Safety-Critical Event Identification on Mountain Roads for Traffic Safety and Environmental Protection Using Support Vector Machine with Information Entropy

Zihao Wen, Hui Zhang and Ronghui Zhang *

Guangdong Provincial Key Laboratory of Intelligent Transport System, School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, China; wenzh6@mail2.sysu.edu.cn (Z.W.); zhanghui@mail.sysu.edu.cn (H.Z.)
* Correspondence: zhangrh25@mail.sysu.edu.cn; Tel.: +86-18138734181

Abstract: Traffic accidents, which cause loss of life and pollution, are a social concern. The complex traffic environment on mountain roads increases the harm caused by traffic accidents. This study aimed to identify safety-critical events related to accidents on mountain roads to understand the causes of the accidents, improve traffic safety, and protect the environment. In this study, a naturalistic-driving data collection system, consisting of approximately 8000 km of naturalistic-driving data from 20 drivers driving on mountain roads, was developed. Using these data, a comparative analysis of the identification performance of the support vector machine (SVM), backpropagation neural network (BPNN), and convolutional neural network (CNN) methods was conducted. The SVM was found to yield optimal performance. To improve the identification performance, the yaw rate and information entropy of the data were added as input variables. The improved SVM method yielded an identification accuracy of 90.64%, which was approximately 15% higher than that yielded by the traditional SVM. Moreover, the false positive and false negative rates of the improved SVM were reduced by approximately 10% and 20%, respectively, compared with the traditional SVM. The results demonstrated that the improved SVM method can identify safety-critical events on mountain roads accurately and efficiently.

Keywords: traffic safety and environmental protection; naturalistic driving study; safety-critical events identification; support vector machine; information entropy

1. Introduction

The rapid expansion of transportation networks has made travel more convenient. However, it has also led to an increased frequency of traffic accidents. The occurrence of traffic accidents directly causes loss of life and property damage. According to the 2018 statistics from the World Health Organization, the number of road traffic accident deaths has risen steadily since 2000, reaching 1.35 million in 2016. Moreover, the road traffic accident death rate was 18 per 100,000 people from 2000 to 2016 [1]. Additionally, traffic accidents lead to traffic jams in many cases, which in turn result in increased travel costs and pollution. The Texas Transportation Institute found that traffic jams cause delays ranging from approximately 12h to 67 h per person every year. Furthermore, a significant portion of pollution attributable to traffic occurs during traffic jams [2]. In particular, the consequences of hazardous material (Hazmat) transportation accidents are relatively more severe. A US Department of Transportation report concluded that injuries and fatalities account for most of the costs associated with Hazmat transportation accidents [3]. With respect to environmental degradation, Hazmat transportation accidents majorly cause air pollution, followed by water pollution and biological damage [4]. In conclusion, traffic accidents cause loss of life and property damage, environmental pollution, and additional travel costs, all of which are not conducive to sustainable transportation.
This necessitates the prevention and mitigation of traffic accidents. Mountain roads typically exhibit poor alignment, blind bends, steep slopes, and high-risk factors. Therefore, most accidents that occur on mountain roads are severe. For example, according to the Annual Report on Road Traffic Accidents, 46 traffic accidents with more than one casualty occurred on mountain roads, and 18 major accidents with more than 10 casualties occurred between 2004 and 2010 in Yunnan Province, China. Consequently, avoiding and reducing traffic accidents on mountain roads has drawn increasing attention in societies.

Researchers have mainly investigated traffic accidents from four perspectives: people [5], vehicles [6], roads [7], and the environment [8]. However, transportation is a human–vehicle–road–environment interaction system. The naturalistic driving study (NDS) has been proposed as a reliable method to study this system. NDS refers to the use of high-precision data acquisition systems to observe and record driving-related data, including the driving status of the driver and vehicle, the road traffic environment under naturalistic driving conditions, and an analysis of traffic problems based on these data [9]. A naturalistic-driving data collection system with high accuracy, robust real-time performance, and multiple data sources can also facilitate a thorough scientific investigation of traffic accidents from multiple aspects. However, most current NDSs focus on traffic environments, such as expressways and urban highways, and not on mountain roads. Therefore, this study aims to develop a method for identifying driving risk on mountain roads based on naturalistic driving data.

2. Review

Following the increase in traffic accidents in recent years, NDS is becoming a widely and frequently researched topic. The National Highway Traffic Safety Administration launched the Next Generation Simulation [10] and the 100-Car-Naturalistic Driving Study [11] projects in early 2000. The United States launched Phase II of the Highway Strategic Research Plan in 2007 [12]. Tongji University, China, conducted a pilot test of the China Field Operational Test [13–15] in February 2013. The Shanghai Naturalistic Driving Research (SH-NDS), a collaboration between Tongji University, General Motors China, and Virginia Tech Transportation Institute [16,17], was conducted from December 2012 to December 2015.

Investigating the causes and natures of traffic accidents is crucial to improving traffic safety and protecting the environment. A considerable amount of time and a wide range of observations, whose acquisition is challenging, are required to obtain a sufficient sample size for the investigation of traffic accidents. Moreover, the occurrence of traffic accidents is a low-probability event; therefore, statistical analyses in naturalistic driving experiments are hindered by the inevitably small sample size of accidents. Accordingly, researchers typically utilize safety-critical events as a substitute indicator of traffic accidents. A safety-critical event is any situation that requires a collision avoidance response from the driver and comprises opposing objects and collision risks, including near and actual crashes [18]. In short, it refers to a situation where the probability of an accident is high. Accidents and safety-critical events have similar causal mechanisms [19] and a strong frequency correlation [20]; hence, safety-critical events can effectively represent accidents.

