Research on Discrimination Method of Pavement Disease Causes Based on Machine Learning

Zhoucong Xu, Haonan Zhou*, Quanlei Wang, Tianlong Chen, Liqin Long, Li Xia
China Merchants Chongqing Communications Technology (CMCT) Research & Design Institute CO., LTD, Chongqing 400067, China
* Corresponding Author: 470880789@qq.com

ABSTRACT

To construct an automatic discrimination method for the causes of pavement diseases, the typical characteristics of different types of asphalt pavement diseases of the Inner Ring Expressway in Chongqing, which was taken as an engineering example, were analyzed, and the feasibility of data dimension reduction analysis was determined based on the correlation characteristics of different types of damage. Then, numerous state information data were subjected to dimension reduction through the principal component analysis (PCA), followed by the automatic cause analysis of pavement diseases using the random forest algorithm. The results show that the cause conclusions acquired through machine learning model training basically accord with the actual field survey conclusions. Thus, it can be deemed that the intelligent discrimination method based on machine learning is reliable, to some extent, for the cause analysis of pavement diseases and can serve as an automatic discrimination method for the follow-up development of an intelligent maintenance decision system.

Key Words: Pavement disease; correlation analysis; cause discrimination; machine learning

1. INTRODUCTION

As the highway network is gradually improved, China's highway transportation industry has developed from a large-scale construction period to a sustainable maintenance period. It is clearly pointed out in the *Outline for the Development of Highway Maintenance Management* formally printed and distributed by the Ministry of Transport to "promote the maintenance transformation and carry out scientific maintenance decision-making; explore and establish a highway disease-oriented backtracking mechanism, analyze the law of performance attenuation and the causes of diseases, and gradually improve the industry supervision system and technical measures; promote the construction of maintenance decision-making support information systems, popularize scientific decision-making technologies, scientifically formulate maintenance investment plans, and rationally select technical solutions of maintenance". Therefore, the maintenance and management of highway traffic in China is becoming increasingly important.

Highway pavement disease is one of the most common diseases in highway maintenance. How to accurately classify highway pavement diseases and judge the causes of pavement diseases is the premise of scientific maintenance decision-making. The causes of pavement diseases have been extensively investigated by both Chinese and foreign scholars. Eighmy *et al.* [1] analyzed and studied the causes and cracking mechanism of cracks on asphalt pavement by combining the climatic environment in Texas. Afterward, Hojat *et al.* [2] explored the correlation between traffic loads and asphalt pavement diseases by combining the properties of asphalt mixtures and acquired the cracking mode of asphalt cracks and the disease mechanism. Combining the Xi'an-Tongchuan Expressway maintenance and reconstruction project, Pan [3] conducted field detection of some road sections, analyzed the causes of such diseases as pavement cracks and ruts, and proposed a mathematical model for the comprehensive evaluation of the pavement. For a long time, the causes of pavement diseases are discriminated mostly by combining field detection and indoor experimental verification and then analyzed and judged by establishing the corresponding mechanical deduction model [4–8]. With the development of artificial intelligence (AI) technology, the causes of pavement diseases are discriminated against using intelligent machine learning algorithms on the basis of mass historical data, providing a new direction for data-driven scientific pavement maintenance decision-making. In this study, therefore, the characteristics of typical pavement diseases were analyzed based on historical detection data, and a discrimination method for the causes of pavement diseases based on the random forest algorithm was proposed.

International Conference on Smart Transportation and City Engineering (STCE 2023),
edited by Miroslava Mikusova, Proc. of SPIE Vol. 13018, 130182W
© 2024 SPIE · 0277-786X · doi: 10.1117/12.3024155

