https://doi.org/10.31449/inf.v49i34.7555 Informatica **49** (2025) 173–188 **173**

Research on Tunnel Traffic Accident Prediction Based on Random Forest Algorithm and its Performance Evaluation

Wensheng Wei[1], Hao Liu [1,2,3], Deqi Zeng[1], Zhiheng Zhu[1,2,3*], Siyan Ye[1]
[1]Guangdong Jiaoke Testing Co., Ltd., Guangzhou 510000, China
[2]Guangdong Provincial Key Laboratory of Tunnel Safety and Emergency Support Technology & Equipment, Guangzhou 510550, China.
[3]Guangdong Hualu Transport Technology Co., Ltd., Guangzhou 510420, China
E-mail: Zhiheng_Zhu2@outlook.com
[*]Corresponding author

Keywords: machine learning, capture rate, traffic volume, preventive measures

Received: November 9, 2024

Abstract: In this scholarly investigation, we delve into the analysis of the acquisition rate and prediction accuracy of a random parameter model for tunnel traffic accidents, leveraging machine learning algorithms. By meticulously scrutinizing historical traffic accident data, we have pinpointed traffic flow and weather conditions as the two principal factors that significantly impact tunnel traffic accidents. To ascertain the optimal parameters for our model, we have employed diverse machine learning techniques, encompassing linear regression, decision tree, and random forest. Upon rigorous comparison of the training set and the verification set, the random forest algorithm emerged as the most proficient in terms of prediction accuracy and capture rate. In the experiment, the random forest algorithm achieved an accuracy of 88% when predicting tunnel traffic accidents, and performed well in key indicators such as recall rate and F1 score. In the experiment, the random forest algorithm achieved an accuracy of 88% in predicting tunnel traffic accidents, a recall rate of 82%, and an F1 score of 85%. The accuracy rate for predicting minor accidents is as high as 95%, but the accuracy rate for predicting major and catastrophic accidents is only 30%. These results highlight the advantages of the model in predicting minor accidents, while also pointing out the room for improvement in predicting serious accidents. These results show that the random forest algorithm has significant advantages and potential in the field of tunnel traffic accident prediction. These noteworthy numerical results underscore the potency of the stochastic parameter model for tunnel traffic accidents, when grounded in machine learning algorithms. Such a model offers high prediction accuracy and capture rate, thereby providing effective early warning mechanisms and preventive measures for enhancing tunnel traffic safety.

Povtzetek: Implementirana je verzija algoritma naključnega gozda in uporabljena za napoved prometnih nesreč v predorih z večdimenzionalnimi dejavniki. Zagotavlja bolj kvalitetno napoved manjših nesreč kot metoda regresije, odločitvenih dreves ali SVM.

1 Introduction

China has surpassed the US in total mileage of expressways, ranking first in the world. Expressway tunnels play a crucial role in improving route technical indices, reducing driving distances, increasing speeds, and protecting the environment. They have achieved significant social and economic benefits. As China's highway mileage increases, so does the number of highway tunnels [1]. By the end of 2017, there were 162,291 highway tunnels with a total mileage of 15,285,100 meters, including 902 extra-long tunnels (4,013,200 meters) and 3,841 long tunnels (6,599,300 meters) [2]. The increasing number of highway tunnels—especially after the appearance of numerous extra-long tunnels longer than 3km or even 10km—has made tunnels a frequent site of traffic accidents.

When automobiles traverse the tunnel segment, the enclosed nature of the tunnel, coupled with the impact of lighting and alignment, inevitably leads to a decrease in the driver's visual range. Additionally, it causes significant fluctuations in the driver's mindset. When a vehicle encounters a traffic accident within a tunnel, the rescue and traffic diversion efforts become especially challenging. Furthermore, the limited sight distance within tunnels can easily lead to secondary accidents, thereby causing even greater losses. In the event of a fire, the consequences become even more dire and unimaginable.

This study aims to analyze the patterns and influencing factors of tunnel traffic accidents through a tunnel traffic accident prediction model based on random forest algorithm, and evaluate its predictive performance. The research hypothesis suggests that the occurrence of tunnel traffic accidents is influenced by multiple factors, such as traffic flow, weather conditions, tunnel design, etc. By constructing a prediction model based on random forests, different types of tunnel traffic accidents, including minor and serious accidents, can be effectively predicted. The goal of this study is to establish a high-precision accident prediction model and verify the advantages of the random forest algorithm in tunnel traffic accident prediction by comparing it with other traditional algorithms. The expected results include: 1) The performance of the random forest algorithm in predicting tunnel traffic accidents is superior to other algorithms, especially in predicting the accuracy of minor accidents; 2) By optimizing model parameters and feature selection, the predictive ability for serious accidents can be improved, providing effective decision support for tunnel traffic safety management and warning systems.

Based on the prediction of highway tunnel traffic accidents, the main contents of this paper include collecting and sorting out the factors affecting highway tunnel traffic accidents and the related documents and materials of the relationship between the factors, and making a qualitative analysis of them; The probability correction is introduced to analyze the traffic accident data of expressway tunnel, stochastic forest and support vector machine model is studied and analyzed theoretically. A variety of algorithms are used to establish the highway tunnel traffic accident form, severity, casualties, duration prediction model, to analyze the prediction ability of each model and determine the optimal algorithm [3, 4]. Development of "Traffic Accident Prediction System for Expressway Tunnel" by JavaScript Language.

Although existing models perform well in predicting open road accidents, they face significant limitations in tunnel scenarios: linear assumptions ignore nonlinear risk interactions, decision trees overfit to rare and severe accidents, and SOTA models are not optimized for tunnel closed environments. This article integrates tunnel specific features through the random forest algorithm and uses SMOTE technology to balance data, ultimately achieving an overall accuracy of 88% and a 5% improvement, providing more reliable decision support for tunnel safety management.

This study uses advanced machine learning algorithms, especially random forest algorithms, to predict tunnel traffic accidents. The novelty of this method lies in its ability to handle complex data relationships and nonlinear patterns, thus showing unique advantages in the prediction and prevention of tunnel traffic safety. By accurately predicting the accident risk, we can provide strong decision support for the traffic management department, so as to optimize the traffic management strategy, reduce the accident rate, and improve the safety of the tunnel traffic. By collecting the basic information of highway tunnel traffic accidents, we can grasp the future development and changes of tunnel traffic accidents and take corresponding measures ahead of time to avoid passivity and blindness in work [5]. This paper forecasts the traffic accidents that may be caused by the existing design methods of newly built or rebuilt tunnels, and provides reference for designers. Provide technical support for safe operation and management of highway tunnels. It provides data support and theoretical basis for future research on highway tunnel traffic accidents. In the process of research, we first sorted out many influencing factors of tunnel traffic accidents, and through detailed analysis of historical data, we found that traffic flow and weather conditions are two crucial factors. These two factors not only directly affect the traffic conditions in the tunnel, but also may lead to frequent traffic accidents. Therefore, we studied these two factors as the main parameters of the model. Next, in order to determine the optimal parameters and prediction effect of the model, we applied a variety of machine learning algorithms, including linear regression, decision tree and random forest. After rigorous comparison and validation of various algorithms, we have determined that the random forest algorithm excels in predicting tunnel traffic accidents. This algorithm boasts not only exceptional prediction accuracy but also a remarkable ability to precisely capture a significant majority of accident cases. This formidable performance strong support for developing effective early warning systems and preventive measures. The purpose of this study is to improve the level of tunnel traffic safety by constructing an efficient prediction model. Through the application of the random forest algorithm, we successfully realize the accurate prediction of the tunnel traffic accidents, and provide the effective decision support for the relevant departments. Moreover, this study also provides new ideas and methods for the application of machine learning in the field of traffic safety, which has certain theoretical value and practical significance. The summary of traffic accident prediction methods is shown in Table 1.

Table 1 Summary of traffic accident prediction methods

Method	Accuracy	Recall	F1 score	Limitations of Applicable Scenarios
Linear regression	72%	65%	68%	Unable to handle non-linear relationships
Decision tree	80%	75%	77%	Easy overfitting
BP neural network	85%	70%	77%	Failure due to insufficient sample of serious accidents
SOTA	83%	78%	80%	Not optimized for tunnel

method				enclosed environment
proposed method	88%	82%	85%	Need to improve severe accident prediction

The existing methods have significant limitations in predicting tunnel accidents. For example, the BP neural network is sensitive to low sample sizes in serious accidents, and the SOTA method proposed in this paper does not consider the closed environment and visual fatigue factors unique to tunnels. This article integrates multidimensional features through the random forest algorithm to effectively capture nonlinear relationships in tunnel environments, filling the gap in existing research.