Presently, based on naturalistic driving data, there are three main methods of identifying safety-critical events: the use of a traditional threshold, the introduction of classification algorithms to improve the traditional threshold, and the use of a classification algorithm. The traditional threshold method achieves the highest false-positive rate. The overall false positive rate of the method exceeds 80%, and substantial manual checking and screening are required in the later period [18,21]. An improved algorithm has been proposed for the traditional threshold method. For example, after the initial screening using the threshold method, Wu and Jovanis [22] used the Chow test to filter out invalid events, as well as the receiver operating characteristic (ROC) curve to determine the optimal threshold of vehicle kinematics. The second round of screening significantly reduces the manual verification work. Kluger et al. [23] combined the discrete Fourier transform and k-means clustering method to identify the law of vehicle acceleration changes with time before
and after safety-critical events and reduced the false positive rate to 22%. All the studies assumed that collected events provide a complete sample for reducing the false positive rate; however, the threshold method exhibited a certain false-negative rate and cannot cover the entire sample.

Consequently, more studies focused on classification algorithms, such as the support vector machine (SVM), backpropagation neural network (BPNN), and convolutional neural network (CNN). The SVM method mainly solves classification and regression problems and has been widely used in recent traffic research, such as traffic flow prediction [24], incident detection [25], and accident frequency prediction [26]. The SVM can operate with small and medium sample sizes. Moreover, it efficiently obtains the nonlinear relationships among data features, but it is more sensitive to the lack of data. The BPNN is a multi-layer feed-forward neural network trained according to the error backpropagation algorithm, and it is a widely used neural network. The algorithm is often used for feature classification in the transportation field [27]. It is highly effective in strong nonlinear mapping and has a remarkable fault tolerance; however, the algorithm easily falls into local minimums and has high selective requirements for learning samples. A CNN is a type of feed-forward neural network that includes convolution calculations and has a deep structure. It is often used in feature extraction and identification of events in traffic [28,29]. The algorithm learns effectively from samples of the corresponding features, avoiding the complicated feature extraction process. The disadvantage of the algorithm is that it requires large samples, and the encapsulation of the feature extraction process complicates performance improvement.

However, the abovementioned naturalistic driving study [10–23] mainly focuses on traffic environments, such as expressways and urban highways, and only a few studies focus on mountain roads. Based on the naturalistic-driving data, the accurate and efficient identification of safety-critical events on mountain roads is vital for analyzing the characteristics and causes of mountain road traffic accidents, as well as the improvement of mountain road traffic safety and environmental protection. Therefore, this study focused on developing a method for identifying safety-critical events in the complex traffic environments of mountain roads based on naturalistic-driving data.

3. Materials and Methods

In this section, key fragments were extracted from the original naturalistic-driving data on mountain roads to create a sample set of safety-critical events. Based on the naturalistic-driving data on mountain roads, an SVM model, a BPNN model, and a CNN model were built and compared to determine the most suitable method to identify safety-critical events on mountain roads.

3.1. Naturalistic Driving Experiment

In this study, a set of vehicle-mounted integrated systems for naturalistic-driving data acquisition was developed, as illustrated in Figure 1. Throughout the experiment, the system was installed on the test vehicle to continuously collect and record the driver’s operating behavior, vehicle driving status, and road traffic environment information. The collected data included three-axis acceleration, longitudinal velocity, three-axis angular velocity, latitude and longitude, altitude, distance to the vehicle ahead, a video of the upper body and feet of the driver, and a video of the front and rear of the vehicle.
Figure 1. Naturalistic-driving data acquisition system.

The specifications of the equipment used in the aforementioned data acquisition system are listed in the Table 1.

Table 1. Equipment specifications.

| Equipment                  | Specification                                      |
|----------------------------|----------------------------------------------------|
| Cameras                    | 1080P image clarity                                |
| Inertial Navigation        | Acceleration resolution: 0.001 g \(^1\)           |
| Millimeter-wave Radar      | Angular velocity resolution: 0.5\(^2\)/s \(^2\)   |
| GPS                        | Measuring range: 1~175 m \(^3\)                   |
| Industrial Computer        | The error radius: 1 m \(^3\)                      |
|                            | Supply voltage: 12 V \(^4\)                      |

The gravitational acceleration, \(^1\) second, \(^3\) meter, \(^4\) volt.

The naturalistic-driving experiment was performed in the mountainous area of Yunnan Province, China. The main road was a two-way, two-lane secondary mountain road. Approximately 8000 km worth of naturalistic driving data were collected from 20 drivers. The data included 1022 data fragments of various traffic events such as meeting, overtaking, and passing through the tunnel. The obtained traffic events were manually verified, then 319 valid safety-critical events and 703 general events were confirmed. The naturalistic driving data of these traffic events were intercepted and labeled (safety-critical events = 1, general events = 0) to create a sample set for the study.

