sequentially to reduce the bias of the combined estimators.

The regularized objective function is as follows:

$$L(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \text{ s. t. } \Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (9)$$

Here, l is a differentiable convex loss function that measures the difference between the prediction, y_i , and the target, \hat{y}_i . The second term, Ω , penalizes the complexity of the model, which can help avoid over-fitting. Each f_k corresponds to an independent tree structure.

When we predict the i -th instance at the t -th iteration, the prediction can be written as:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x) \quad (10)$$

Then, we minimize the following objective:

$$L^{(t)} = \sum_{i=1} l(y_i, (\hat{y}_i^{(t-1)} + f_t(x_i))) + \Omega(f_t) \quad (11)$$

Second-order approximation can be used to quickly optimize the objective above. Then, we can get the following objective:

$$L^{(t)} = \sum_{i=1} \left[l(y_i, (\hat{y}_i^{(t-1)})) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (12)$$

where g_i and h_i are first- and second-order gradient statistics on the loss function.

Obviously, the regularization term is introduced to control the complexity of the model and avoid over-fitting, and we can define the objective function as needed.

3. Model Construction and Application

In this study, data of 68,055 segment lining rings of six subway lines in a city were fed into the machine learning models mentioned above. According to defects records from 2014 to 2016 and corresponding convergence and characteristic data, defects conditions in 2017 were predicted and compared with real defects conditions in 2017 to analyze the prediction effect of the models.

3.1. Data Preprocessing

In machine learning field, the data determines the upper limit of machine learning, and the algorithm is just as close to that limit as possible. Thus, the selection of features is very important for machine learning models. Fortunately, although tunnel structure defects are affected by multiple factors, we can get the features base on engineering experience.

Combined with engineering experience, various factors including convergence, buried depth, initial convergence, PS value and sensitivity value of the soil layer, geological division, hydrogeological division, cover condition and open time were determined and corresponding data were collected (as shown in Table 1, Table 2, and Table 3).

Table 1. Sample data of tunnel segments' time-invariant features.

ring_id	263517	260196	260012	148570	321966	9535	207927
ps_value	0.4679	0.7598	0.599	0.6995	0.5012	0.5677	1.6795
buried_depth(m)	8.103	8.655	6.714	14.201	9.071	11.842	24.258
init_convergence(m)	0.0200	0.0183	0.0276	0.0256	0.0347	0.034	0.0280
Sensitivity_value	4.4738	1.9422	2.0805	4.3165	2.5142	2.5011	1.9164
open_time(year)	1993	1993	1993	2010	2007	2010	2013
Relationships_with_bypass	A	E	A	E	E	E	E
cover_condition	road	road	park	water	yard	road	buiding
geological_division	IA	IIB	IIB	IIA	IIIAE	IIIBE	IIB
hydrogeological_division	first	fourth	fourth	third	seventh	fifth	sixth

Table 2. Sample data of tunnel segments with diseases.

Ring number	Defect types	Discovery date of defect
263517	leakage	2014
263517	leakage	2017
263518	leakage	2014
263518	leakage	2017
263592	leakage	2014
263736	cracking	2017
264061	leakage	2014
264064	leakage	2015
264139	spalling	2014
264139	spalling	2016
264139	leakage	2016

Table 3. Sample data of tunnel segments' time-variant features (inner diameter and deformation against inner diameter).

Ring number	Measure time	Inner diameter(m)	Deformation(m)
263517	2014	5.55417	0.05417
263517	2015	5.55218	0.05218
263517	2016	5.5514	0.0514
263517	2017	5.5595	0.0595
263518	2014	5.55417	0.05417
263518	2015	5.55218	0.05218
263518	2016	5.5514	0.0514
263518	2017	5.5595	0.0595
263521	2014	5.54813	0.04813
263521	2015	5.54602	0.04602
263521	2016	5.547	0.047
263521	2017	5.5467	0.0467

Before training machine learning models, the data need to be preprocessed, and continuous variables need to be normalized. Normalization refers to limiting features' data within the range of 0 to 1. The calculation is as follows:

$$X = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (13)$$

where X_i denotes structural feature data before normalization, X_{\max} , X_{\min} denote the maximum and minimum values of the feature, respectively; and X denotes the normalized structural feature data.

Categorical variables are processed with One-Hot Encoding. One-Hot Encoding uses 0 and 1 to represent parameters and N status for a state code. For example, tunnel segment lining rings in the geological division zone can be divided into four district, and the first district can be expressed as [1,0,0,0] and the second district can be expressed as [0,1,0,0], the third district and fourth district can be expressed as the same way.

In the paper, convergence, buried depth, initial convergence, open time, PS value and sensitivity value of the soil layer were calculated as continuous variables and geological division, hydrogeological division, cover condition were calculated as categorical variables.