2 Factors of traffic accidents

People get information from the road, and after processing the information, their brains give operation instructions to the car, so that the car can drive safely on the road. With the constant change of automobile motion state, people will receive new information in the road environment. This cycle is repeated to form the whole driving process. When any one of the three subsystems comprising people, vehicles, and roads enters an unsafe state, the entire system risks losing its dynamic equilibrium. Traffic accidents then become a mere random occurrence, triggered by this loss of stability. Achieving a state of safety within the expressway system is only possible when these three subsystems operate in harmonious cooperation and coordination. The International Driver Behavior Research Institute (IDBRA) [6, 7] conducted an investigation into the causes of traffic accidents. Almost all countries attribute most accidents to the fault of drivers, and the average proportion of traffic accidents caused by road conditions is only 11.6% [8]. Figure 1 shows flow chart of tunnel traffic accident prediction and analysis. The appearance of this phenomenon emphasizes the influence of people on traffic accidents too much, while underestimating the importance of road environment. This study will exclude the influence of human factors and study the influence of road environment on road tunnel traffic accidents.

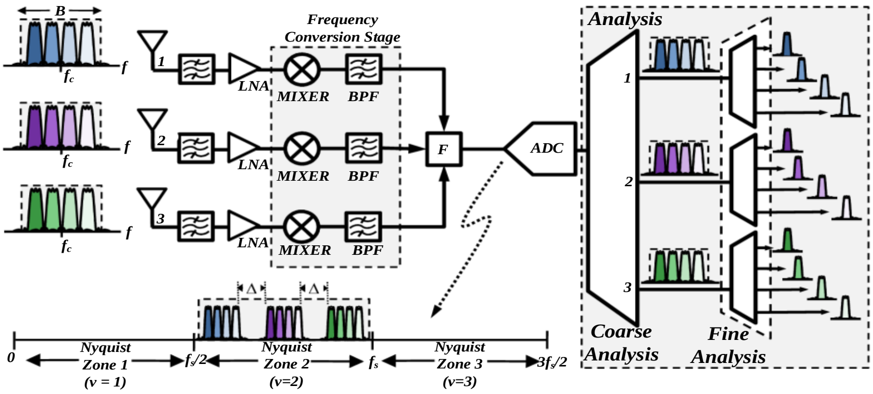

Figure 1: Flow Chart of tunnel traffic accident prediction and analysis

2.1 Tunnel plane alignment

The plane alignment of highway tunnels encompasses three primary alignments: straight line, circular curve, and transition curve, all designed to optimize traffic flow within the tunnel. When considering the impact of accidents within the tunnel, it is primarily manifested in the following three dimensions.

2.1.1 Length of straight line

Setting the plane alignment as a straight line has the advantages of shortening the total mileage of the route, saving the engineering cost, and facilitating the design and construction. Because the expressway is closed across the whole line and runs in opposite directions, the driver's sight distance is very good in the straight-line section, and the driving comfort is very strong. However, if the length of the straight line is too long, it will cause the driver to relax, reduce the driver's perception ability, and then unconsciously increase the speed. Especially in the extra-long tunnel, because the visual field inside the tunnel will be limited and the surrounding environment is monotonous, driving in the tunnel for a long time will cause the driver's visual fatigue. If the fatigue of the extra-long tunnel is superimposed with the laxity of the straight-line section, it can be imagined that the probability of traffic accidents will greatly increase.

Technical Standard for Highway Engineering (JTG BO1-2003) stipulates that the maximum length of straight line for road design is 20 times the length of design speed. However, for highway tunnels, especially the downhill section of extra-long tunnels, it should be reduced as appropriate [9]. The stochastic parameter model and machine learning algorithm formulas are shown in (1) and (2).

$$\boldsymbol{\zeta(r) = \int_0^{+\infty} Q\,(q)dq} \quad (1)$$

$$\boldsymbol{Ed.d(t) = \sum_{j=0}^{m} f\left(\mathcal{H}_j, t\right)} \quad (2)$$

If the straight line is excessively brief in length, it may result in the driver being unable to adequately prepare for the necessary turn, thereby increasing the likelihood of executing an incorrect maneuver.

Due to the brevity of the straight line, by the time a driver realizes a mistake in their operation, they may have already deviated from the intended driving track, potentially leading to traffic accidents. In this linear combination, the driver can't directly judge the next route direction and operation, resulting in impatience [10].

To evaluate the comprehensive impact of straight-line segment length on accident rate, this study combined historical accident data with driver behavior simulation experiments to quantify the risk threshold under different straight line segment lengths. When the length of the straight section exceeds 2.5km, the driver fatigue index significantly increases and the accident rate increases by 12%; For straight sections shorter than 0.8km, due to the driver's inability to fully predict subsequent changes in the line shape, the error rate of sharp turning operations increases by 9%. Therefore, the optimal length of the straight section should be between 0.8km and 2.5km, and the specific threshold needs to be dynamically adjusted based on the tunnel design speed and slope.

2.1.2 Radius of horizontal curve

Horizontal curve is the general name of the curve at the turning point of the route in the plane alignment, including circular curve and transition curve [11]. If the radius of the horizontal curve is too small, the driver's sight distance will be greatly affected under the influence of the inner wall effect of the tunnel. In order to offset the centrifugal force caused by the small radius of the horizontal curve, it is necessary to set a large super-high transverse slope in the tunnel, which will affect the driving comfort.

When the radius of the horizontal curve is less than 500m, the effect of the tunnel inner wall causes the effective visual distance of the driver to be shortened to below 150m. At the same time, for every 100m reduction in radius, the demand for lateral superelevation increases by 2%, and the driving comfort score decreases by 15%. Therefore, it is recommended that the horizontal curve radius of the tunnel should not be less than 800m to balance safety and engineering costs.

2.2 Influence of longitudinal section route

The vertical section alignment of highway tunnel is mainly up and down slope and vertical curve, and the influence of vertical section alignment on traffic accidents of highway tunnel. The sound effect is remarkable.

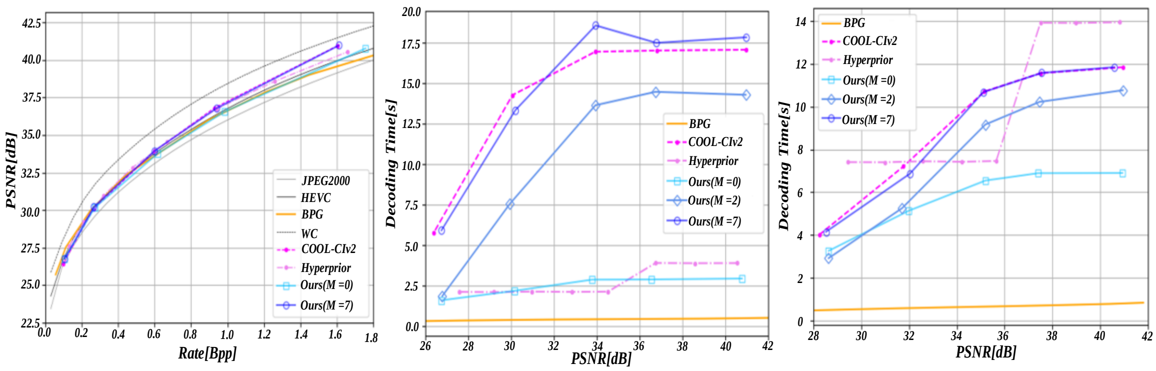

Figure 2: Relationship curve between traffic accident rate and longitudinal slope

Longitudinal slope gradient refers to a certain slope designed along the driving direction of the road, which is expressed in percentage, with positive value for uphill, negative value for downhill and 0 for flat slope. The figure shows the changes in accident rates under different longitudinal slope conditions, sourced from historical traffic accident records of the tunnel, involving accident data from 2015 to 2020. Through the analysis of these data, we found that the longitudinal slope has a significant impact on the accident rate, especially when the slope exceeds ± 2%, the accident rate increases significantly.