3.2. Identification Methods

In methods’ studies, the original data was two-dimensional sequence data. The longitudinal direction was the time series, and the lateral direction was the vehicle kinematics parameters corresponding to the current moment, including longitudinal acceleration, lateral acceleration, vertical acceleration, roll rate, pitch rate, yaw rate, roll angle, pitch angle, yaw angle, and velocity. The identification methods were all based on the same original data set (naturalistic-driving data on mountain roads).
3.2.1. Support Vector Machine

The SVM method is a generalized linear classifier that performs binary classification on data according to supervised learning. The basic idea is to determine the separating hyperplane that can divide the dataset correctly and with the largest geometric interval. For a linearly separable dataset, there are infinitely many separating hyperplanes, but the separating hyperplane with the largest geometric interval is unique [30]. The hyperplane is expressed in Equation (1):

\[ W^T \cdot X + b = 0 \]  

(1)

where \( X \) is the vector composed of input variables, and \( W^T \) and \( b \) are the parameters sought.

The SVM model seeks to solve the optimization problem shown in Equation (2) [31]:

\[
\begin{align*}
\min_{W, b, \epsilon} & \left( \frac{1}{2} W^T W + C \sum_{i=1}^{N} \epsilon_i \right) \\
y_i (X_i W^T + b) & \geq 1 - \epsilon_i \\
i = 1, 2, \ldots, N
\end{align*}
\]

(2)

where \( \epsilon_i \) is the slack variable of the sample \( i \), and the constant \( C \) is the penalty factor.

Using the Lagrange multiplier method, the optimization problem is transformed to Equation (3):

\[
\max W(\alpha) = -\frac{1}{2} (\alpha_i y_i \alpha_j y_j X_i \times X_j) + \sum_{i=1}^{N} \alpha_i \\
0 \leq \alpha_i \leq C \\
\sum_{i=1}^{N} \alpha_i y_i = 0
\]

(3)

where \( \alpha \) is the Lagrange multiplier.

These formulas are based on a linear classification. The nonlinear classification problem in the input space can be transformed into a linear classification problem in a certain dimensional feature space through a nonlinear transformation, as shown in Figure 2.

![Figure 2. Nonlinear transformation of the support vector machine (SVM).](image)

A kernel function and a linear SVM were trained in a high-dimensional feature space. The Gaussian kernel (radial basis function) was used in this study, and the inner product of the sample points \( X_i \) and \( X_j \) was recorded as \( k(x, z) \). The kernel function used in this study is shown in Equation (4):

\[ k(x, z) = \exp\left(-\frac{\| x - z \|^2}{2\sigma^2}\right) = \exp\left(-\gamma \| x - z \|^2\right) \]

(4)

where \( \gamma \) is the kernel function parameter.

In summary, the SVM model has two undetermined parameters, \( C \) and \( \gamma \). \( C \) is the penalty coefficient, which adjusts the weight of the interval size and identification accuracy.
preference in the optimization direction; that is, it is tolerance to error. Therefore, the larger C is increasing the chances of over-fitting, the less error is tolerated. The smaller C is increasing the chances of easier under-fitting. If C is too large or too small, the generalization ability is attenuated. γ is a parameter of the radial kernel function and determines the distribution of the data mapped to the new feature space. The larger γ is the fewer of the support vectors; the smaller γ is the more of the support vectors. The number of support vectors affects the speed of training and prediction.

The ten-fold cross-validation method was used in this study to compare the classification errors of the training set and select the best combination of parameters [27]. This validation method divides the original data into 10 groups and creates a validation set for each subset. Meanwhile, the remaining nine groups of data are used as the training set to obtain 10 models. The average identification accuracy of the validation sets of the 10 models is taken as the performance index of the model under this parameter combination. This method can effectively avoid over-learning and under-learning.

Since traffic events are a continuous process, the corresponding data in the sample set are time series data containing many vehicle kinematics parameters. The feature vector of the sample data needs to be inputted when SVM is used to identify the safety-critical events in the sample set. Therefore, it is necessary to parameterize the time series in the sample to form a parameter vector to input into the SVM model. Hence, the statistical values of the vehicle kinematics data were used as the input variables of the SVM model. The input variables were average (avg), standard deviation (std), maximum (max), minimum (min), and amplitude (amp) of the longitudinal acceleration (Xaccel), lateral acceleration (Yaccel), and velocity (V), based on related research [18,32], as shown in Table 2. The calculation time domain of the corresponding parameters is the duration of the event.

**Table 2. Summary of input variables of the support vector machine model.**

| Vehicle Dynamic Parameters | Input Variables |
|---------------------------|-----------------|
| Xaccel, Yaccel, V         | Xaccel_avg, Xaccel_std, Xaccel_max, Xaccel_min, Xaccel_amp; Yaccel_avg, Yaccel_std, Yaccel_max, Yaccel_min, Yaccel_amp; V_avg, V_std, V_max, V_min, V_amp. |

Due to the requirements for input variables of the SVM model, the SVM model’s input variables were the feature vectors constructed by extracting basic features from the original data (average, standard deviation, maximum, minimum, and amplitude of the longitudinal acceleration, lateral acceleration, and velocity, based on related research [18,32]). Therefore, the data used in the SVM model were the original data after the preliminary feature extraction. The preliminary feature extraction of the original data was also a part of the method.