3.2. Model Training and Evaluation

After data processing, the data were divided into a training set (data from 2014–2016) and test set (data from 2017). Then the training set was fed into Decision Tree model, Random Forest model and XGBoost model for training and the test set was used for evaluating the models. Additionally, we used the module called GridSearchCV [17] from scikit-learn to tune the hyper-parameters of the three machine learning models.

Model evaluation generally includes accuracy rate, recall rate, and F1 value in classification tasks. The confusion matrix has the following four markers in binary classification in Table 4. True Positive (TP) means predicting a positive sample as a positive sample; False Negative (FN) means predicting a positive sample as a negative sample; False Positive (FP) means predicting a negative sample as a positive sample; and True Negative (TN) means predicting a negative sample as a negative sample.

Table 4. Four markers in binary classification results.

Predict (column) / label (row)	Positive	Negative
Positive	TP	FP
Negative	FN	TN

The precision rate is the ratio of the correct prediction among predicted positive samples. It can be calculated as follows:

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

The recall rate is the ratio of the correct prediction among positive label samples. It can be calculated as:

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

The F1 value is the comprehensive evaluation of precision rate and recall rate. It is the harmonic average value of the two values, and its calculation is as follows:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (16)$$

Receiver operating characteristic curve (ROC curve) is another way to evaluate machine learning model. For a ROC curve of machine learning model, x axis denotes the False Positive rate and y axis denotes the True Positive rate. As the False Positive rate is only related to negative samples and the True Positive rate is only related to positive samples, ROC curve doesn't change as the class distribution (the ratio of the positive samples and the negative samples) changes, that is, it is very suitable for evaluating models of imbalance dataset. AUC is the area under ROC curve and it shows the quality of a classifier.

3.3. Results and Discussion

As most of the tunnel segment lining rings were free of disease, the accuracy could not truly reflect the effect of the model. Therefore, we randomly selected a subset of instances with no defect to match the number of instances with defects to reduce the imbalance of the dataset and used four indexes, accuracy, precision, recall, and F1 values, to comprehensively evaluate the effect of the model. We compared the real condition of tunnel segment defects with the predicted condition to generate the confusion matrix (Figures 2–4), which represents the effect of the models. We also plotted the ROC curves of the three models and the corresponding AUC values were calculated in the Figure 5.

In this study, the segment lining rings' diseases were set as positive samples. In Figures 2–4, the horizontal axis denotes the label of predicted value, and the ordinate denotes the label of true value, in which 0 means no disease and 1 means disease. As the evaluation values in Table 5 show, the training model worked well. The results of the three models were very close, and all demonstrated quite accurate predictions on the tunnel diseases; the accuracy rates were all above 90%, and the F1 values were all above 80%. As we can see, we have got high accuracy rate and the models still leave room for improvement according to the results of recall rate. From ROC curves and AUC values in Figure 5, we can see that all models got high AUC values which means the algorithms we used are robust though the dataset is imbalanced. Furthermore, we can see that XGBoost model and Random Forest model are better than Decision Tree model as the ensemble learning methods are superior to a single estimator.

Table 5. Evaluation indexes of the machine learning models.

Algorithms	Accuracy	Precision	Recall	F1 value
Decision tree	93%	98.4%	74%	83%
Random forest	93%	98.4%	69.5%	81%
XGBoost	93%	99.8%	69%	81.4%

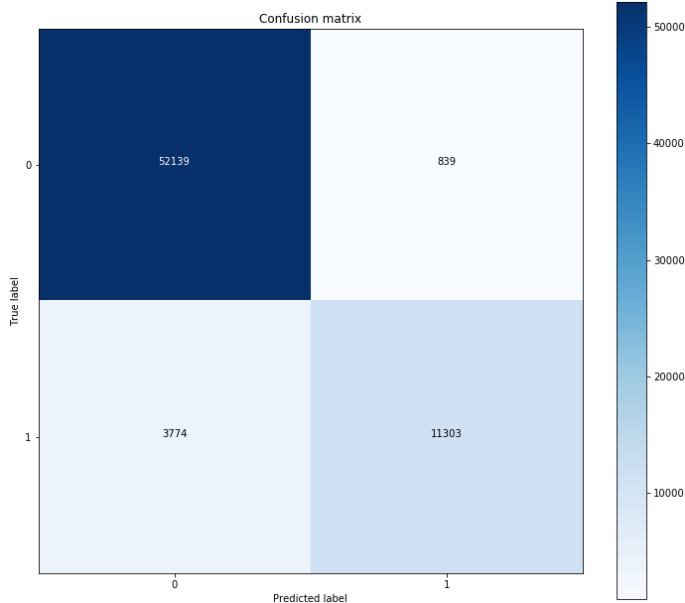


Figure 2. Prediction results of Decision Tree model of binary classification (confusion matrix).