2. DATA PREPROCESSING

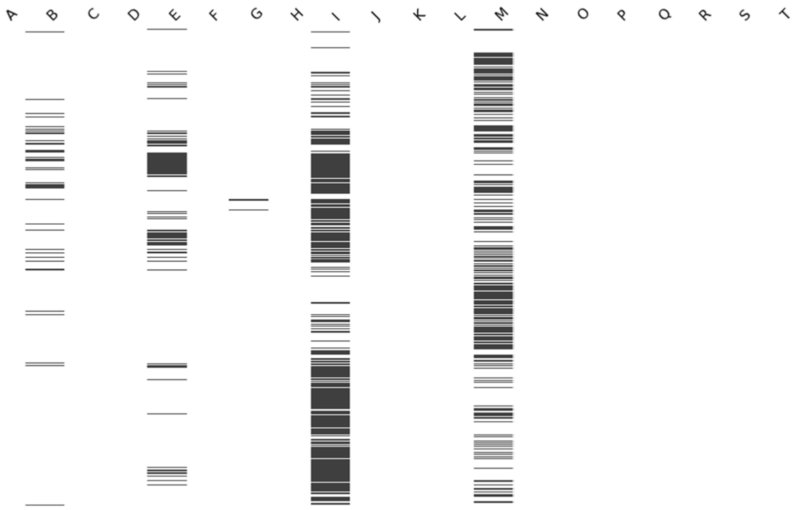
2.1 Correlation analysis of typical pavement disease data characteristics

The high characteristic dimension of the original disease data will lead to excessively high requirements for computing power or too low model generalization ability, so dimension reduction is required before data input, i.e., the correlation analysis of typical pavement disease data characteristics should be done first to prove the feasibility of dimension reduction. In this study, the detection data of typical road sections of the Inner Ring Expressway in Chongqing during 2015–2021 were adopted, and the pavement structure type of each road section is exhibited in Table 1. Then, an intelligent CMCT road operation and maintenance database (CMCT-ROMD), which included basic road information, typical pavement disease data, disease causes, and a maintenance scheme knowledge base, was established according to the detection data and historical maintenance data. The disease data, detection data, and environmental data involved in the follow-up data input all came from this database.

Table 1. Pavement Structure Types of Typical Road Sections

Typical road section	Layer thickness/mm	Structure type	Material composition
Typical road section 1 (white+ black/sealing coat/modified asphalt AC surface course)	40	Surface course	Modified asphalt AC-13
	60	Surface course	AC-20
	8	Sealing coat	Modified asphalt slurry sealing coat
	250	Base course	Cement concrete
	200	Subbase course	Lime powder-coal ash-macadam subbase course
	300	Subbase course	Lime-coal cinder-macadam subbase course
Typical road section 2 (white + black/sealing course-free/modified asphalt AC surface course)	40	Surface course	Modified asphalt AC-13
	70	Surface course	AC-20
	250	Base course	Cement concrete
	200	Subbase course	Lime powder-coal ash-macadam subbase course
	300	Subbase course	Lime-coal cinder-macadam subbase course
Typical road section 3 (white + black/sealing course/SMA surface course)	40	Surface course	SMA-13
	35	Surface course	AC-13
	50	Surface course	AC-20
	250	Base course	Cement concrete
	200	Subbase course	Lime powder-coal ash-macadam subbase course
	300	Subbase course	Lime-coal cinder-macadam subbase course
Typical road section 4: tunnel pavement	50	Surface course	Modified asphalt SMA-13
Typical road section 5: steel bridge deck pavement	40	Surface course	Modified asphalt SMA-13
	40	Surface course	Pouring-type GA-10

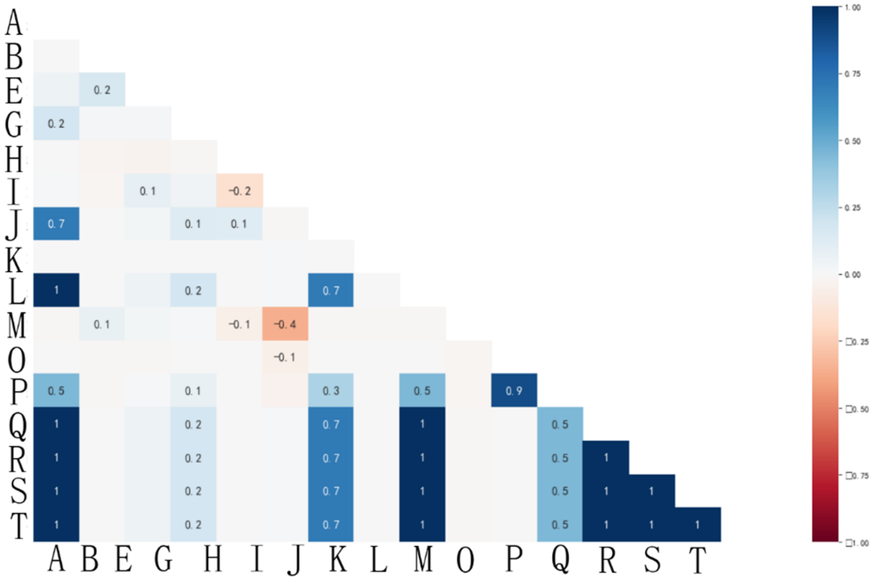
First of all, the existence of all kinds of disease types on all road sections and the correlation between such diseases were analyzed, as shown in Figures 1 and 2.