### 3.2.2. Back Propagation Neural Network

A BPNN is a multilayer feed-forward network trained according to the error back-propagation algorithm [27]. The BPNN is divided into the input layer, hidden layer (fully connected layer), and output layer. In the forward pass, the input signal is processed by the hidden layer to the output layer. If the output layer does not yield the desired output, it switches to backpropagation. According to the prediction error, it adjusts the network weights and thresholds in order to yield an output that is closer to the expected output. In this study, a BPNN was built based on the Tensorflow framework by setting the number of hidden layers, the number of neurons in each layer, and the neuron activation function. According to the characteristics of the input data of the neural network, the structure of the BPNN was determined to be 12-24-1; that is, the input layer dimension is 12, the number of hidden layer neurons is 24, and the number of output categories is 1, as shown in Figure 3.
Figure 3. Structure of the backpropagation neural network (BPNN).

The number of neurons and hidden layers are mainly selected according to the data characteristics [27]. Commonly used neuron activation functions include sigmoid, linear rectification function (ReLU), and Tanh functions. The ReLU function can considerably accelerate the convergence speed and does not encounter the gradient disappearance problem. Therefore, the ReLU function was selected as the hidden layer activation function in this study. The ReLU function is expressed in Equation (5):

$$f(x) = \max(0, x)$$  \hspace{1cm} (5)

where $x$ is the current neuron input value $\sum w_i x_i + b_j$.

The sigmoid function is monotonous and continuous, the output range is limited, and the data are not easily diverged in the process of transmission. In this study, the sigmoid function was used as the activation function of the output layer and is expressed in Equation (6):

$$g(x) = \frac{1}{1 + e^{-x}}$$  \hspace{1cm} (6)

where $x$ is the current neuron input value $\sum w_i x_i + b_j$.

The BPNN continuously updates the weight $w$ and deviation through error backpropagation to optimize the network model. The optimization process is mainly affected by factors such as the error function, optimization algorithm, and learning rate. For the neural network training, the binary cross-entropy loss function was selected in this study for error backpropagation. The function is expressed in Equation (7):

$$loss = -\frac{1}{n} \sum_i \left( t[i] \times \log(o[i]) + (1 - t[i]) \times \log(1 - o[i]) \right)$$  \hspace{1cm} (7)

where $t$ is the label and contains only 0 or 1; and $o$ is the input, which contains decimals between 0 and 1; both variables are similar in size. The RMSprop optimization algorithm was used in this study to iteratively update the ownership value and deviation, and the learning rate was set to 0.001. Since the BPNN model had the same requirements for input variables as the SVM model, the input variables of the BPNN model were the same as those of the SVM model.

3.2.3. Convolutional Neural Network

A CNN is a type of feed-forward neural network that includes convolution calculations and has a deep structure [28]. The input data of CNN is processed by convolution and pooling of multiple convolutional layers. By conveying the extracted feature vector
into the fully connected layer, the classification result can be obtained. Convolutional neural networks are mainly composed of convolutional layers and other connected layers. Convolution is a mathematical operation of weighted average. The weight function of $w$ is usually expressed as Equation (8):

$$s(t) = (x \times w)t$$

where function $x$ in the convolution is called the input, and $w$ is called the kernel function.

The CNN used in this study was based on the Tensorflow framework. According to the characteristics of the input data of the neural network, the structure of the CNN neural network was selected, as shown in Figure 4, and the size of the convolution kernel was $3 \times 1$.

![Figure 4. Structure of the convolutional neural network (CNN).](image)

The activation function of the convolutional layer and the fully connected layer was the ReLU function (Equation (5)); the activation function of the output layer was the sigmoid function (Equation (6)); the loss function was a binary cross-entropy loss function (Equation (7)). The SGD optimization algorithm was used in this study to iteratively update the ownership value and deviation, and the learning rate was set to 0.001.

CNN is a neural network specially used for processing data with similar grid structures. The CNN model could automatically extract features from the original data. The input variables of the CNN model were the original data. In this study, the input variables of the CNN comprised 1022 entries of $800 \times 10$ two-dimensional sequence data with labels of 0 and 1. The longitudinal direction was the time series, and the lateral direction was the vehicle kinematics parameters corresponding to the current moment, including longitudinal acceleration, lateral acceleration, vertical acceleration, roll rate, pitch rate, yaw rate, roll angle, pitch angle, yaw angle, and velocity.
3.3. Evaluation Metrics

The three methods were all based on the same original data set (naturalistic-driving data on mountain roads). All three methods were designed to identify safety-critical events. It was meaningful to compare their identification performances.

We used the calibrated traffic events described in Section 3.1 as the full sample set and randomly divided the data into training and test sets in ratio 8:2. The training set was used to train the model, and the test set was used to test the identification effect of the trained model. After training the SVM, BPNN, and CNN models, predictions were made, and the classification performances of the three models were compared based on the identification accuracy, false-positive rate, false-negative rate, and the receiver operating characteristic (ROC) curve. This study addresses a binary classification problem (safety-critical events or general events), and the possible classification results are shown in Table 3.

| Actual Outcomes | Predicted Outcomes |
|-----------------|--------------------|
| 1 (safety-critical events) | True positive, \( T_p \) |
| 0 (general events) | False negative, \( F_N \) |
| 1 (safety-critical events) | False positive, \( F_P \) |
| 0 (general events) | True negative, \( T_N \) |

The evaluation metrics presented in Table 3 for the identification effect are calculated as follows:

1. Identification accuracy \( A_{CC} = \frac{T_p + T_N}{T_p + F_p + F_N + T_N} \);
2. False-positive rate \( R_{FP} = \frac{F_P}{F_P + T_N} \);
3. False-negative rate \( R_{FN} = \frac{F_N}{T_p + F_N} \);
4. AUC (Area under Curve) of ROC curve.