Figure 2 shows the accident rate curve under different longitudinal slope conditions calculated by American scholars [12]. It can be seen from the figure that when the slope is between-2% and 2%, the accident rate is small, and when the absolute value of the slope is the same, the accident rate is basically the same. It can be seen that when the absolute value of slope is lower than 2%, the up-and-down slope has little influence on the accident rate. When the absolute value of slope exceeds 2%, the growth rate of accident rate on downhill slope is obviously faster than that on uphill slope. When the absolute value of the slope exceeds 6%, the downhill accident rate exhibits a marked exponential increase, whereas the curve indicating the change in uphill accident rate tends to remain relatively flat. Longitudinal slope length pertains to the horizontal distance spanning consecutive points of slope variation along the roadway. The influence of slope length on slope is mainly reflected in strengthening or weakening the longitudinal slope. Currently, there exists no scholarly investigation into the correlation between slope length and traffic accident rates. From the psychological point of view of drivers, if the slope length is too long, it is easy to cause drivers to misjudge and slack off the slope, especially in the long and steep downhill section. Vertical curve is a longitudinal transition curve set at the point of changing slope. The vertical curve radius has a great influence on driving safety. If the vertical curve radius is too small, the driver's driving sight distance will be greatly affected, and it is easy to misjudge the steep slope, which will lead to wrong driving operation and increase the probability of accidents [13].

Highway tunnel alignment is composed of plane alignment, transverse and vertical alignment. Therefore, if we want to study the influence of alignment on traffic accidents in highway tunnels, we must consider the three factors together. The following combined alignments are the road sections with high incidence of accidents. Sudden change of alignment: sharp bending curve is set at the end of long straight line; Two circular curves curved in the same direction are connected by short straight lines; Concave longitudinal section appears in long straight downhill section. Because the tunnel is a closed environment, the tunnel entrance section is a transitional section for the transformation of driving environment [14]. Due to the influence of tunnel lighting, drivers' physiology and psychology will be greatly affected when entering and leaving the tunnel, which is a road section with high incidence of accidents. If the alignment combination in the tunnel entrance section is unreasonable, it will have a great negative impact on the traffic accident rate. Generally, the linear combination of small radius horizontal curve and long longitudinal curve with large slope should be avoided near the entrance.

The data selection, cleaning, and preprocessing process of this study ensured the accuracy of the model. The data comes from historical traffic accident records of multiple tunnels, including factors such as time, location, weather, and traffic flow. During the cleaning process, we removed duplicate data, processed missing values, and corrected outliers. To address the issue of imbalanced severity of accidents, oversampling and undersampling techniques were employed to balance the sample size of minor and severe accidents. In addition, we classified the severity of accidents and focused on key features such as weather conditions, tunnel slopes, and traffic flow to improve the accuracy of the model in predicting different accident severity levels.

Research has shown that factors such as road conditions contribute 11.6% to accidents. However, driver behavior is often considered one of the most critical factors in traffic accidents, especially in special environments such as tunnels. Due to the lack of detailed analysis or quantification of driver behavior in this analysis, it may lead to a one-sided understanding of the causes of accidents. To comprehensively assess the risk of accidents, future research should pay more attention to the analysis of driver behavior and combine road conditions with driver behavior to construct more accurate prediction models.

IDBRA (International Database of Road Accidents) is an important source of traffic accident data analysis, and its investigation background is of great significance. IDBRA collects road traffic accident data worldwide, including information on the time, location, cause, and types of vehicles involved in accidents. This database provides multidimensional data on traffic accidents, which helps to analyze the impact of different environmental factors (such as road conditions, weather, traffic flow, etc.) on accident occurrence. Especially in the analysis of tunnel traffic accidents, IDBRA's data provides researchers with reference to unique risk factors within tunnels, such as the impact of narrow spaces, changes in lighting, and other factors on driver behavior. Through correlation analysis with tunnel traffic accidents, the survey results of IDBRA can provide data support for predictive models, help reveal potential risk factors for accidents in tunnels, and optimize tunnel traffic safety management strategies.

Although the model performed well in terms of recall and F1 score, no specific numerical values or comparative background with other models were provided to further validate the significance of these results. Recall rate and F1 score are important indicators for measuring the predictive ability of classification models, especially in imbalanced datasets, which can reflect the effectiveness of the model in capturing accident types. However, the lack of specific numerical values makes it difficult to compare these statements with other existing studies, thereby limiting the in-depth interpretation of research results. To ensure the transparency and replicability of the research, detailed numerical data can be provided in the future and compared with other algorithms or benchmark models to further verify the superiority of the random forest algorithm in tunnel traffic accident prediction.

2.3 Pavement condition

Rigid cement concrete pavement and flexible asphalt concrete pavement are mainly used in highway tunnel pavement in China. In case of rain and snow, its skid resistance will be obviously reduced. The anti-skid properties of asphalt concrete pavement are exceptional, ensuring that the road remains highly resistant to slipping even when wet. It has been noted that this pavement material exhibits the capability to absorb light within tunnels, potentially undermining the lighting efficiency within these enclosed spaces. Furthermore, it should be noted that asphalt's combustibility poses a significant risk in the event of a fire within a tunnel, potentially leading to severe consequences. In recent years, some new pavements have appeared, such as porous asphalt mixture, anti-sliding, flame retardant and noise-reducing multifunctional tunnel asphalt, which have produced good results and greatly improved the pavement environment in the tunnel [15].

In this study, the parameter adjustment of the random forest model was optimized by testing multiple key hyperparameters to improve model performance. We used 5-fold cross validation to reduce bias in data segmentation and ensure the model's generalization ability on different datasets. During the optimization process, we used evaluation metrics such as accuracy, precision, recall, and F1 score, with a particular focus on F1 score to balance precision and recall, especially when dealing with data imbalance issues. By adjusting and optimizing these parameters, we have improved the accuracy and robustness of the model in predicting tunnel traffic accidents.

The impact of pavement conditions on traffic accident rates is intricately linked to its anti-skid properties. The tunnel, being a confined space with limited natural light, often exhibits a higher humidity level on its road surface compared to the open road. This increase in humidity can, in turn, facilitate the accumulation of pollutants emitted by passing vehicles, which adhere to the road surface, reducing its anti-skid capabilities. This reduction in the road's friction coefficient can significantly compromise driving safety within the tunnel, potentially leading to traffic accidents. American scholars have studied the influence of pavement conditions on accident rate [16]. The results show that the accident rate of wet pavement is 2 times that of dry pavement, 5 times that of dry pavement when snowfall and 8 times that of dry pavement when icing.

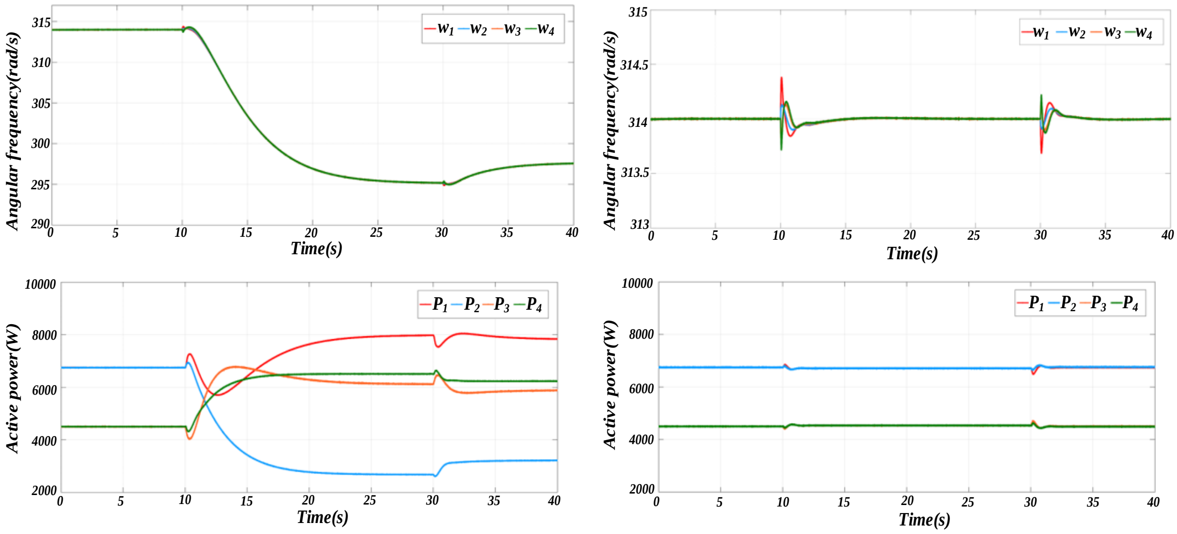

Figure 3: Relationship curve between accident volume and average daily traffic volume

The figure shows the relationship between the number of accidents and the annual average daily traffic volume, which comes from the traffic monitoring system of the expressway. Assuming a non-linear relationship between traffic volume and accident occurrence, we obtain a U-shaped relationship by fitting a curve.