A-Slight fissure B-Moderate fissure C-Serious fissure D-Slight block fracturing E-Serious block fracturing F- Slight longitudinal cracking G-Serious longitudinal cracking H-Slight transverse cracking I-Serious transverse cracking J-Slight subsidence K-Heavy subsidence L-Slight rutting M-Serious rutting N-Slight wavy upheaval O-Serious wavy upheaval P-Slight potholes Q-Serious potholes R-Diffusion of oil S-Blocky patching T-Strip-like patching

Figure 1. Existence Analysis of Disease Characteristics on All Road Sections

It could be seen from Figure 1 that the disease types in the whole section of the Inner Ring Expressway in Chongqing included: fissure, block fracture, longitudinal cracking, transverse cracking, subsidence, rutting, upheaval, potholes, and patching. Among them, fissure, block fracture, transverse cracking, and rutting were the main diseases with high frequency, while other diseases existed at a relatively low frequency.



A-Slight fissure B-Moderate fissure C-Serious fissure D-Slight block fracturing E-Serious block fracturing F- Slight longitudinal cracking G-Serious longitudinal cracking H-Slight transverse cracking I-Serious transverse cracking J-Slight subsidence K-Heavy subsidence L-Slight rutting M-Serious rutting N-Slight wavy upheaval O-Serious wavy upheaval P-Slight potholes Q-Serious potholes R-Diffusion of oil S-Blocky patching T-Strip-like patching

Figure 2. Correlation Analysis of Disease Characteristics on All Road Sections

It could be observed from Figure 2 that all diseases were correlated to some extent, i.e., the occurrence of one disease was usually accompanied by another one. For instance, fissure was highly correlated with rutting and potholes. In the case of the fissure disease on a road section, rutting would co-exist with potholes at a considerable probability.

Through the analysis of the disease data of the whole road section from 2015 to 2021, it was found that: 1) a high proportion of disease characteristics had missing values; 2) some disease characteristics might have zero or very low contribution to cause analysis and decision-making; 3) all types of diseases were specifically correlated. On this basis, the following conclusion could be drawn: the pavement disease data could be subjected to dimension reduction in the decision-making analysis, which could facilitate the subsequent algorithm design.

In addition, the damage characteristic maps of different structure types (typical road sections 1–5) were compared by the same method, as shown in Figure 1, and the correlation analysis was conducted, as shown in Figure 2. Finally, the following conclusions were drawn: 1) The main damage characteristics of different structure types of pavements were quite different, along with typical damage characteristics corresponding to typical structures; 2) No significant correlation was observed between crack damage and deformation damage; 3) The correlation between some damage types was evident, and damage characteristic indexes could be preprocessed through dimension reduction.

2.2 Data dimension reduction

Considering the certain data loss and the sparse input sample matrix, data were preprocessed by invoking the csc_matrix sparsification module before fitting. In addition, the number of samples applied to machine learning was relatively small, but there were many sample characteristics, so the dimension should be stipulated before algorithm design. In this study, the data characteristics were subjected to dimension reduction using the principal component analysis (PCA) method.

PCA is a data analysis technique whose great advantage is that the dimension reduction process is not restricted by parameters [9–12], aiming to reduce the dimension of high-dimensional data, remove the noise and redundancy in data, and find the important elements and structures therein, so as to reveal the simple structure hidden behind complex data.