The ROC curve can reflect the trend of sensitivity and specificity of the model and is often used to evaluate the performance of classifiers in the transportation field [27,33,34]. The abscissa of the ROC curve is specificity, and the value is \( 1 - R_{FN} \); the ordinate is sensitivity, and the value is \( 1 - R_{FP} \). The trained model can obtain the predicted probability for each test sample. The set threshold was \( p \in [0, 1] \). If the predicted probability of a sample is less than \( p \), it is classified as a general event; if it is greater than \( p \), it is classified as a safety-critical event. Different values of \( p \) produce different specificities and sensitivities. When \( p \) changes from 0 to 1, several sets of specificities and sensitivities form a ROC curve. The prediction effect of the model can be measured by the area (AUC) enclosed by the ROC curve and the coordinate axis. The larger the AUC, the better the prediction effect.

4. Results and Model Enhancement

4.1. Identification Results

The ROC curves of the three learning models after training and prediction are shown in Figure 5. As shown in the figure, the AUC values of the SVM, BPNN, and CNN were 0.8055, 0.7020, and 0.6336, respectively, indicating that the SVM model yielded optimal prediction results.

As shown in Table 4, the SVM model achieved the highest identification accuracy and lowest false-positive and false-negative rates. This result indicates that SVM is more suitable for the identification and classification of safety-critical events on mountain roads.

The following is the analysis of the result. Advanced models, such as BPNN and CNN, generally have higher requirements for application conditions. For example, a large number of samples are required for training to improve generalization [27,28]. Basic models, such as SVM, are generally not so harsh on the application conditions. An SVM with a small sample size can also achieve improved classification results [33]. In the naturalistic-driving study, the sample size of safety-critical events is generally limited, and it is difficult to reach tens of thousands or hundreds of thousands of big data. Therefore,
in this study’s application condition, the SVM model is more suitable compared with the other two models.

![ROC Curve](image)

**Figure 5.** Receiver operating characteristic (ROC) curve of the support vector machine (SVM), backpropagation neural network (BPNN), and convolutional neural network (CNN).

**Table 4.** Comparison of prediction effects of SVM, BPNN, and CNN.

| Models | Predictive Performance Metrics |
|--------|-------------------------------|
|        | $A_{CC}/\%$ | $A_{UC}$ | $R_{FN}/\%$ | $R_{FP}/\%$ |
| SVM    | 75.66          | 0.8055   | 32.28        | 17.14        |
| BPNN   | 66.34          | 0.7020   | 55.56        | 53.97        |
| CNN    | 60.09          | 0.6336   | 68.42        | 27.37        |

4.2. SVM Model Enhancement

Although the SVM model yielded optimal identification accuracy, the accuracy is still low at 80%, and the underreporting rate is high at 30%. This shows that the model did not achieve a remarkable predictive effect. Additionally, the current input variables did not sufficiently cover the characteristics of safety-critical events. Therefore, the data were further analyzed, and mountain roads were found to have poor alignment, narrow roads, several sharp bends, steep slopes, short visual distances, and mostly two-way double lanes. Therefore, it was estimated that driving stability would have a greater impact on driving risk identification. For example, during a high-risk meeting (safety-critical event), an oncoming vehicle would be relatively closer, and the speed would be low, but the driver’s mental load would be at a high level and result in unstable speed. When occupying a lane during overtaking (safety-critical event), there would likely be an oncoming vehicle, or the vehicle ahead fails to give way, and the acceleration would be low. However, the driver’s hesitation during the operation would lead to unstable longitudinal and lateral acceleration. Therefore, this study introduced more relevant input variables to train the models for improved identification effects.

A vehicle kinematics parameter (yaw rate) was added to the input variable of the SVM because mountain roads have poor alignment and many curves, and the probability of safety-critical events occurring in curve roads is relatively high. Based on the study by Dingus et al. [18], Sudweeks et al. [35] established a yaw rate classifier that filtered 42% of invalid events identified by the yaw rate threshold, indicating that the yaw rate can be used as a basis for identifying safety-critical events.

Furthermore, we used the sample data of naturalistic driving traffic events on mountain roads to analyze the basic characteristics of vehicle motion parameters under different levels of risk. After several analyses, the stability of longitudinal acceleration, lateral
acceleration, yaw rate, and vehicle speed data showed a certain correlation with the risk level. The vehicle kinematics parameter curve for safety-critical events had lower stability, whereas the general events had higher stability, as shown in Figure 6. It is necessary to obtain a stability evaluation index to show that stability effectively reflects driving risk.