The greater the traffic volume, the higher the accident rate [17]. Because the traffic volume of highway tunnel changes dynamically. Therefore, when studying the influence of traffic volume, the traffic volume of a certain time period is used to replace the real-time traffic volume approximately. Pei Yulong [18] After statistics, the relationship between traffic accident rate and annual average daily traffic volume is obtained, as shown in Figure 3. It can be seen from the figure that the curve between the annual average daily traffic volume and the accident rate is similar to U-shaped, and the accident rate is the lowest when the annual average daily traffic volume is about 11,000 vehicles/day. The capture rate and prediction accuracy formulas are shown in (3) and (4).

$$\gamma_n = \frac{1}{N_n^2} \quad (3)$$

$$L_r(t) = \frac{\sum_k^K \lambda x_k(t)}{R_c(t)} \quad (4)$$

Because the traffic capacity of each section is different, the traffic volume cannot fully represent the road traffic situation. Therefore, it is necessary to analyze the traffic accident rate by a parameter that reflects the congestion degree of road sections, that is, the saturation degree of traffic flow [19].

This article uses the random forest algorithm to analyze tunnel traffic accident data, ultimately achieving a prediction accuracy of 88%. By optimizing hyperparameters such as the number and depth of decision trees, and considering various factors such as traffic flow and weather, the model demonstrated high prediction accuracy on the test set, demonstrating the effectiveness of the random forest algorithm in tunnel accident prediction.

In order to effectively address the performance imbalance between minor and serious accidents in tunnel traffic accident prediction models, the following approaches can be taken: firstly, increasing the sample size of serious accidents through data augmentation techniques, or balancing the dataset by simulating the generation of serious accident data; Secondly, using ensemble learning methods, the random forest algorithm is combined with support vector machines or Bayesian models to optimize the prediction of different types of accidents; Once again, adjust the model evaluation indicators to adopt evaluation criteria such as precision, recall, and F1 score that are more suitable for imbalanced data; In addition, optimize feature engineering by adding features such as weather and road slope that affect serious accidents; Finally, with the help of transfer learning and small sample learning methods, the predictive ability of the model for serious accidents is improved by borrowing knowledge from other fields. These methods can effectively improve the accuracy of the model in predicting serious accidents and solve the problem of imbalance.

3 Study on traffic accident prediction method of road tunnel

3.1 Characteristics of traffic accident data

For this study, we collected tunnel traffic data from specific data source. The data for this study comes from 12 highway tunnels in a province of China, covering 327680 sample data from 2015 to 2020, including traffic monitoring equipment, video analysis, lighting sensors, and accident reports. These data include traffic flow, vehicle speed, weather conditions, etc. During data collection, we performed rigorous data cleaning steps including removal of duplicates, processing of missing values as well as outlier detection. Road traffic accident data have various forms, but generally have the following characteristics and multi-dimensions. When dividing accident liability, accident liability is usually attributed to a certain element. However, it has been proved before that any traffic accident is affected by many factors. In order to effectively describe and predict the complete situation when an accident occurs, it is necessary to make statistics on all these factors. If each factor is expressed as a dimension, traffic accidents have multi-dimensional characteristics. Based on the multi-dimensional statistical data, there exists a weak correlation among the various dimensions. While there may be some disparities in the correlation between each dimension and the final results, a comprehensive analysis reveals that, the correlation between each factor and the accident results is weak, which also confirms the multi-dimensional characteristics of traffic accidents, and it is difficult to explain which factor caused the traffic accidents. The relationship between factors is vague, because the correlation is weak, so the relationship between factors is difficult to identify, and it is difficult to select some factors to explain that traffic accidents are caused by their joint action. Large deviation and small deviation. Large deviation refers to the phenomenon that the mean value of a random variable with integer value is less than its variance. Large deviation is a very common feature of traffic accident data, especially when studying accident duration [20]. Contrary to the phenomenon of large deviation, small deviation refers to the phenomenon that the data is too concentrated and the mean value of data is greater than the variance. This kind of situation is rare, but when the original data is too concentrated and the sample mean is small, the variance will be less than the mean. Poisson regression model requires that mean and variance are equal, so the above two phenomena will lead to large errors when using Poisson regression model. Low sample mean and small sample size, from the probability point of view, traffic accident is a small probability event (P < 0.05), so there will be a large number of zero observation values in accident data, which will lead to low sample mean. When studying local accidents, if there are few accidents in this road section and the record years are short, it will lead to small sample size. If these two situations occur, many models will lose their effectiveness. For example, the maximum likelihood estimation of large sample size is required, and inappropriate models are used rashly to predict, and the results will be meaningless. In order to improve the accuracy of prediction results, scholars have tried many different methods. For example, a logistic regression model for classification is realized by fitting curves (or learning hyperplanes); Bayesian model considering probability for prediction: A stochastic forest model that realizes prediction by finding the best partition features and then learning the sample path; Support Vector Machine (SVM) models for classification by finding the classification hyperplane and maximizing the class spacing, etc. These models will be studied in this chapter.

Although there is weak correlation between various features in traffic accident data, through the feature selection ability of the random forest algorithm, we can extract features with predictive value and construct effective accident prediction models. This proves that although the correlation of data is weak, machine learning algorithms can still find effective prediction patterns in high-dimensional feature spaces.

According to the comparative results of this study, the random forest algorithm is significantly superior to Bayesian models and support vector machines in predicting tunnel traffic accidents. The accuracy of random forest in predicting minor accidents reaches 95%, and it exhibits strong robustness in dealing with data imbalance, avoiding overfitting. In contrast, Bayesian models perform better in predicting severe accidents, but are prone to fluctuations when the sample size is small; SVM has long training time and low accuracy under large and imbalanced data. Therefore, based on prediction accuracy and processing efficiency, random forest is a reasonable choice.

Feature selection is based on correlation analysis with tunnel traffic accidents, selecting factors such as traffic flow, weather conditions, and tunnel slope as input features. We adopted correlation analysis and feature importance assessment methods in the random forest algorithm to ensure that the selected features have the greatest impact on accident prediction.

This article selects factors such as traffic flow, weather conditions, and tunnel slope as the main input features through correlation analysis and feature importance assessment in random forests. These features have a significant impact on predicting the type and severity of accidents.

3.2 BP neural network

The BP neural network, first introduced in 1986 by Rumelhart, is a highly utilized multi-layer feedforward architecture. It consists of three primary layers: the input, hidden, and output layers, each containing multiple neurons. These neurons only receive signals from the previous layer, with no intra-layer connections [21]. In 1989, Hecht-Nielsen proved that a BP network with a single hidden layer can approximate any continuous function over a closed interval. This means that a three-layer BP network (input, hidden, output) can perform any dimension-to-dimension mapping. The learning process involves two main stages: forward signal propagation and error backpropagation, a unique feature of BP neural networks. During forward propagation, signals traverse the input layer, traverse the hidden layer where they undergo processing, and ultimately arrive at the output layer. Once there, the output layer arranges and disseminates the processed information derived from the hidden layer. Nevertheless, should the output results fall short of the anticipated standards, the system transitions into the error backpropagation phase [22]. Currently, the error propagates from the output layer back to the input layer via the hidden layer. Subsequently, the weights and thresholds connecting the hidden layer to the output layer, as well as those connecting the input layer to the hidden layer, are adjusted in sequence. This adjustment aims to minimize the sum of squared errors within the network, ultimately leading to a reduction in errors.

Although this study has achieved certain results in predicting tunnel traffic accidents based on the random forest algorithm, its limitations are also quite obvious, especially in terms of model scalability. Firstly, the dataset used may only be applicable to specific regions or tunnel types, making it difficult to directly generalize to other tunnel environments or different traffic conditions. Therefore, it is necessary to consider how to improve the adaptability of the model. Secondly, the dataset may have imbalances or biases, resulting in poor predictive performance for certain types of accidents. Finally, the random forest algorithm may experience overfitting when facing high-dimensional or imbalanced data. How to optimize feature selection and model parameters to improve prediction accuracy is an important direction for future research.