The steps of the PCA algorithm were described as follows:

1) The pavement structure data (six characteristics: road age, structural layer number, structural layer thickness, mixture type, adhesive layer type, and years of maintenance), pavement detection data (8 characteristics: *PCI*, *RQI*, *RDI*, *SRI*, *PSSI*, water permeability coefficient, core sample strength, and extraction aging), pavement disease data (21 characteristics: 21 diseases of the asphalt pavement), and traffic environment data (5 characteristics: traffic volume, cumulative load, temperature, humidity, and precipitation) of the discriminated road sections were read and preprocessed (including normalization and centralization).

2) A sample set $D=\{x_1,x_2,...x_m\}$was established, and the dimension $d^{'}$ of the low-dimensional space was set.

3) The covariance matrix XX^T of samples was calculated, and the eigenvalues and eigenvectors were solved.

4) The calculated eigenvectors were sorted according to the values of eigenvalues, and the first $d^{'}$ eigenvectors were chosen as per the dimension obtained after dimension reduction to constitute the characteristic space $W^* = (w_1, w_2, ..., w_{d^{'}})$.

5) After the dimension $d^{'}$ of the low-dimensional space was determined after dimension reduction, a better $d^{'}$ value was selected through the cross validation of *k*-nearest neighbor classifiers in the low-dimensional space with different $d^{'}$ values.

In this study, the influencing factor data of 40 characteristics, including pavement structure data, pavement detection data, pavement disease data, and traffic environment data, in typical sections of the Inner Ring Expressway in Chongqing were dimension-reduced to 12 principal component data, and an algorithm model was designed based on the dimension-reduced data.

3. ALGORITHM DESIGN

3.1 Logic framework for discriminating the causes of pavement diseases

Conventionally, the causes of pavement diseases have been analyzed depending on multiple types of data and artificial reconnaissance and discrimination. Experts and engineers usually consider such factors as the structural composition of roads, the environment, and loads, all of which should be included as factors influencing the cause analysis. In this study, the logic framework for the cause analysis of pavement diseases was established, as shown in Figure 3. The steps of judging the cause of the disease are divided into three times. The primary discrimination mainly aimed to discriminate the damage degree while the secondary discrimination aimed to further determine the damage degree and preliminarily speculate about the cause, and the cause type was determined by the final discrimination. The parameters used in the primary discrimination and secondary discrimination were data acquired in accordance with the industry standard currently in force [11,12], as seen in Tables 2 and 3. The final discrimination results were acquired through the random forest algorithm.