![Figure 6. Distribution characteristics of kinematics parameters of different events.](image)

Information entropy is a quantitative index used in information theory to measure the amount of information and degree of order. The more orderly a system is, the lower the information entropy; conversely, the more chaotic a system is, the higher the information entropy. In this study, information entropy was used as an evaluation indicator to characterize the vehicle kinematics parameters’ stability under different risk levels. If a random system can be modeled by a discrete random variable $X$, then the overall information volume of this random system can be represented by information entropy $H(X)$, as shown in Equation (9):

$$H(X) = -\sum_{i} p(x_i) \log p(x_i)$$  \hspace{1cm} (9)

where $p(x_i)$ represents the probability that a random event $X$ is $x_i$.

Vehicle kinematics parameters are discrete time series data, and the grid method is used to calculate information entropy. The X-axis is divided into several grids, where the length of each grid is $t$, the number of non-zero data points $t$ is denoted as $T$; the Y-axis is divided into several grids, where the length of each grid is $s$, and the $i$th grid is marked as $s_i$. The mean value of the data points in each $t$ is calculated. If the mean value falls in $s_i$, the number of data points in $s_i$ is increased by 1, then $p(x_i)$ is the quotient of the number of data points in $s_i$ and $T$, as shown in Equation (10). After testing, we finally chose $t = 1$ and $s = 1$.

$$p(x_i) = \frac{\text{the number of data points in } s_i}{T}$$  \hspace{1cm} (10)

Information entropy was used in this study as the stability evaluation index of the vehicle dynamics parameters and was added to the input variables of the SVM model as the statistical value of vehicle kinematics parameters.
The input variables of each SVM are summarized in Table 5. The vehicle kinematics parameters used were longitudinal acceleration (Xaccel), lateral acceleration (Yaccel), velocity (V), and yaw rate (Z). The statistical values of the vehicle kinematics parameters were average (avg), standard deviation (std), maximum (max), minimum (min), amplitude (amp), and information entropy (entropy). SVM\(_Y\) is the SVM that adds only yaw rate, SVM\(_E\) only information entropy, and SVM\(_{Y&E}\) both yaw rate and information entropy.

Table 5. Summary of input variables of each SVM.

| Model       | Vehicle Dynamic Parameters | Input Variables                                                                 |
|-------------|----------------------------|---------------------------------------------------------------------------------|
| SVM         | Xaccel, Yaccel, V          | Xaccel\_avg, Xaccel\_std, Xaccel\_max, Xaccel\_min, Xaccel\_amp; Yaccel\_avg, Yaccel\_std, Yaccel\_max, Yaccel\_min, Yaccel\_amp; V\_avg, V\_std, V\_max, V\_min, V\_amp; |
| SVM\(_Y\)   |                            | Xaccel\_avg, Xaccel\_std, Xaccel\_max, Xaccel\_min, Xaccel\_amp; Yaccel\_avg, Yaccel\_std, Yaccel\_max, Yaccel\_min, Yaccel\_amp; V\_avg, V\_std, V\_max, V\_min, V\_amp; Z\_avg, Z\_std, Z\_max, Z\_min, Z\_amp; |
| SVM\(_E\)   |                            | Xaccel\_avg, Xaccel\_std, Xaccel\_max, Xaccel\_min, Xaccel\_amp, Xaccel\_entropy; Yaccel\_avg, Yaccel\_std, Yaccel\_max, Yaccel\_min, Yaccel\_amp, Yaccel\_entropy; V\_avg, V\_std, V\_max, V\_min, V\_amp, V\_entropy; |
| SVM\(_{Y&E}\)|                          | Xaccel\_avg, Xaccel\_std, Xaccel\_max, Xaccel\_min, Xaccel\_amp, Xaccel\_entropy; Yaccel\_avg, Yaccel\_std, Yaccel\_max, Yaccel\_min, Yaccel\_amp, Xaccel\_entropy; Yaccel\_avg, Yaccel\_std, Yaccel\_max, Yaccel\_min, Yaccel\_amp, Xaccel\_entropy; V\_avg, V\_std, V\_max, V\_min, V\_amp, V\_entropy; Z\_avg, Z\_std, Z\_max, Z\_min, Z\_amp, Z\_entropy; |

Finally, to verify the influence of the two additional parameters on the SVM identification effect, further calculations were performed, as shown in Table 6.

Table 6. Comparison of the prediction effects of SVM, SVM\(_Y\), SVM\(_E\), and SVM\(_{Y&E}\).

| Models           | Predictive Performance Metrics |
|------------------|--------------------------------|
|                  | A\(_{CC}\)% | A\(_UC\) | R\(_FN\)% | R\(_FP\)% |
| SVM              | 75.66      | 0.8055   | 32.28     | 17.14     |
| SVM\(_Y\)        | 79.03      | 0.8515   | 28.35     | 14.29     |
| SVM\(_E\)        | 83.52      | 0.8749   | 23.62     | 10.00     |
| SVM\(_{Y&E}\)    | 90.64      | 0.9585   | 11.81     | 7.14      |

The ROC curves of the four SVMs are shown in Figure 7. Table 6 and Figure 7 show the results of adding the yaw rate and information entropy as input variables. The trained models SVM\(_Y\) and SVM\(_E\) achieved better identification effects than the original SVM. This shows that the yaw rate and stability can be used as effective features to identify safety-critical events. Comparison of SVM\(_Y\) and SVM\(_E\) showed that the SVM\(_E\) identification effect was superior, showing that the smoothness is an important reflection of driving risk on mountain roads.