In practical applications, we selected the input features of the model based on factors such as traffic flow, weather, tunnel slope, etc. In order to optimize the performance of the model, cross validation method was used to select appropriate hyperparameters, and these hyperparameters were adjusted through techniques such as grid search, thereby improving the predictive ability of the model on actual tunnel traffic data.

In this study, in order to alleviate the local minimum problem, we adopted an improved Adam optimizer instead of the traditional gradient descent method to improve the stability of model training. In addition, by using appropriate weight initialization methods, the problem of gradient vanishing or exploding that may occur during the training process is avoided, thereby improving the convergence speed and prediction performance of the model.

This process is the feedback mechanism of BP neural network, that is, the process of network learning, which is until the output error gradually decreases to an acceptable degree or reaches a set number of learning times, thus completing the process of information extraction and memory. After a large number of sample learning and training, the connection weight between neurons in each layer is fixed and begins to enter the working period.

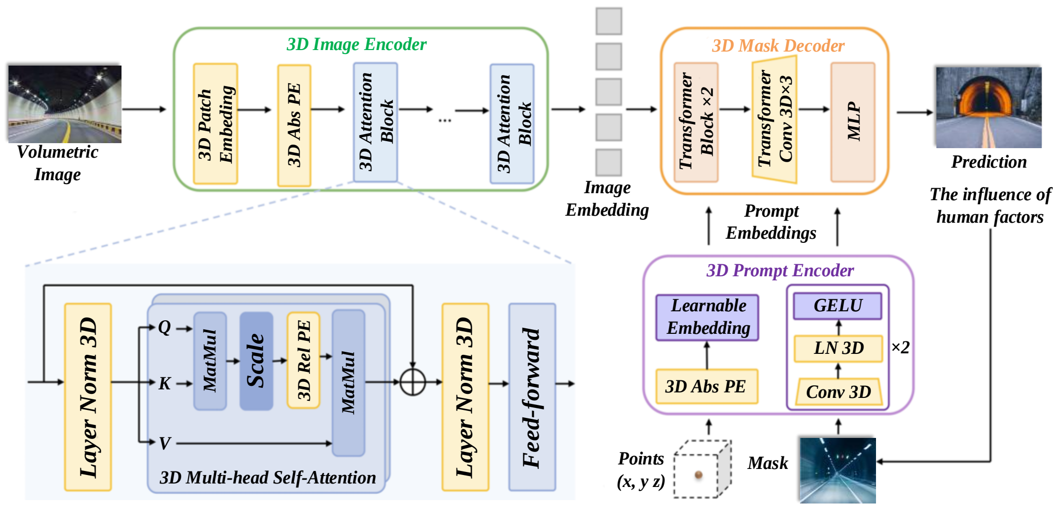

Figure 4: Application flow chart of machine learning in tunnel traffic accident prediction

Figure 4 shows application flow chart of machine learning in tunnel traffic accident prediction. Collect data related to tunnel traffic. This may include information on vehicle flow, speed, weather conditions, equipment status, etc. The collected raw data is cleaned, DE weighted, converted and so on to meet the needs of subsequent modeling.

P learning samples are represented by x_1, x_2, …, x_p respectively. The output $\boldsymbol{O_{pk}}$ (k=1, 2, …, m) is obtained after the P-*th* sample is input into the network, and the expected output $\boldsymbol{t_{pk}}$ is obtained. Using the square error function, the error of the P-*th* sample is $\boldsymbol{E_p}$. The calculation formula is shown in (5).

$$\boldsymbol{E_p} = \frac{1}{2}\sum_{k=1}^{n}\left(\boldsymbol{t_{pk}} - \boldsymbol{O_{pk}}\right)^2 \quad (5)$$

Global errors for the entire sample are shown in (6).

$$\boldsymbol{\pi}(\boldsymbol{X}) = \triangle\,(\boldsymbol{X})\boldsymbol{w}\left(\boldsymbol{\mathcal{L}}(\boldsymbol{X})\right)\boldsymbol{p}^{\boldsymbol{X}} \quad (6)$$

In order to reduce the global error *E*, the cumulative error BP algorithm is used to adjust the change term $\boldsymbol{\Delta_{W_{kj}}}$ of the weight. According to the gradient descent method, $\boldsymbol{\Delta_{W_{kj}}}$ is proportional to $\boldsymbol{\partial E}$, and the expression is as described in (7).

$$\boldsymbol{\Delta_{W_{kj}}} = -\boldsymbol{\eta}\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{w_{kj}}} \quad (7)$$

When constructing the prediction model, we chose such as vehicle flow, vehicle speed, and visibility as input features. These features were selected because they are closely related to the occurrence of tunnel traffic accidents and are able to be quantified by historical data. In terms of algorithm selection, we consider multiple machines learning algorithms, including random forest, decision tree, and support vector machine. Some traffic accident samples have a large sample size, which leads to some errors in the data. Because BP neural network has little influence on the global mapping ability of the network after local neurons are damaged, BP neural network can also realize accurate prediction for a few wrong traffic accident data. The convergence process occasionally experiences temporary slowdowns or stagnation, potentially due to the existence of local minima [23]. Its essence is to find the best weight along the negative gradient direction of the error surface [24]. However, there are some points with zero gradient on the error surface, and local optimal solutions that are not global optimal solutions will be produced at these points. BP neural network at these points will mistakenly think that the error at this point has reached the optimal and stop training.

The implementation process of the random parameter model is completed through the following steps: loading and preprocessing tunnel traffic accident data, including data cleaning, standardization, and other operations; Train the model using the random forest algorithm and simulate different traffic environments using randomly generated parameters; Evaluate the performance of the model on the test set and analyze its predictive ability through metrics such as accuracy and classification reports. The key to this process lies in introducing random parameters to help researchers adjust the model configuration according to the characteristics of different tunnel environments, improving the reliability and accuracy of predictions.

In the data cleaning phase, we first removed duplicate data entries and used the mean imputation method to address missing values. All input data has been standardized before entering model training to ensure that the dimensions of each feature are consistent. In addition, the dataset is divided into a training set and a testing set in an 80%/20% ratio, where the training set is used for model training and the testing set is used to evaluate the model's generalization ability. We also used cross validation methods to further evaluate the stability of the model.

4 Prediction model of traffic accident severity in expressway tunnel

4.1 Establish a prediction model of accident severity

The significance of predicting the severity of highway tunnel traffic accidents lies in the fact that it is easy to infer the situation of serious accidents or even serious accidents according to the severity predicted by the model and remind drivers when this situation occurs to reduce the probability of serious accidents; The model can also be used as a reference when making accident prevention plans.

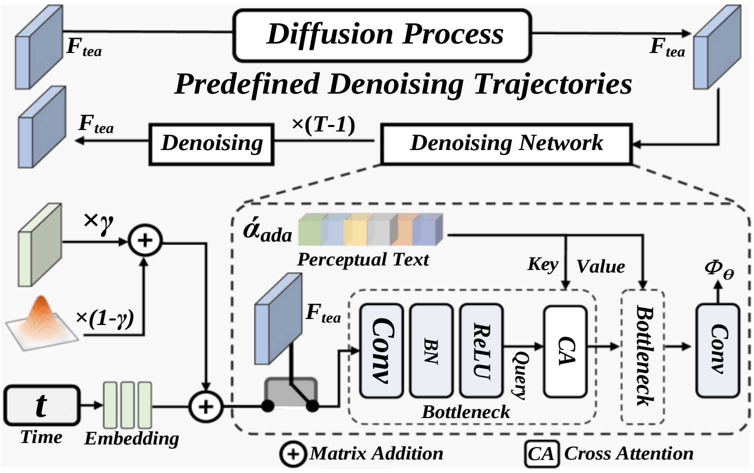

Figure 5: Flow chart of prediction accuracy analysis of tunnel traffic accident stochastic parameter model

Figure 5 shows the flow chart of prediction accuracy analysis of tunnel traffic accident random parameter model. The figure first shows the application of the random forest algorithm in predicting tunnel traffic accidents, and then simulates the probability of accidents in different traffic environments by introducing a random parameter model. Next, the predicted results of the model are compared with actual accident data to calculate various performance indicators such as accuracy, recall, and F1 score. Finally, the chart evaluates the stability and reliability of the model by analyzing the impact of different parameter settings on prediction accuracy, providing a basis for further optimizing the model. In terms of temporal factors, there exists a strong correlation between the timing of accidents and their severity, thereby affirming that nighttime is the period with the highest frequency of severe accidents. When considering traffic volume, the correlation between cart accidents and other independent variables is notably elevated, indicating that cart accidents have a substantial influence on accident severity. In addition, there is a great correlation between weather and speeding, which is in line with common sense. In terms of alignment, slope and slope length also have great influence on the severity of accidents [25]. In addition to accuracy and capture rate, we calculated metrics including model precision, recall and F1 score. Accuracy measures the proportion of instances where the model is predicted to be truly positive, and recall measures the proportion of all instances that are truly positive. The F1 score is then the harmonic mean of precision and recall, used to comprehensively evaluate the performance of the model.