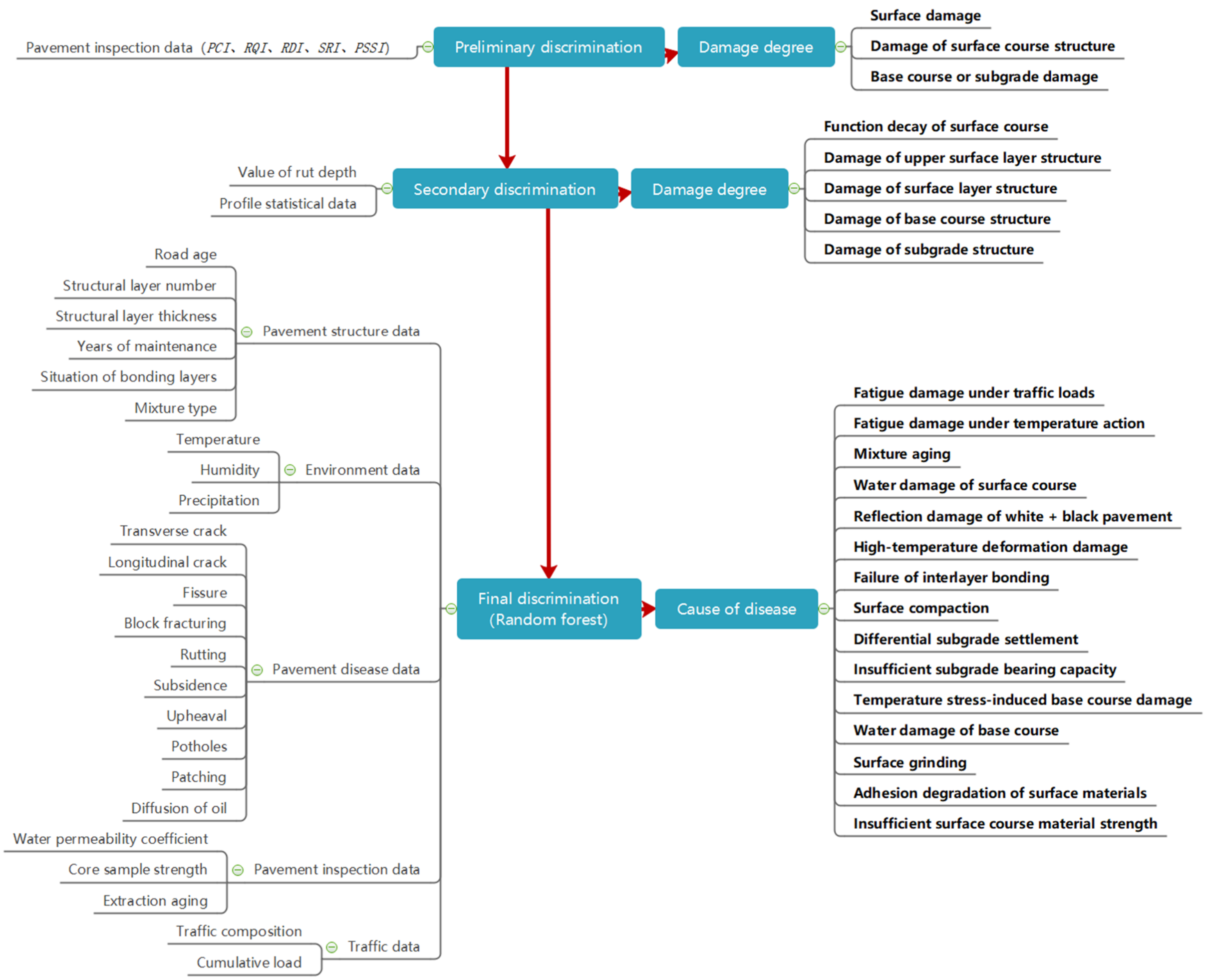

Pavement inspection data（PCI、RQI、RDI、SRI、PSSI）
Preliminary discrimination
Damage degree
Surface damage
Damage of surface course structure
Base course or subgrade damage
Value of rut depth
Profile statistical data
Secondary discrimination
Damage degree
Function decay of surface course
Damage of upper surface layer structure
Damage of surface layer structure
Damage of base course structure
Damage of subgrade structure
Road age
Structural layer number
Structural layer thickness
Years of maintenance
Situation of bonding layers
Mixture type
Pavement structure data
Temperature
Humidity
Precipitation
Environment data
Transverse crack
Longitudinal crack
Fissure
Block fracturing
Rutting
Subsidence
Upheaval
Potholes
Patching
Diffusion of oil
Pavement disease data
Water permeability coefficient
Core sample strength
Extraction aging
Pavement inspection data
Traffic composition
Cumulative load
Traffic data
Final discrimination (Random forest)
Cause of disease
Fatigue damage under traffic loads
Fatigue damage under temperature action
Mixture aging
Water damage of surface course
Reflection damage of white + black pavement
High-temperature deformation damage
Failure of interlayer bonding
Surface compaction
Differential subgrade settlement
Insufficient subgrade bearing capacity
Temperature stress-induced base course damage
Water damage of base course
Surface grinding
Adhesion degradation of surface materials
Insufficient surface course material strength

Figure 3. Logic Framework for Cause Discrimination of Pavement Diseases

Table 2. Critical Values of Initial Discrimination Indexes

Evaluation index/discrimination conclusion	Subgrade/base course damage	Surface course damage	Surface performance decay
PCI	<75	<85	<95
RQI	<76	<84	<94
RDI	<73	<82	<91