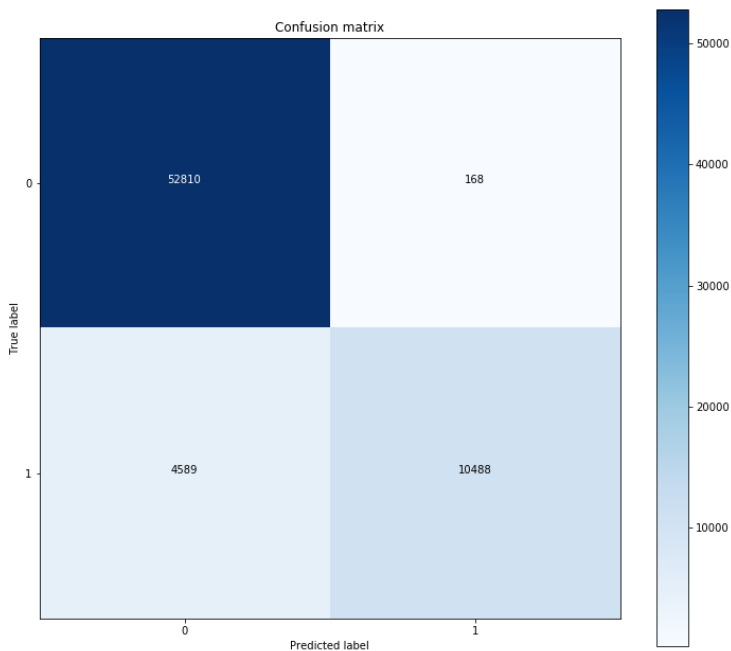


Figure 3. Prediction results of Random Forest model of binary classification (confusion matrix).

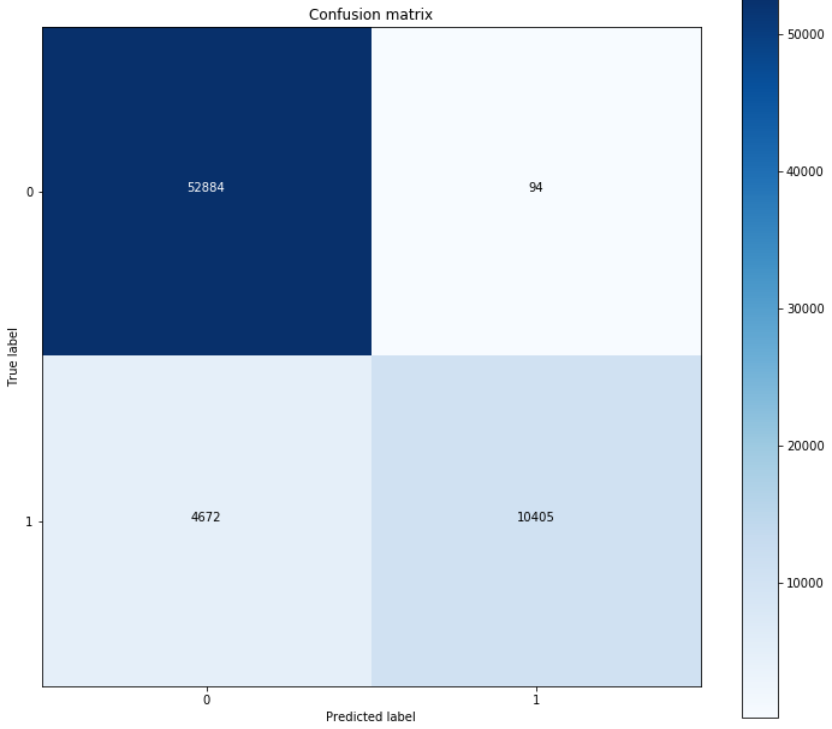


Figure 4. Prediction results of the XGBoost model of binary classification (confusion matrix).