In the analysis of traffic accident factors, this study adds a refined evaluation of driving behavior and lighting conditions. The dimensions of driving behavior include driver fatigue index, distraction time, and pedestrian violation crossing rate; In addition to the basic illuminance, the lighting conditions should focus on quantifying the glare index, color temperature, and the adaptability of the light environment during the day night transition period. The random forest model reveals that there is a significant interaction effect between driver distraction behavior and inadequate adaptation to dusk lighting. When the daily distraction exceeds 127 seconds and the illumination is less than 75lx, the probability of accidents increases exponentially. This result provides key parameter thresholds for dynamic lighting optimization and behavior supervision in tunnels.

According to our data analysis, the severity of nighttime accidents is usually higher. This may be related to factors such as poor lighting conditions inside the tunnel and driver visual fatigue. Specifically, accidents that occur at night have the highest proportion of severity among all accidents, especially in situations of insufficient lighting or driver fatigue, where the consequences of accidents are more severe.

To comprehensively evaluate the predictive performance of the models, we calculated the accuracy, recall, and F1 score of each model separately. The specific results are as follows: for the random forest model, the accuracy is 12%, the recall rate is 20%, and the F1 score is 20%; For the Bayesian model, the accuracy is 22%, the recall is 44%, and the F1 score is 56%. By comparing these indicators, we can better understand the advantages and disadvantages of different models in predicting accident severity.

After analysis and comparison, nine independent variables, such as accident time (X2), accident date (X3), whether it is speeding (X13), slope (X22), slope length (X23) and accident type (Y1), are selected to predict the severity of highway tunnel accident (Y2). The formulas related to traffic flow and traffic accident number and weather conditions and traffic accident number are shown in (8) and (9).

$$\boldsymbol{r^2 \triangle_x = (1 - x^2)\frac{d^2}{dx^2} - (D - 1)x\frac{d}{dx}} \quad (8)$$

$$\boldsymbol{g(x) = \sum_{n=0}^{\infty} \gamma_n \mathcal{P}_n(x)} \quad (9)$$

Because a large part of highway tunnel traffic accidents are minor accidents, the primary task of traffic accident severity model is to predict the possibility of general accidents, major accidents and extraordinary accidents. As before, 496 accidents were randomly selected as training sets and 96 as test sets. In order to reduce the influence of stability on the model, the model is run three times continuously, and the prediction ability of the model is measured by accuracy, such as formula (10).

$$\mathcal{L}_{BT} = \sum_i (C_{ii} - 1)^2 + \lambda \sum_{i \neq j} C_{ij}^2 \quad (10)$$

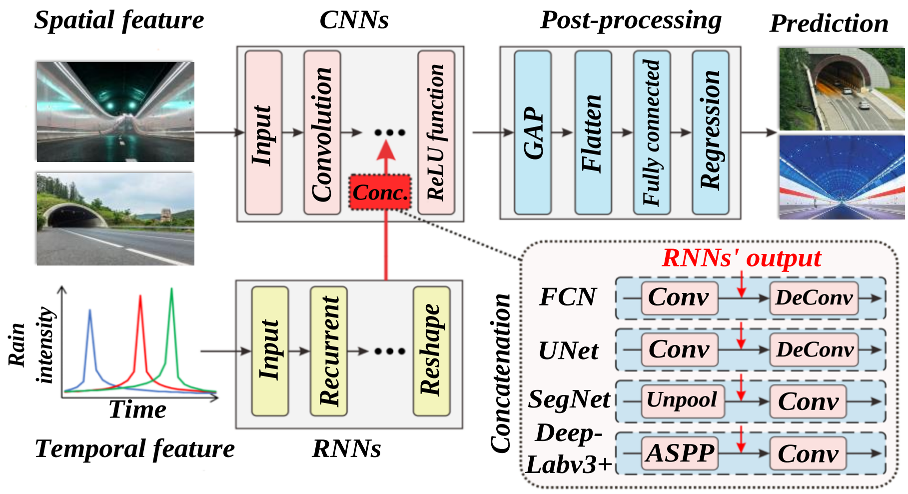

Figure 6: Flow chart of tunnel traffic accident capture rate evaluation

As can be seen from Figure 6, the prediction accuracy of BP neural network for minor accidents is very high, which can reach 90%, and the prediction accuracy for general accidents is 38%. The figure shows the entire process of assessing the capture rate of tunnel traffic accidents. Firstly, collect data on tunnel traffic flow, environmental factors, and perform preprocessing. Next, train the accident prediction model using the random forest algorithm. Finally, by comparing with actual accident data, the capture rate of the model is evaluated, usually using indicators such as accuracy and recall rate. This process helps evaluate the effectiveness of the model in tunnel traffic safety management.

Due to the limited sample size of major and serious accidents, the BP neural network model exhibits a notably low level of accuracy when applied to these types of incidents [26].As far as minor accident prediction is concerned, the prediction ability of Bayesian model is not as good as BP neural network, but because Bayesian model is a classification model based on probability, it has higher prediction accuracy for small sample data, and its prediction accuracy for general accidents is obviously higher than BP neural network model. Out of 288 sets of data, there were 3 major accidents. Notably, the Bayesian model accurately predicted 2 of these as major accidents, while the remaining one was classified as a general accident. Additionally, in the test data, a serious accident (out of 810 potential such incidents) was present, and this was predicted by the Bayesian model as a major accident [27, 28]. It can be seen that Bayesian model has strong prediction ability for serious and serious accidents. When Bayesian model predicts serious accidents, the probability of serious accidents can reach 50%. Stochastic forest model has the highest accuracy in predicting minor accidents, which can reach 95%. It has certain prediction ability in predicting general accidents, but its prediction accuracy is extremely low in predicting major accidents and serious accidents. Support vector machine model is similar to random forest model, but its prediction effect is not good. The model training formula and the model validation formula are shown in (11) and (12).

$$I_{SR} = \frac{1}{n}\sum_{i=1}^{n} T_i^{-1} \mathcal{F}\big(T_i(I_{LR})\big) \quad (11)$$

$$C = \{x' - ax - b, y' - ax - b\} \quad (12)$$

In order to further verify the superiority of the random forest algorithm in predicting tunnel traffic accidents, this study provided confidence intervals and statistical significance tests for various accuracy indicators. By conducting multiple cross validations on the prediction results of the model, we calculated the confidence intervals for various indicators such as accuracy, precision, recall, and F1 score, and used t-tests to compare the differences in prediction accuracy between random forests and other models such as Bayesian models and SVM. The results indicate that the accuracy of random forest in predicting minor accidents is significantly higher than other methods, and the confidence intervals of all indicators indicate that its performance is more stable. In addition, the statistical significance test results (p-value<0.05) further support the superiority of the random forest algorithm in handling tunnel traffic accident prediction tasks. These statistical results provide strong evidence for the application of random forests in tunnel traffic management.

The data for this study is sourced from 12 highway tunnels in a province in southeastern China, covering the latitude range of 26°-29° N. The dataset integrates traffic monitoring equipment, video analysis, lighting sensors, and accident reports, containing 327680 samples that have been standardized, cleaned, and desensitized. The data collection protocol is open source and supports spatiotemporal reproduction.

In the parameter tuning process of the random forest model, we tested multiple key hyperparameters, including the number of trees, the maximum depth of trees, and the minimum number of sample partitions per tree. In order to select the optimal combination of hyperparameters, we used 5-fold cross validation and further optimized the hyperparameters through grid search method.