For the SVM\(_{Y&E}\) model (penalty factor C = 1, radial kernel function parameter \(\gamma = 0.93\), according to the method in Section 3.2.1), which is the addition of the yaw rate and information entropy as input variables, the high identification accuracy and AUC values were 90.64% and 0.9585, respectively, were achieved. The false-negative and false-positive rates were both low at 11.81% and 7.14%, respectively. Compared with the traditional SVM method, the identification accuracy rate was increased by approximately 15%, the false-negative rate was reduced by approximately 20%, and the false-positive rate was reduced by 10%. Thus, the identification effect of SVM\(_{Y&E}\) was significantly improved compared with that of the original SVM. The SVM model was optimized as a more suitable method for identifying the safety-critical events on mountain roads by introducing input variables related to the driving risk.
Conclusions

This study aimed to investigate traffic accidents on mountain roads. We designed a complete naturalistic-driving experiment platform and collected 8000 km worth of naturalistic-driving data, from which a sample set of traffic incidents was constructed. The classification performances of SVM, BPNN, and CNN on the data were compared, and SVM was identified as the most effective in identifying safety-critical events on mountain roads. Information entropy was introduced as an indicator of driving stability and an additional input variable for the SVM (SVM(E)) model, and the identification accuracy increased by approximately 8% compared with the traditional SVM. This performance improvement showed that stability is an important indicator of safety-critical events on mountain roads. Furthermore, the yaw rate and information entropy were introduced simultaneously as input variables of the SVM. The improved model (SVM(Y&E)) yielded an identification accuracy of 90.64%, which was approximately 15% higher than that yielded by the traditional SVM in identifying safety-critical events. Additionally, the false-positive and false-negative rates of the improved SVM(Y&E) were reduced by approximately 10% and 20%, respectively, compared with the traditional SVM. These results demonstrate that the method developed in this study can accurately and efficiently identify safety-critical events on mountain roads, characterize and analyze traffic accidents, and provide a reliable technical reference for preventing and mitigating the occurrence of traffic accidents. Prevention and mitigation of traffic accidents can reduce loss of life, property damage, environmental pollution, and travel costs, which is conducive to sustainable transportation.

The general conclusions of this study are as follows. 1. Compared with BPNN and CNN, SVM's classification effect is better in the specific scene of mountain roads. 2. In the specific scene of mountain highways, the stability index (information entropy) can be used as an effective feature for safety-critical event identification.