Table 3. Critical Values of Secondary Discrimination Indexes

Characteristic index/discrimination conclusion	Subgrade damage	Base course damage	Surface course damage	Surface performance decay
Structural strength *PSSI*	<65	<75	<85	<95
85 quantiles of rut depth data distribution	>20 mm	>15 mm	>9 mm	>6 mm
Median of rut depth data distribution	>9 mm	>6 mm	>6 mm	>3 mm
Statistical index of height difference between longitudinal sections	>120 mm	>80 mm	>50 mm	>20 mm

3.2 Random forest algorithm

The random forest algorithm is a machine learning algorithm based on Bagging. Therein, the basic unit is a decision tree, and the random forest is composed of several decision trees. In the random forest, special classification and regression trees (CARTs) are selected as weak learners. Improvements have been made by the random forest in the establishment of decision trees: for ordinary decision trees, the optimal characteristic is generally chosen from all n sample characteristics from the node to divide the left and right subtrees of the decision tree; for the random forest, some sample characteristics (the number is smaller than n, assumed to be n_t) on the node are randomly selected, and the optimal characteristic is chosen from such sample n_t characteristics to divide the left and right subtrees of the decision tree, aiming to further enhance the model generalization ability. Generally, the smaller the value of n_t, the smaller the variance of the model, but the fitting degree of the training set will be lower, so an appropriate value of n_t should be chosen to achieve a better model fitting effect [13,14].

The application steps of the random forest algorithm are as follows: 1) The sample set D is taken as the input, $D=\{(x_1,y_1),(x_2,y_2),(x_3,y_3),...,(x_m,y_m)\}$, and the iteration number of the weak learner is T; 2) The sample set is randomly sampled, the number of randomly sampling times is $t=1,2,3,...,T$, and m samples are randomly extracted to obtain the sample D_t; 3) The obtained sample D_t is trained in the CART decision tree model $G_t(x)$, namely, learning is performed in the t-th weak learner; 4) the final strong learner $f(x)$ is output.

The classification effect of the random forest is determined by the correlation between the classification ability of each tree in the random forest and the decision tree. The stronger the classification ability of each tree, the better the classification effect. The greater the correlation between decision trees, the worse the classification effect. For the selection of feature quantity, the greater the characteristic quantity, the stronger the classification ability of each tree, and the greater the correlation between decision trees. Therefore, the choice of the characteristic quantity has a great influence on the final classification effect [15].

3.3 Cause analysis model of pavement diseases based on the random forest algorithm

The cause analysis of pavement diseases based on the random forest algorithm is, in essence, to classify the influencing factors of the cause and get the cause types by analyzing the characteristic data of different types of diseases, among which the extraction of disease characteristic quantity is one of the key factors. It is necessary to extract an appropriate characteristic quantity from many state information data of pavement diseases and select the types that can accurately reflect the causes of diseases.

The state information data of pavement diseases is derived from pavement structure data, pavement detection data, pavement disease data, and external traffic environment data. After PCA-based dimension reduction, the result was taken as the input of the random forest algorithm while the cause type as its output. In the random forest algorithm, a lot of characteristic quantity data should be repeatedly trained to finally acquire a relatively stable model, i.e., the cause analysis model of pavement diseases based on the random forest algorithm, and the model accuracy should be verified by testing the sample data.

The training process of the random forest model is shown in Figure 4.

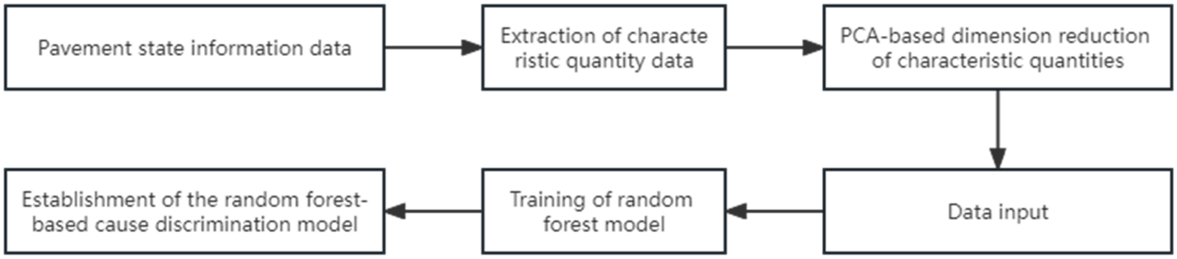

Pavement state information data
Extraction of characte ristic quantity data
PCA-based dimension reduction of characteristic quantities
Data input
Training of random forest model
Establishment of the random forest-based cause discrimination model