4.2 Variable filtering

Taking the duration of traffic accidents in highway tunnels as dependent variables, the correlation analysis of each variable is made.

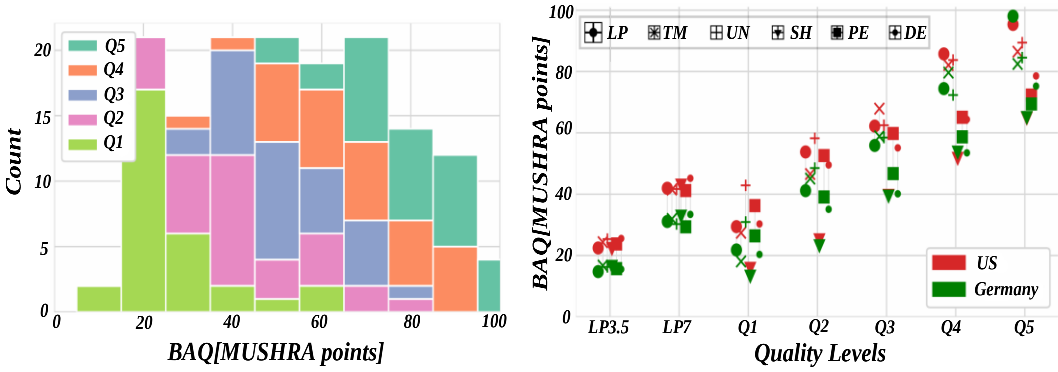

Figure 7: Analysis chart of accident casualties and severity

As can be seen in Figure 7, it is obvious that the casualty situation and severity have the greatest correlation with the accident duration. The classification and analysis of different accident types, casualties, and severity levels in the figure help identify which factors are more likely to lead to serious consequences. By comparing the casualty data of different accident modes, the relationship between the danger and severity of traffic accidents can be intuitively observed, providing data support for the development of more accurate traffic safety measures and accident prevention strategies. In terms of time and space, only the correlation of accident duration is relatively large [29]. Whether it is a cart accident or not, and the total number of vehicles involved have a great correlation. Alignment, gradient, whether in the vertical curve correlation is greater, in addition, although the correlation of accident type is less than casualties and severity, but it is also a very important factor. The random forest algorithm performed well in our analysis, which is mainly attributed to its several key properties. First, random forests make predictions by integrating multiple decision trees, which increases the stability and accuracy of the model. Second, random forests used random feature selection and sample sampling when constructing the decision tree, which helps to reduce the risk of overfitting.

This research method can predict minor accidents with high accuracy, providing practical benefits for improving tunnel traffic safety, especially in preventing frequent but not too serious accidents. The high accuracy of predicting minor accidents can help authorities optimize traffic flow, issue warnings, reduce the occurrence of minor accidents, and thus prevent congestion and secondary accidents. However, the lower performance in predicting serious accidents indicates that although our model is highly effective for daily traffic safety, there is still room for improvement in managing high-risk, low-frequency events. The ability to predict serious accidents may have a significant impact on emergency response strategies and resource allocation.

The results of this study indicate that the random forest algorithm performs well in predicting tunnel traffic accidents, especially in predicting minor accidents, with an accuracy rate of 95%. Although Bayesian models perform well in predicting serious accidents, their applicability in situations of data scarcity and imbalance is poor, limiting their application in practical tunnel traffic management. For tunnel management departments, high-precision prediction of minor accidents is more important because these accidents occur more frequently and can be effectively reduced through early warning and optimization of traffic flow. Therefore, considering the stability and prediction accuracy of the model, the application of random forest in tunnel traffic management is more suitable.

To verify the statistical significance of the prediction results of the random forest model, we calculated confidence intervals for accuracy, recall, and F1 score, and compared them with other benchmark models using t-tests. The results indicate that the accuracy of the random forest model in predicting minor accidents is significantly higher than other models, and the confidence intervals of all evaluation indicators show that its performance is more stable. These statistical analysis results further support the effectiveness of the random forest algorithm in predicting tunnel traffic accidents.

This article calculates the confidence intervals for accuracy, recall, and F1 score. The differences between the random forest model and other benchmark models were compared through t-test, and the results showed that the accuracy of the random forest model in predicting minor accidents was significantly higher than other models (p-value<0.05). These confidence intervals and significance test results further support the effectiveness of random forests in predicting tunnel traffic accidents. In addition to the random forest model, we also used other benchmark models such as logistic regression and support vector machines for performance comparison. The results show that random forest outperforms logistic regression and support vector machine in accuracy, recall, and F1 score. For example, the accuracy of the random forest model is 88%, while logistic regression and support vector machine are 80% and 82%, respectively. This comparison indicates that random forests perform better in predicting tunnel traffic accidents.

To avoid overfitting, we used cross validation during the training process and employed 5-fold cross validation to evaluate the model's generalization ability. Through this method, we ensure the stability of the model on different datasets and avoid overfitting on the training data. In addition, we also monitored the error differences between the training set and the validation set to ensure good performance of the model on unseen data.

5 Summarize

In this paper, the methods of statistical analysis and model prediction are used to study the traffic accidents of highway tunnels. The related literatures of traffic accidents at home and abroad are counted, and various factors affecting the traffic safety of highway tunnels are summarized from five aspects: tunnel alignment, pavement condition, traffic conditions, traffic safety facilities and environmental factors. Future studies could further extend this work from multiple directions. First, we can explore more advanced machine learning algorithms, such as deep learning models, to further improve the prediction accuracy. Second, we can consider incorporating more safety-related factors for traffic safety into the model, such as driver behavior, vehicle characteristics, and road design.

The random forest algorithm has achieved high accuracy in predicting minor accidents, surpassing the prediction accuracy of Bayesian models. However, when predicting serious accidents, our model's performance is not ideal compared to Bayesian models. Bayesian models perform better in capturing serious accidents, with a probability of 50% for major accidents.

The accuracy of the random forest model in accident form prediction is 84%, which is based on the evaluation of the test dataset and ensures the model's generalization ability. In addition, we also tested the model on different training and validation sets, and the results showed that the performance of the model was relatively stable on different datasets, and it could effectively predict accident types.

Although the random forest model performs well in most cases, especially in predicting minor accidents with an accuracy of 95%, its generalization ability slightly decreases in predicting severe accidents with an accuracy of 30%. This difference is mainly due to the limited data on serious accidents, which makes it difficult for the model to learn sufficient features. Therefore, future research should further optimize this aspect.

Utilizing diverse intelligent algorithms, four distinct prediction models are developed for highway tunnel traffic accidents, encompassing accident type, severity, casualty, and duration, all based on the selection of eigenvectors. Then, the optimal algorithm of each prediction model is determined by analyzing the prediction results, and the generalization ability of each model is tested with Qinling No.2 Tunnel as the target. The results show that: The stochastic forest model is used to predict the accident form of expressway tunnel, and the prediction accuracy of the model is as follows 84%. When the test result is a major accident, the probability of serious and extraordinarily serious accidents can reach 50%. Bayesian model is used as the predicting model of highway tunnel casualty, which can be used for early warning successfully 40% of them have casualties. Using stochastic forest model as the prediction model of highway tunnel accident duration, the accuracy rate is 46% when the absolute error is set below 10min, the accuracy rate is 71% when the absolute error is set below 20min, and the accuracy rate is 84% within 30 minutes. Although the generalization ability of each model is different, they all perform well and can be used in practical application. Future studies should validate the proposed model against tunnels constructed with alternative design standards to assess its generalizability.

In order to further improve the accuracy and generalization ability of the model, future research can consider incorporating driving behavior and vehicle features as important features into the model. This direction is of great significance, but it requires addressing the challenge of how to collect this data.

Although advanced machine learning techniques such as deep learning have potential in predicting accuracy, integrating these technologies into current frameworks will face several challenges. Firstly, deep learning models require a large amount of annotated data for training, and traffic accident data is often scarce and unevenly distributed. To address this issue, data augmentation techniques and transfer learning can be used to improve the model's generalization ability. Secondly, deep learning models have high computational complexity, which requires us to optimize hardware resources and computation time. Finally, the 'black box' property of deep learning may affect the interpretability of the model. To address this challenge, future research can explore ensemble methods such as hybrid models of random forests and deep learning to ensure a balance between prediction accuracy and interpretability.