Author Contributions: Conceptualization, Z.W., H.Z., and R.Z.; methodology, Z.W.; software, Z.W.; validation, Z.W., H.Z., and R.Z.; formal analysis, Z.W.; resources, H.Z. and R.Z.; data curation, Z.W.; writing—original draft preparation, Z.W.; writing—review and editing, H.Z. and R.Z.; visualization, Z.W. and R.Z.; supervision, H.Z. and R.Z.; project administration, H.Z. and R.Z.; funding acquisition, R.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Guangzhou Science and Technology Plan Project (Grant No.202007050004) and the National Key Research and Development Program of China (Grant No. 2017YFC0803906).
Data Availability Statement: The data presented in this study are available on request from the corresponding author or the other author. The data are not publicly available due to privacy restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. *Global Status Report on Road Safety 2018*; World Health Organization: Geneva, Switzerland, 2018.
2. Ubiergo, G.A.; Jin, W.L. Mobility and environment improvement of signalized networks through Vehicle-to-Infrastructure (V2I) communications. *J. Transp. Eng.* 2016, 128, 182–190. [CrossRef]
3. Chakrabarti, U.K.; Parikh, J.K. Applying HAZAN methodology to hazmat transportation risk assessment. *Process Saf. Environ. Prot.* 2012, 90, 368–375.
4. Yang, J.; Li, F.; Zhou, J.; Zhang, L.; Huang, L.; Bi, J. A survey on hazardous materials accidents during road transport in China from 2000 to 2008. *J. Hazard. Mater.* 2010, 184, 647–653. [CrossRef] [PubMed]
5. Sikander, G.; Anwar, S. Driver fatigue detection systems: a review. *IEEE Trans. Intell. Transport. Syst.* 2019, 20, 2339–2352.
6. Gao, Z.; Ren, H.; Chen, S.; Chen, Y.; Li, Z.; Zhang, H.; Liu, B. Prediction of Vehicle Rollover State Based on Online Learning of Limit Learning Machine. In Proceedings of the Annual Meeting of China Society of Automotive Engineers, Shanghai, China, 27 October 2020.
7. Meng, X.H.; Guan, Z.Q.; Zheng, L. Safety evaluation of mountain expressway based on geometric alignment index. *China J. Highway Transp.* 2011, 24, 103–108.
8. Trifunović, A.; Pešić, D.; Čičević, S.; Antić, B. The importance of spatial orientation and knowledge of traffic signs for children’s traffic safety. *Accid. Anal. Prev.* 2017, 102, 81–92. [CrossRef] [PubMed]
9. Fitch, G.M.; Hanowski, R.J. Using Naturalistic Driving Research to Design, Test and Evaluate Driver Assistance Systems. In *Handbook of Intelligent Vehicles*; Springer: London, UK, 2012; pp. 559–580.
10. Alexiadis, V.; Colyar, J.; Halkias, J.; Hranac, R.; McHale, G. The next generation simulation program. *ITE J.* 2004, 74, 22–26.
11. Neale, V.L.; Dingus, T.A.; Klauer, S.G.; Sudweeks, J.; Goodman, M. An Overview of the 100-car Naturalistic Study and Findings; National Highway Traffic Safety Administration: Washington, DC, USA, 2005; pp. 1–10.
12. Pedersen, N.J.; Nourredlin, S. The Second Strategic Highway Research Program; Transportation Research Board: Washington, DC, USA, 2013.
13. Zhu, X.; Liu, Z.; Li, L. Open-loop model of driver’s emergency lane changing behavior based on naturalistic driving data. *Automot. J. Saf. Energy Conserv.* 2015, 6, 328–332.
14. Zhu, X.; Liu, Z.; Li, L. Steering collision avoidance control strategy based on dangerous conditions of vehicles and pedestrians. *Automot. J. Saf. Energy Conserv.* 2015, 6, 217–223.
15. Li, L.; Zhu, X.; Dong, X.; Ma, Z. Research on collision avoidance strategy of autonomous emergency braking system. *Autom. Eng.* 2015, 37, 168–174.
16. Wang, X.; Li, Y. Analysis of driver’s lane change characteristics based on naturalistic driving data. *Traffic Inf. Saf.* 2016, 34, 17–22.
17. Zhu, M.; Wang, X.; Wang, X. Car-following headways in different driving situations: a naturalistic driving study. In Proceedings of the 16th COTA International Conference of Transportation Professionals, Shanghai, China, 1–6 July 2016; pp. 1419–1428.
18. Dingus, T.A.; Klauer, S.G.; Neale, V.L. The 100-Car Naturalistic Driving Study. Phase 2: Results of the 100-Car Field Experiment; National Highway Traffic Safety Administration, Department of Transportation: Washington, DC, USA, 2006.
19. Guo, F.; Klauer, S.G.; Hankey, J.M.; Dingus, T.A. Near crashes as crash surrogate for naturalistic driving studies. *Transp. Res. Rec.* 2010, 2147, 66–74.
20. Wu, K.F.; Aguero-Valverde, J.; Jovanis, P.P. Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. *Accid. Anal. Prev.* 2014, 72, 210–218.
21. Perez, M.A.; Sudweeks, J.D.; Sears, E.; Antin, J.; Lee, S.; Hankey, J.M.; Dingus, T.A. Performance of basic kinematic thresholds in the identification of crash and near-crash events within naturalistic driving data. *Accid. Anal. Prev.* 2017, 103, 10–19. [PubMed]
22. Wu, K.F.; Jovanis, P. Screening naturalistic driving study data for safety-critical events. *J. Transp. Res. Board* 2013, 2386, 137–146.
23. Kluger, R.; Smith, B.L.; Park, H.; Dailey, D.J. Identification of safety-critical events using kinematic vehicle data and the discrete Fourier transform. *Accid. Anal. Prev.* 2016, 96, 162–168.
24. Zhang, Y.; Xie, Y. Forecasting of short-term freeway volume with v-support vector machines. *Transp. Res. Rec.* 2007, 2024, 92–99.
25. Chen, S.; Wang, W.; Van Zuylen, H. Construct support vector machine ensemble to detect traffic incident. *Expert Syst. Appl.* 2009, 36, 10976–10986.
26. Li, X.; Lord, D.; Zhang, Y.; Xie, Y. Predicting motor vehicle crashes using support vector machine models. *Accid. Anal. Prev.* 2008, 40, 1611–1618.
27. Duan, L. Driver Mental Workload Evaluation and Application in Driver Assistance System. Ph.D. Thesis, Jilin University, Changchun, China, 8 December 2013.
28. Gao, Z.; Liu, Y.; Zheng, J.Y.; Yu, R.; Wang, X.; Sun, P. Predicting hazardous driving events using multi-modal deep learning based on video motion profile and kinematics data. In Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 3352–3357.
29. Ma, D.; Song, X.; Li, P. Daily traffic flow forecasting through a contextual convolutional recurrent neural network modeling inter-and intra-day traffic patterns. *IEEE Trans. Intell. Transp.* 2020, **99**, 1–10.
30. Breiman, L. Random forests. *Mach. Learn.* 2001, **45**, 5–32.
31. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* 1995, **20**, 273.
32. Chuan, S. Identification and Control for Driving Risk of Vehicle Using Naturalistic Driving Data. Ph.D. Thesis, Wuhan University of Technology, Wuhan, China, 29 November 2016.
33. Wang, X.; Xu, X. Dangerous event recognition method based on naturalistic driving data. *J. Tongji Univ.* 2020, **48**, 51–59.
34. Zhang, R.H.; He, Z.C.; Wang, H.W.; You, F.; Li, K.N. Study on self-tuning tyre friction control for developing main-servo loop integrated chassis control system. *IEEE Access* 2017, **5**, 6649–6660.
35. Sudweeks, J.D. *Using Functional Classification to Enhance Naturalistic Driving Data Crash/near Crash Algorithms*; Virginia Tech Transportation Institute: Blacksburg, VA, USA, 2015.