Figure 4. Training Process of the Random Forest Model

4. EXAMPLE TEST

In this study, the disease data of typical sections of the Inner Ring Expressway in Chongqing were used for testing, and the cause analysis conclusions and field exploration results of 5 typical pavement diseases with high frequency in maintenance data were used for verification. Then, the historical case analysis conclusions were compared with the learning conclusions based on the random forest algorithm, specifically as follows:

1) Line crack

The line crack of the pavement was mainly involved. The line crack of this road section was attributed to the following aspect: the original pavement structure was a composite pavement, and the interface of construction joints in the original cement concrete layer was not well treated in the pavement of the asphalt course, which led to a lot of reflection cracks in the asphalt course.

Learning conclusion: The cracks actually detected were divided into transverse and longitudinal cracks, where the latter was ascribed to insufficient shear strength. Some slight transverse cracks were discriminated against to be fatigue cracks.

2) Fissure

Fissures mainly referred to the problem of fatigue, which was attributed to the short local fatigue life of the asphalt course due to the poor bonding between asphalt concrete layers, the partially thin surface course, and overload traffic. Fissures gradually diffused, the core filling was performed beside the fissure for detection, and poor bonding between the upper and lower asphalt concrete surface courses was found.

Learning conclusion: Some fissures resulted from the poor bonding between structural layers, which accorded with the field investigation conclusion.

3) Rutting and upheaval

Rutting not only resulted in poor road driving comfort but also affected driving safety. Rutting and upheaval were mainly attributed to the large traffic volume and insufficient thermal stability of the asphalt mixture.

Learning conclusion: Asphalt aging and high-temperature shearing were detected, which was basically identical to the field investigation conclusion.

4) Potholes and spalling

The asphalt mixture peeled off mainly due to the aging of the pavement asphalt mixture or the poor adhesion of aggregate and asphalt. Because of the high porosity of asphalt concrete, rainwater permeated into the pavement structure layer, and dynamic absorption was generated under the action of traffic loads. As a result, the fine aggregate was gradually adsorbed out of asphalt concrete, generating potholes at weak positions.

Learning conclusion: Poor interlayer adhesion and water damage were detected, which was in line with the field investigation conclusion.

5) Subsidence

Cement concrete, which was the base course, was subjected to such diseases as slab breaking and faulting of slab ends under the loading action. In the case of a large traffic volume on the Inner Ring Expressway, cement concrete was fragmented, which resulted in structural instability and subsidence.

Learning conclusion: Structural problems occurred to the asphalt course under the action of traffic loads and rainwater, which was basically consistent with the field investigation conclusion.

Generally speaking, the random forest discrimination method proposed in this study for disease causes achieved a favorable effect in the example test, and the field investigation conclusion was largely identical to the analysis conclusion of historically recorded data. However, certain differences were still found in the cause discrimination of specific subdivided fields (line cracks), which might be attributed to the insufficient training of reflection crack data. This could be solved after the datasets were subsequently supplemented.

5. CONCLUSION

In this study, the disease data of the asphalt pavement on the Inner Ring Expressway in Chongqing were subjected to the characteristic analysis, followed by dimension reduction of numerous information data through PCA. Then, a cause discrimination model for pavement diseases based on the random forest algorithm was established. Finally, the following conclusions were drawn:

1) The disease characteristics of the detection data of typical sections of the Inner Ring Expressway in Chongqing during 2015–2021 were analyzed. The results revealed that the main diseases were fissure, block fracturing, transverse cracking, and rutting, all of which were correlated in some way, and a high proportion of disease characteristics had missing values, which might make zero or extremely low contribution to the cause analysis and decision-making. Hence, it was feasible to perform dimension reduction of pavement disease data. Furthermore, the influencing factors of 40 types of characteristics, including pavement structure data, pavement detection data, pavement disease data, and traffic environment data, were dimension-reduced into 12 pieces of principal components through the PCA method.