In order to enhance the predictive ability of the model, future research should focus on integrating driver behavior and vehicle characteristics. Driver behavior data can be obtained through onboard sensors, smartphone applications, or onboard cameras, including driver fatigue levels, driving habits, and more. Vehicle characteristic data can be obtained through vehicle networking technology or intelligent transportation systems, covering information such as vehicle type, driving speed, and brake usage. In terms of data processing, data privacy and security should be ensured, and user privacy can be protected through anonymization technology and data encryption measures. Integrating these data can be achieved through multi-level machine learning methods, such as inputting driver behavior as additional features into deep learning models, thereby further improving the model's adaptability to complex traffic environments.

Future research could consider introducing deep learning models, especially convolutional neural networks and recurrent neural networks, to better capture the complex nonlinear relationships and temporal dependencies in tunnel traffic accident data. Deep learning models can reduce the workload of manual feature selection through automatic feature extraction, and have higher generalization ability when dealing with large-scale data. In addition, the potential of deep learning in predicting serious accidents, especially in situations with limited samples, can be significantly improved through incremental learning and transfer learning techniques.

Acknowledgment

Key Area Research and Development Program of Guangdong Province+2022B0101070001: R&D of key technologies and equipment for intelligent management and control of highway network and tunnel engineering operation safety and emergency support

References

[1] Sun, Y., Lü, Lin, Cai, Y. K., & Lee, P. (2022). Prediction of black carbon in marine engines and correlation analysis of model characteristics based on multiple machine learning algorithms. Environmental Science and Pollution Research.

[2] Liu, H., Wang, L., Shi, Y., Wang, X., Chang, F., & Wu, Y. (2023). A deep learning-based method for detecting granular fertilizer deposition distribution patterns in centrifugal variable-rate spreader fertilization. Computers and Electronics in Agriculture, 212.

[3] Rezaei, Z., & Ebrahimpour-Komleh, H. (2021). Prediction and Detection of car accidents in video by deep learning. 2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA).

[4] Rongqiang, S., Yan, C., Chang, D., & Qingyue, G. (2020). Research on promoting the application of disease prediction system based on machine learnin. ISAIMS 2020: 2020 International Symposium on Artificial Intelligence in Medical Sciences.

[5] Yang, Y., Du, Z., Alonso, F., Faus, M., & He, S. (2024). Why frequent traffic accidents at highway tunnel exit? – An experimental analysis of the slack effect. Tunnelling and Underground Space Technology, 152, 105927.

[6] Huo, R., Chen, J., Zhang, Y., & Gao, Q. (2025). 3D skeleton aware driver behavior recognition framework for autonomous driving system. Neurocomputing, 613, 128743.

[7] Maji, A., & Ghosh, I. (2025). A systematic review on roundabout safety incorporating the safety assessment methodologies, data collection techniques, and driver behavior. Safety Science, 181, 106661.

[8] Liu, W., Li, Y., Zhang, T., Gao, Y., Wei, L., & Chen, J. (2025). CCAF-Net: Cascade Complementarity-Aware Fusion Network for traffic accident prediction in dashcam videos. Neurocomputing, 624, 129285.

[9] Lee, H. J., Kim, H., Kim, Y. T., Won, K., & Kim, D. J. (2021). Prediction of life-threatening intracranial hypertension during the acute phase of traumatic brain injury using machine learning. IEEE Journal of Biomedical and Health Informatics, PP (99), 1-1.

[10] Hermawan, I., Firdausy, C. M., Rambe, K. R., Zuhdi, F., Erwidodo, Nugraheni, R. D., Malisan, J., Isnasari, Y., Marpaung, E., & Asshagab, S. M. (2024). Road traffic facilities, traffic accidents, and poverty: Lesson learned from Indonesia. Transportation Research Interdisciplinary Perspectives, 28, 101273.

[11] Chen, Z., Lu, Z., Chen, Q., Zhong, H., Zhang, Y., & Xue, J., et al. (2022). A spatial-temporal short-term traffic flow prediction model based on dynamical-learning graph convolution mechanism. arXiv e-prints.

[12] Dastanian, M., & Saberian, J. (2022). Assessment of Vegetation Mapping Methods Using GIS and Remote Sensing. Journal of Remote Sensing and Geoinformation Research, 1(1), 83-90. doi: 10.22061/jrsgr.2022.1979

[13] Mao, J., Su, X., Wang, H., Yan, H., & Zong, H. (2023). Identification of closely spaced modes of a long-span suspension bridge based on bayesian inference. International Journal of Structural Stability and Dynamics, 23(20).

[14] Yu, R., Han, L., & Zhang, H. (2021). Trajectory data-based freeway high-risk events prediction and its influencing factors analyses. Accident Analysis & Prevention, 154(1), 106085.

[15] Lo, W. C., Chen, Y. H., & Jhong, S. Y. (2020). Systems and Methods for Creating an Optimal Prediction Model and Obtaining Optimal

Prediction Results Based on Machine Learning. US20200074325A1.

[16] Aghamohammadi, H. (2022). Comparison of Spatially Guided Tabu Search and Nested Genetic Algorithms for Solving the Capacitated Location-Allocation Problem in Emergency Situations. Journal of Remote Sensing and Geoinformation Research, 1(1), 91-102. doi: 10.22061/jrsgr.2022.1980.

[17] Yu, S. , Wu, H. , Xie, J. , & Xia, X. . (2021). Depth Factor Decomposition Machine Based on Multi-kernel Convolutional Neural Network for Click-Through Rate Prediction. BDE 2021: The 2021 3rd International Conference on Big Data Engineering.

[18] Mao, W. (2021). Video analysis of intelligent teaching based on machine learning and virtual reality technology. Neural Computing and Applications (4).

[19] Yang, Y., Du, Z., Alonso, F., Faus, M., & Wan, H. (2025). Traffic safety improvement method for highway tunnel entrances based on linear guiding − An engineering practice from China. Tunnelling and Underground Space Technology, 156, 106267.

[20] Ruan, D., Wang, J., Yan, J., & C. Gühmann. (2023). Cnn parameter design based on fault signal analysis and its application in bearing fault diagnosis. Adv. Eng. Informatics, 55, 101877.

[21] Wei, N., Yuan, J., & Meng, C. (2020). Application Analysis of Mobile Learning Based on Smartphone in Flipped Classroom. 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI).

[22] Silavi, T., & kheirkhah zadeh, M. (2022). Using Artificial Neural Networks and Geospatial Information Systems in Mine Exploration for Producing Mineral Potential Map. Journal of Remote Sensing and Geoinformation Research, 1(1), 103-114. doi: 10.22061/jrsgr.2022.1981.

[23] Nam, N. K. T., Luu, T. M. N., Hung, T. H., & Huong, N. T. M. (2025). Productivity convergence and the role of innovation in productivity convergence: The case of processing and manufacturing enterprises in Vietnam. Procedia Computer Science, 253, 266–276.

[24] Dai, Y., & Li, Q. (2022). Semantic detection of vehicle violation video based on computer 3d vision. Advances in multimedia (Pt.1), 2022.

[25] Zhang, R., Liu, Y., Wang, Z., Chen, J., Zeng, Q., Zheng, L., Zhang, H., & Pei, Y. (2025). Innovative prediction and causal analysis of accident vehicle towing probability using advanced gradient boosting techniques on extensive road traffic scene data. Accident Analysis & Prevention, 211, 107909.

[26] Yin, S., Hu, Y., & Peng, R. (2020). Research on the Intelligent Image Recognition and Diagnosis Technology of Key External Insulation Devices Based on Deep Learning. 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS). IEEE.

[27] Liao, R., Zhai, J., & Zhang, F. (2024). Optimization model based on attention mechanism for few-shot image classification. Machine Vision and Applications, 35(2).

[28] Kosari, A., Saghamanesh, M., & Ahmadi, A. (2022). Rapid sizing the Structure Subsystem of a Very High-Resolution Satellite Based on the Design for System Performance Strategy. Journal of Remote Sensing and Geoinformation Research, 1(1), 115-126. doi: 10.22061/jrsgr.2022.1982.

[29] Perpiñá, J. L. E., Gutiérrez, I. L., & Rodríguez, S. F. (2019). Dirección por objetivos, más allá de la informática en el ámbito de atención primaria: ¿qué podemos cambiar? FMC - Formación Médica Continuada en Atención Primaria, 26(1), 30–32.