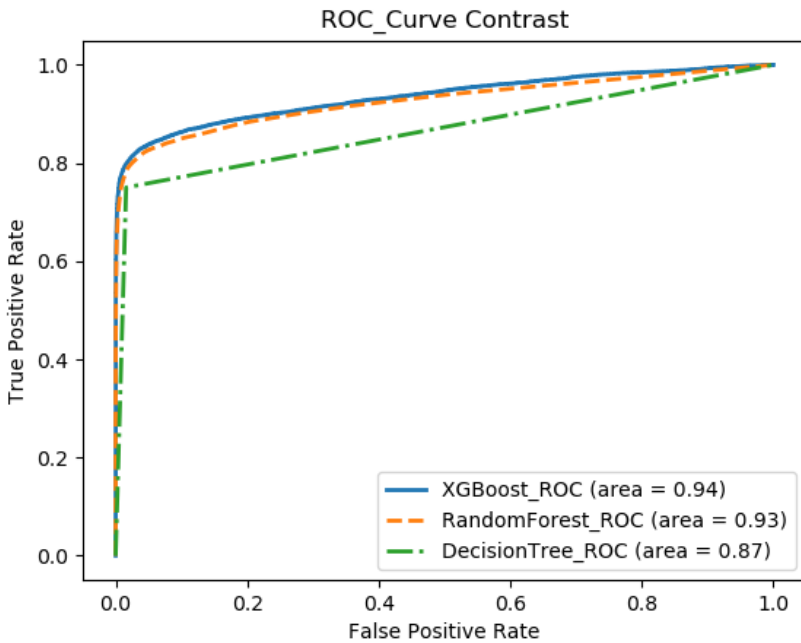


Figure 5. ROC curve and AUC values of the three models.

4. Conclusion

Given the complexity of the influencing factors of metro tunnel structural diseases, the study used three machine learning methods to predict the occurrence of the defects in metro tunnels. Through the three models, we predicted tunnel structural defects and obtained satisfactory results. The accuracy rates were all above 90%, and the F1 values were all above 80%. It is important to note that it is nearly impossible to achieve completely accurate prediction using machine learning methods. Furthermore, ROC curve of the three models were plotted and AUC values were calculated. The AUC values of XGBoost model, Random Forest model, Decision Tree model are 0.94, 0.93 and 0.87 respectively which means that the algorithms we chose are robust for imbalanced engineering data. Overall, our proposed method can be useful to aid tunnel maintenance departments owing to the high prediction accuracy rate.

The prediction results can provide auxiliary decision-making assistance for tunnel maintenance departments and relevant government regulatory departments to prevent and control tunnel structural diseases and focus on tunnel sections where serious diseases may occur so as to further clarify the development trend of tunnel diseases. This study provides a new idea for the prediction of tunnel disease, but the current research can only give the probability of the occurrence of the disease (binary classification). Subsequent research can be extended to the prediction of the occurrence of the disease (multiple classification) and its probability, which will provide more accurate auxiliary guidance for tunnel maintenance and supervision authorities.

Acknowledgements

The authors would thank Social Development Projects of Shanghai ‘Science and Technology Innovation Action Plan’ (18DZ1205900, 19DZ1200801) and National Natural Science Foundation of China (42002272) for support.

References

- [1] Chen X, Zhu H, Li X, Lin X and Wang X. Probabilistic performance prediction of shield tunnels in operation through data mining. *Sustainable Cities and Society*. 2019; 44: 819-29.
- [2] Yuan Y, Bai Y and Liu J. Assessment service state of tunnel structure. *Tunn. and Undergr. Space Technol.* 2012; 27: 72-85.
- [3] Alimoradi A, Moradzadeh A, Naderi R, Salehi M Z and Etemadi A. Prediction of geological hazardous zones in front of a tunnel face using TSP-203 and artificial neural networks. *Tunn. and Undergr. Space Technol.* 2008; 23:711-7.
- [4] Adoko A, Jiao Y, Wu L, Wang H and Wang Z. Predicting tunnel convergence using Multivariate Adaptive Regression Spline and Artificial Neural Network. *Tunn. and Undergr. Space Technol.* 2013; 38:368-76.
- [5] Mahdevari S, Hamid S and Seyed R. A dynamically approach based on SVM algorithm for prediction of tunnel convergence during excavation. *Tunn. and Undergr. Space Technol.* 2013; 38: 59-68.
- [6] Feng X, Jimenez R, Zeng P and Senent S. Prediction of time-dependent tunnel convergences using a Bayesian updating approach. *Tunn. and Undergr. Space Technol.* 2019; 94: 103-118.
- [7] Breiman L, Friedman J, Olshen R, and Stone C. *Classification and regression trees*. Wadsworth, Belmont, CA. 1984.
- [8] Quinlan J. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers. 1993.
- [9] Hastie, Tibshirani R and Friedman J. *Elements of Statistical Learning*. Springer. 2009.
- [10] Quinlan J. Introduction of decision trees. *Machine Learning*. 1986; 1:81-106.
- [11] Breiman L. Bagging predictors. *Machine Learning* 1996; 24 :123-40.

- [12] Breiman L. Random forests *Machine Learning*. 2001;45 5-32.
- [13] T Ho. The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence*. 1998;20: 832-44.
- [14] Pedregosa et al. [Scikit-learn: Machine Learning in Python](#). *JMLR*. 2011; 12:2825-30.
- [15] G Louppe and P Geurts. Ensembles on random patches. *Machine Learning and Knowledge Discovery in Databases*. 2012; 346-61.
- [16] Chen T, Carlos G. XGBoost: A scalable tree boosting system [C]. *SIGKDD*. 2016.
- [17] Buitinck et al [API design for machine learning software: experiences from the scikit-learn project](#) *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*. 2013;108-22.