2) The logic framework for the cause analysis of pavement diseases was established, and the disease causes were discriminated through 3 steps: the preliminary discrimination mainly aimed to judge the damage degree, the secondary one further determined the damage degree and preliminarily speculated about the case, and the final one determined the cause type. Moreover, a cause analysis model for pavement diseases based on the random forest algorithm was constructed and used in the final discrimination stage.

3) The disease data of typical sections of the Inner Ring Expressway in Chongqing, as an example, were used for testing, and the cause analysis conclusions and field exploration results of 5 typical pavement diseases with high frequency in maintenance data were used for verification. The proposed discrimination method based on random forest achieved a good effect in the example test. The cause conclusions acquired through model learning and training largely accorded with the actual field investigation conclusion, but certain differences were found in the cause discrimination of subdivided fields (line cracks). This study can, on the whole, provide an automatic cause discrimination method of pavement diseases for the follow-up development of an intelligent maintenance decision system.

REFERENCES

[1] Eighmy, T. T., Cook, R. A., Gress, D. L., Coviello, A., Spear, J., & Hover, K., et al. (2002). Use of accelerated aging to predict behavior of recycled materials in concrete pavements: physical and environmental comparison of laboratory-aged samples with field pavements.

[2] Hojat Shamami, V., & Khiavi, A. K. (2017). Effect of temperature on geosynthetic rutting performance in asphalt pavement. Petroleum Science & Technology, 35(11): 1104–1109.

[3] Pan X. Study on interlayer treatment technology of full life cycle asphalt pavement [D]. Xi'an: Chang'an University, 2012.

[4] Dou M. J., Hu C. S., Dorgirob, et al. Analysis of the causes of Qinghai-Tibet highway [J]. Journal of Glaciology and Geocryology, 2003, 25 (4): 439–444.

[5] Yang S. P., Lu Z. S., Cheng X. C. Causes and prevention of diseases of asphalt pavement at semi-rigid base level [J]. Journal of Hefei University of Technology, 2002,25 (5): 748–752.

[6] Zhang Y., Shi H. J., Li W. S. et al. Analysis and treatment measures [J]. Journal of Inner Mongolia Agricultural University (Natural Science Edition), 2005,26 (4): 80–82.

[7] Gao L. B. The causes and prevention measures of early pavement diseases of asphalt pavement [C]// International Road and Airport Pavement Technology Conference, Kunming: Chinese Highway Society, 2002:173–175.

[8] Chang Z. P. Genuses and treatment methods of asphalt pavement diseases [J]. Transpo World, 2009 (1): 80–81.

[9] Zhang M., Li T. S., Zhong S. Y. Implementation of Matlab-based principal component analysis method (PCA) [J]. Journal of Guangxi University, 2005 (S2): 74–77.

[10] Li B., Han S., Xu O. M., et al. Evaluation of the use performance of asphalt pavement based on principal component analysis method [J]. Journal of Chang'an University, 2009,29 (3): 15–18.

[11] Liu M. M., Pan J. P., Yang H. M. Study on the early warning and evaluation model of snow and ice disaster in mountain roads [J]. Technology of Highway and Transport, 2011 (3): 27–30.

[12] Li C., Huang K., Li X. Prediction of short-time traffic flow based on K-nearest neighbor and principal component analysis [J]. Technology of Highway and Transport, 2022 (3): 138–144.

[13] Xu Y. C., Du Y. B. Using the random forest algorithm to build the asphalt pavement rut prediction model [J]. Journal of Henan University of Urban Construction, 2022 (2): 31.

[14] Zhang J. X, Guo W. D., Song B., et al. Prediction of asphalt pavement performance based on random forest [J]. Journal of Beijing University of Technology, 2021,47 (11): 1256–1263.

[15] Li L. F., Gao X. X., Sun R. Y. A random forest-based pavement crack detection method and its evaluation method [P]. Shanxi: CN108520278A, 2018.